"SUBSTRATE SPECIFICITY EVOLUTION OF THE ADP-DEPENDENT SUGAR KINASE FAMILY: STRUCTURE-ACTIVITY STUDIES ON ADP-DEPENDENT KINASES FROM HYPERTHERMOPHILIC ARCHAEA"

Tesis entregada a la Universidad de Chile

en cumplimiento parcial de los requisitos

para optar al Grado de

Magíster en Ciencias con mención en

Biología Celular, Molecular y Neurociencias

Facultad de Ciencias

Universidad de Chile

por

Felipe Ignacio Merino León

Septiembre de 2011
Santiago, Chile

Director de Tesis: Dr. Victoria Guixé Leguia

FACULTAD DE CIENCIAS
UNIVERSIDAD DE CHILE
INFORME DE APROBACIÓN
TESIS DE MAGÍSTER.

Se informa a la Escuela de Postgrado de la Facultad de Ciencias de la Universidad de Chile que la Tesis de magíster entregada por el candidato

FELIPE IGNACIO MERINO LEÓN

ha sido aprobada por la comisión de evaluación como requisito para optar al grado de Magíster en Ciencias con mención en Biología Molecular, Celular y Neurociencias, en el examen de defensa privada de tesis realizado el día 6 de Junio de 2011.

Dr. Victoria Guixé Leguia
Directora de Tesis

Comisión de Evaluación de Tesis

Dr. Osvaldo Álvarez
Presidente Comisión

Dr. Francisco Chávez
Corrector

Espero que aquellos que, por alguna razón, se interesen en el conjunto de ideas y discusiones reunidas en este escrito disfruten las siguientes páginas tanto como yo lo he hecho al crearlas

Dedicado a Rebecca

Acknowledgements

Each of the three chapters that contain the bulk of the research inside this thesis have their own acknowledgments section, which principally mention the funding sources and some researchers that were involved in the experimental work, either through their critical opinion or by providing us critical equipment. Nevertheless, i wanted to add this section to acknowledge those who, beyond the mere experimental work, have contributed to my scientific development.

First of all, i would like to express my gratitude to Professor Jorge Babul and especially to Professor Victoria Guixé for their constant guidance and support as well as for the profound trust that they have deposited in my during all the years we have worked together. Also, i am grateful to Ricardo Cabrera and Mauricio Baez for their great scientific contribution as well as their personal support during this work and all the other projects that we have undertaken these years. Of course, i thank all the members of our laboratory, amongst them César Ramírez, Andrés Caniuguir (now working at the Catholic University of Chile), Christian Wilson (now working at the University of California, Berkeley), Víctor Castro, Pablo Villalobos, Diego Quiroga, Andrés Rivas, and Jazmín Herrera who have helped this work by means of experimental support and importantly through their friendship. (I am very sorry if i forgot to mention somebody).

Part of this work was performed at the Physics Institute of São Carlos (University of São Paulo, Brazil). This was possible essentially through our collaboration with

Professor Richard Garratt. His help as well as the collaboration of all the members of the Crystallography and the Molecular Biophysics groups made our visits there highly enjoyable and productive.

Of course, i am grateful for the huge support, trust, and unconditional love that my family has given to me.

Finally, i would like to thank all of those who made this work possible through their financial support. Amongst them, i thank the projects 1070111 and 1110137 of the Fondo Nacional de Desarrollo Científico y Tecnológico. I also thank the Comisión Nacional de Investigación Científica y Tecnológica (Conicyt) for their partial covering of the cost of my Master studies. For this work i visited several times Brazil. Those travel expenses were covered, in part, by a travel award of the AMSUD-PASTEUR network and a travel award given by the Vicerrectoría de Asuntos Académicos of the University of Chile.

Agradecimientos

Cada uno de los tres capítulos que contienen el grueso de la investigación en esta tesis contienen una sección de agradecimientos, las cuales mencionan principalmente las fuentes de financiamiento así como algunos investigadores que han contribuido particularmente a la realización de los experimentos, ya sea con su opinión crítica así como con equipamiento. Sin embargo, quisiera agregar esta sección para agradecer a quienes más allá del mero trabajo experimental hayan contribuido a mi formación como científico.

En primer lugar me gustaría agradecer al profesor Jorge Babul y especialmente a la profesora Victoria Guixé por su constante apoyo y guía así como por la inmensa confianza que han depositado en mí durante los años que hemos trabajado juntos. También, quiero agradecer a Ricardo Cabrera y a Mauricio Baez por su gran contribución científica así como su apoyo personal durante el desarrollo de este trabajo y otros proyectos que hemos llevado a cabo durante estos años. Por supuesto, agradezco a todos los miembros del Laboratorio de Bioquímica y Biología Molecular, entre ellos César Ramírez, Andrés Caniuguir (actualmente en la Universidad Católica), Christian Wilson (actualmente en la Universidad de California, Berkeley), Víctor Castro, Pablo Villalobos, Diego Quiroga, Andrés Rivas y Jazmín Herrera quienes han apoyado este trabajo ya sea directamente con trabajo experimental o con su amistad (Si he olvidado a alguien mil disculpas).

Una parte importante de este trabajo fue realizado en el Instituto de Física de São Carlos (Universidad de São Paulo, Brasil). Esto no habría sido posible sin la colaboración del Profesor Richard Garratt. Él junto con el resto del grupo de cristalografía y el grupo de biofísica molecular del instituto hicieron nuestras visitas allá muy agradables y provechosas.

Por supuesto, quiero agradecer el eterno apoyo que mi familia me ha dado, su confianza en mi juicio y su cariño incondicional.

Finalmente, quisiera agradecer a aquellos que han hecho posible esta tesis con su apoyo monetario. Entre ellos es necesario mencionar los proyectos del Fondo Nacional de Desarrollo Científico y Tecnológico (Fondecyt) 1070111 y 1110137 y la beca de la Comisión Nacional de Investigación Científica y Tecnológica (Conicyt) que costeo parcialmente mis estudios de Magíster. Para este trabajo además realicé varios viajes a Brasil. Ellos fueron costeados en parte por una beca de la Vicerrectoría de Asuntos Académicos de la Universidad de Chile y una beca entregada por la red AMSUD-PASTEUR.

Contents

List of Figures

List of Tables

List of Abbreviations

A_{280}	Absorbance at 280 nm
A_{580}	Absorbance at 580 nm
ADP-GK	ADP-dependent glucokinase
ADP-GK/PFK	ADP-dependent glucokinase/phosphofructokinase
ADP-PFK	ADP-dependent phosphofructokinase
AIRK	Aminoimidazole riboside kinase
ELCS	Explicit likelihood of subset covariation
ewMCM	Ensemble weighted maximally correlated motion
FMA	Functional mode analysis
HPLC	High performance liquid cromatography
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry/calorimeter
k_{cat}	Catalytic constant
K_d	Dissociation constant
KDGK	2-keto-3-deoxygluconate kinases
K_i	Inhibition constant
K_M	Michaelis-Menten constant
LB	Luria Bertani
LINCS	Linear constraint solver
NADH	Reduced nicotinamide adenine dinucleotide
$NAD(P)^+$	Oxidized nicotinamide adenine dinucleotide (phosphate)
NPT	Constant pressure, temperature, and number of particles
NVT	Constant volume, temperature, and number of particles
PCR	Polymerase chain reaction
perGK	C-terminal circular permutant of the ADP-dependent glucokinase from *Thermococcus litoralis*

*Pf*GK	ADP-dependent glucokinase from *Pyrococcus furiosus*
*Ph*GK	ADP-dependent glucokinase from *Pyrococcus horikoshii*
*Ph*PFK	ADP-dependent phosphofructokinase from *Pyrococcus horikoshii*
R_h	Hydrodynamic radius
RMSD	Root mean square deviation
rvET	Real value evolutionary trace
TEV	Tobacco etch virus
*Tl*GK	ADP-dependent glucokinase from *Thermococcus litoralis*

Abstract

In some archaea of the *Euryarchaeota* the glycolytic flux proceeds through a modified version of the Embden-Meyerhof pathway where only four of the ten canonical enzymes are conserved. One of its most important modifications is that the glucokinase and phosphofructokinase activities are ADP-dependent. Sequence and structural studies have demonstrated that these enzymes are homologous to each other. They belong to the ribokinase superfamily, where all the remaining enzymes use ATP as the phosphoryl donor. The most noticeable difference between ADP and ATP dependent kinases of the superfamily is a non-cyclic permutation of the nucleotide binding site. While the ADP-dependent kinases have been mostly studied in thermophilic organisms, the presence of them has been demonstrated in mesophilic archaea and even in eukaryotic organisms. Some explanations for the ADP dependence have been proposed in the literature, yet neither of them have been proven experimentally. Interestingly, the hyperthermophilic archaeon *Methanocaldococcus jannaschii* has just one homolog of these enzymes which presents both, glucokinase and phosphofructokinase activities. On the basis of this feature, it has been proposed that this enzyme represents an ancestral form in the family, which later gave rise to both specificities through a gene duplication event. In this work, we have studied the evolution of the ADP-dependent sugar kinase family, with special emphasis on the structural determinants of the substrate specificity. The results show that this family is closely related to the ATP-dependent

phosphofructokinases of the superfamily. Surprisingly, neither the topology of the nucleotide binding site nor temperature affect the nucleotide specificity of these enzymes. However, the results suggest that the circular permutation has a strong effect on the unfolding kinetics. Inside the ADP-dependent kinase family, the most ancient activity is the phosphorylation of glucose which rejects the bifunctional ancestor hypothesis. Here, the phosphofructokinase specificity seems to be related with the presence of two positively charged side chains which stabilize the negative charge of the ligand in the binding site, while the glucokinase specificity is related with the presence of a glutamic acid residue which interacts with the C2 hydroxyl group of glucose through a hydrogen bond.

Resumen

En algunas arqueas pertenecientes al grupo de las *Euryarchaeotas* el flujo glicolítico procede a través de una versión modificada de la vía de Embden-Meyerhof en donde sólo se conservan cuatro de las diez enzimas canónicas de esta vía. Una de sus principales modificaciones es que tanto las actividades glucoquinasa como fosfofructoquinasa utilizan ADP en vez de ATP como dador de fosforilo. Estudios de secuencia así como estructurales han demostrado que estas quinasas son homólogas entre sí y pertenecen a la superfamilia riboquinasa, donde el resto de sus miembros utilizan ATP como dador de fosforilo. La mayor diferencia entre las quinasas dependientes de ADP y ATP de la superfamilia, es una permutación no cíclica en el sitio de unión a nucleótidos. Si bien las quinasas dependientes de ADP han sido principalmente estudiadas en organismos termófilos, la presencia de ellas también ha sido demostrada tanto en arqueas mesófilas como en eucariontes. En la literatura se han propuesto algunas explicaciones para la dependencia de ADP sin embargo, ninguna ha sido explorada experimentalmente. De manera interesante, la arquea hipertermófila *Methanocaldococcus jannaschii* presenta sólo un homólogo de estas enzimas, el cual presenta actividades glucoquinasa y fosfofructoquinasa. En base a esto se ha propuesto que esta enzima representa una forma ancestral en la familia que luego dio origen a ambas especificidades a través de una duplicación génica. En este trabajo hemos estudiado la evolución de la familia de quinasas dependientes de ADP, con especial énfasis en los determinantes estruc-

turales de la especificidad por sustrato. Los resultados muestran que esta familia está estrechamente relacionada con las fosfofructoquinasas dependientes de ATP presentes en la superfamilia. Sorprendentemente, ni la topología del sitio de unión a nucleótido ni la temperatura afectan la capacidad de estas enzimas para discriminar los nucleótidos. Sin embargo, los resultados sugieren que la permutación circular tiene un gran efecto en la cinética de desplegamiento. Dentro de la familia de quinasas dependientes de ADP, la actividad más ancestral parece ser la fosforilación de glucosa, lo que descarta la hipótesis del ancestro bifuncional. Aquí, la especificidad por fructosa-6-fosfato parece depender de la presencia de dos cadenas laterales con carga positiva que estabilizan la carga negativa en el fosfato del sustrato, mientras que la especificidad por glucosa está relacionada con la presencia de un residuo de ácido glutámico que interacciona mediante puente de hidrógeno con el hidroxilo en el carbono dos del azúcar.

Preface

The ADP-dependent kinases are a poorly understood group of enzymes present in some Archaea and Eukaryotes. Over the years, highly misleading interpretations to the question "why ADP?" has been given in the literature without really answering the question. Interested in this particular issue, we proposed this thesis as a way to gain a better understanding of that question. In the way, we noticed that other problems about the evolutionary history of this group of enzymes were highly interesting and ultimately captured our attention. This, of course, expanded (and somehow twisted) the scope of this work. In particular, we detour our interest from the nucleotide specificity problem to the sugar specificity problem. This leads us to the publication of two articles dealing with the sugar specificity problem from an evolutionary and biophysical perspective.

This thesis contains most of the knowledge that we have gained over the last few years about the ADP-dependent sugar kinase family. As some of the work has already been published, the original articles are included. As such, most of the chapters can be read independently resulting in a highly modular document. However, the chapters will be presented in a way that, while anachronistic, will produce a document that could be read as a whole starting with the most general questions. The document is organized as follows:

Chapter 1. This chapter summarizes briefly what is known to date about the ADP-dependent enzymes with emphasis in their evolutionary relation with other kinases. A brief introduction to the Archaea domain is also given.

Chapter 2. This chapter describes our work related to the phylogenetic relation of the ADP-dependent family with the rest of the ribokinase superfamily and how the nucleotide specificity is related with the protein structure.

Chapter 3. This chapter describes our initial work with the sugar specificity problem from an evolutionary perspective. Here, we particularly deal with the problem of a glucokinase/phosphofructokinase bifunctional enzyme from the archaeon *Methanocaldococcus jannaschii.*

Chapter 4. This chapter describes a collaborative work between the Midwest Center for Structural Genomics, Queen's University, and our laboratory. Here, we fully characterized the ADP-dependent phosphofructokinase from the archaeon *Pyrococcus horikoshii* with special emphasis on the sugar discrimination problem.

Chapter 5. Finally, this chapter discusses the main common issues of the preceding chapters just emphasizing those aspects not mentioned on the specific discussions.

CHAPTER 1

Introduction

1.1 The Archaea domain

Once upon some decades every field of science experiences some sort of revolution which shifts the current paradigm governing the field. Traditionally, biology dealt with the diversity of organism by classifying them. However, without a hypothesis governing the work this resulted in a large collection of knowledge that did not answer any particular question. It was not until the work of people like Lamarck and Darwin that a true solid hypothesis came to biology – evolution.

Up until the late 70's cells were simply divided into prokaryotes and eukaryotes based, as their name suggest, in the presence of a cellular nuclei amongst other characteristics. Even then, it was known that some special kind of, then known as bacteria, were able to grow in extreme conditions such as high temperature and low pH (de Rosa *et al.*, 1975). However, it was not until the pioneering work of Carl Woese that it was noticed that the prokaryotes group was not as simple as it was initially thought. In 1977 Woese and Fox (1977) demonstrated that it was possible to distinguish, phylloge-netically, two main groups of prokaryotes based on ribosomal sequence. One of them contained typical bacteria and the second contained mostly methanogenic organisms (Woese and Fox, 1977). On the basis of this, they were named Eubacteria ("true bac-

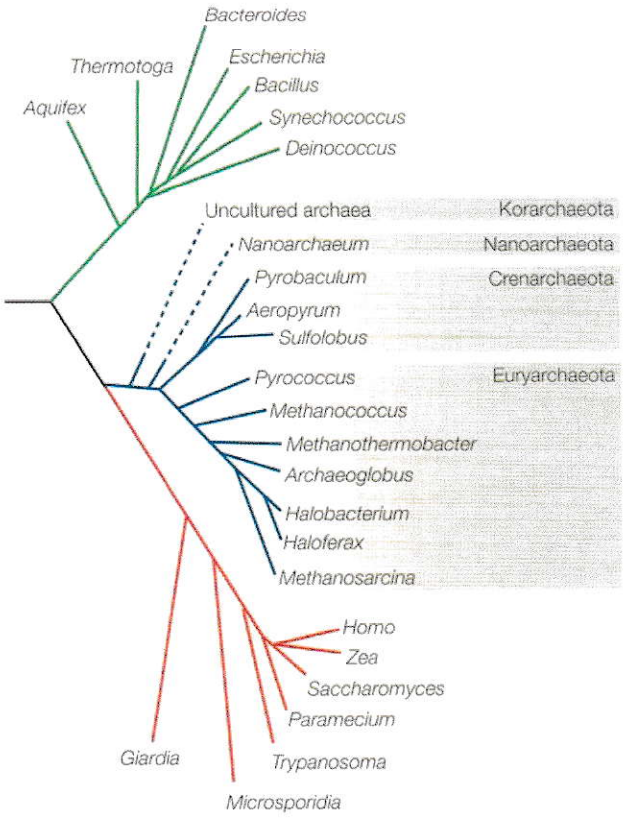


Figure 1.1: Schematic representation of the phylogenetics of the three domains of life. Bacteria are shown in green, Eukarya in red, and Archaea in blue. In dotted lines are shown those branches of the Archaea group that have been recently proposed based on sequence data. This figure was modified from (Allers and Mevarech, 2005).

teria") and Archaeabacteria respectively. In fact, these groups are so different that in 1990, Woese *et al.* (1990) renamed them simple as Bacteria and Archaea to avoid confusions.

Nowadays, these two groups or domains of life, together with the eukaryotes, can be phylogenetically classified as shown in Figure 1.1 (Allers and Mevarech, 2005). Interestingly, although there are some known archaea that growth in mesophilic conditions, most of them are extremophiles. Two main phylogenetic groups can be found inside Archaea: *Euryarchaeota* and *Crenarchaeota* (Allers and Mevarech, 2005) (Figure 1.1). Also, recently based in environmental samples, a third group called *Korarchaeota* has been proposed.

Considering the potential for technological applications, most of the attention has been directed to study those archaea able to grow in extreme temperature conditions (known as thermophiles or hyperthermophiles), extremely high salinities (known as halophiles), extremely low pH (known as acidophiles), and most commonly a combination of them. From the *Crenarchaeota*, the *Sulfolubus* and *Aeropyrum* genera receive lots of attention since both are aerobic thermophilic organisms. In the *Euryarchaeota*, the methanogenic organisms are intensively studied. One of the most studied organism here is *Methanocaldococcus jannaschii*[1] (Jones *et al.*, 1983) since it is one of the few organisms known to be able to produce methane at extreme temperatures. Besides it, the *Halobacterium* and *Haloferax* genera are used as models for halophilic organisms while organisms from the *Thermococcus* and *Pyrococcus* genera are used as models of hyperthermophilic organisms. Here, by far, the most studied organism is *Pyrococcus furiosus*.

One of the hallmarks of the Archaea domain is the unique composition of their cellular membrane. In all known organisms the membrane is built by a glycerol molecule bound to two hydrophobic tails and a polar head. In Bacteria and Eukarya, the hydrophobic tails are composed of fatty acids (generally unbranched) attached to the sn-1 and sn-2 positions of glycerol through an esther linkage. The polar head (a phosphate derivative) is then attached to the sn-3 position. In Archaea on the other hand, the hydrophobic tails are now phyntanyl chains. They are attached to the sn-2 and sn-3 positions of the glycerol moiety through ether bonds and the polar head is attached to the sn-1 position (Koga *et al.*, 1993). Generally, archaeal membranes are composed either of 2,3-O-biphyntanyl-sn-glycerol (known as archaeol), tetra-O-di(biphytanyl)-sn-diglycerol (known as caldarchaeol) or a mixture of them. The latter is really special since it is a very large macrocycle. As such, when membranes are completely com-

[1]This organism was initially named *Methanococcus jannaschii* and was later renamed as *Methanocaldococcus jannaschii* to aknowledge the fact that those organisms from the *Methanococcus* genus are not thermophilic.

posed of this lipid, they are really monolayers. Given this highly specialized structure, these membranes are highly impermeable to small molecules and ions (Yamauchi *et al.*, 1993) which is needed to resist the huge electrochemical potential generated by living in environments with extremely low pH or high salinity. Also, considering the highly branched structure of these lipids, they remain in the liquid phase over an enormous range of temperatures, which enables the use of the same core lipids by psychrophilic and thermophilic organisms (Freisleben *et al.*, 1992). Here, the fluidity of the membrane is controlled by the proportion of archaeol and caldarchaeol (Sprott *et al.*, 1991). In fact, it has been shown that upon temperature or pressure changes, *M. jannaschii* can change from a membrane almost purely made of archaeol to a membrane made of caldarchaeol and a unique macrocyclic archaeol lipid found only in *Methanocaldococcus sp.* (Sprott *et al.*, 1991; Kaneshiro and Clark, 1995).

Metabolically, the members of the Archaea have some features that resemble the eukaryotes while others are more closely related to that of the bacteria. For instance, while the DNA processing machinery is very similar to that of the eukaryotes (e.g. the polymerase) ribosomes are very similar to that of the bacteria. Based on these and other considerations, Margulis (1996) proposed the hypothesis that Eukarya is the results of an endosymbiotic interaction between Bacteria and Archaea.

Sugars are one of the most common carbon sources used by heterotrophic organisms from the three domains of life. There are several known pathways used to produce pyruvate from the incoming sugar (like glucose or galactose). Amongst them, the Embden-Meyerhof pathway, or glycolysis, seems to be the most commonly used. Some microorganisms can also use the Entner-Doudoroff pathway. In the Archaea, both pathways have been observed. For instance, members of the *Thermoproteus*, *Thermoplasma*, and *Sulfolobus* genera degrade glucose through a modified version of the Entner-Doudoroff pathway where sugars are phosphorylated only at the 2-keto-3-deoxygluconate or glycerate level. While the former version is still able to produce one ATP molecule per

glucose the later does not produce any ATP (for a review see Verhees *et al.* (2003)). On the other hand, some members of the *Euryarchaeota* present a highly modified version of the Embden-Meyerhof pathway. Here, only four of the ten textbook enzymes are conserved. In this pathway, the redox reactions are carried out by ferredoxin containing enzymes which latter use the electrons to reduce protons (producing hydrogen) to couple the proton motive force to ATP synthesis by means of a membrane bound hydrogenase enzyme (Sapra *et al.*, 2003). Between the oxido-reductases present in these organisms, perhaps the most interesting is the glyceraldehide-3-phosphate oxido-reductase. This enzyme is responsible for the single-step conversion of glyceraldehide-3-phosphate to 3-phosphoglycerate in a phosphate independent manner (Mukund and Adams, 1995). Besides redox reactions, one of the most striking modifications seen in this version of the Embden-Meyerhof pathway is that the phosphorylation of glucose and fructose-6-phosphate is carried out by enzymes that use ADP and not ATP as the phosphoryl donor (Kengen *et al.*, 1994).

1.2 The ADP-dependent sugar kinase family

Phosphofructokinases have been recognized as one of the key enzymes of glycolysis. Interestingly, while not phylogenetically related, both phosphofructokinase-1 (a member of the PfkA family) and phosphofructokinase-2 (a member of the PfkB family) from *Escherichia coli* present a strong inhibition at high concentrations of their substrate MgATP (Atkinson and Walton, 1965; Kotlarz and Buc, 1981), which suggests that this is a key requirement of this metabolic step. Indeed, it has been already demonstrated that the substrate inhibition is needed for the avoidance of a futile cycle of phosphorylation/dephosphorylation of fructose-6-phosphate/fructose-1,6-bisP which will ultimately lead to a net hydrolysis of ATP (Torres *et al.*, 1997).

Some microorganisms present phosphofructokinases (also members of the PfkA fam-

ily) which use polyphosphates as a source of phosphate and hence they do not appear to be regulated (Peng and Mansour, 1992).

Most of the *Euryarchaeota* members present phosphofructokinases which use ADP as the phosphoryl donor (Verhees *et al.*, 2001). These enzymes are also non-regulated. To avoid net nucleotide hydrolysis these organisms have a much rougher balance between glycolysis and gluconeogenesis. Here, either the phosphofructokinase or the fructose-1,6-bis phosphatase genes are expressed depending on the carbon source (Schut *et al.*, 2003). Besides this ADP dependent enzyme, the phosphorylation of glucose is also ADP-dependent in these organisms. These two kinases are homologous to each other. Since they show no sequence identity, over the noise level, with any other known kinase, they were initially classified as a new protein family, PfkC (Tuininga *et al.*, 1999). The presence of the ADP-dependent kinases has been demonstrated in several members of the *Pyrococcus*, *Thermococcus*, *Methanosarcina*, *Methanosaeta*, *Methanococcoides*, *Methanococcus*, *Methanocaldococcus*, and *Archaeoglobus* genera (Tuininga *et al.*, 1999; Kengen *et al.*, 1994; Koga *et al.*, 2000; Verhees *et al.*, 2001; Hansen and Schönheit, 2004). Also, it has been possible to identify a distant homolog of these enzymes in the genome of higher eukaryotes, which has been proven to be an ADP-dependent glucokinase (Ronimus and Morgan, 2004). Interestingly, the genome of the archaeon *M. jannaschii* presents just one copy of these genes. Surprisingly, the enzyme is able to catalyze the transfer of the β-phosphate of ADP to either glucose or fructose-6-phosphate (Sakuraba *et al.*, 2002). Based on this feature, it was proposed that this enzyme represent an ancestral state of the family, which later gave rise to the separate specificities through a gene duplication event (Sakuraba *et al.*, 2002). However, this hypothesis was never tested.

To date, the crystallographic structures of the ADP-dependent glucokinases from *Thermococcus litoralis* (Ito *et al.*, 2001), *Pyrococcus horikoshii* (Tsuge *et al.*, 2002), *Pyrococcus furiosus* (Ito *et al.*, 2003), and the ADP-dependent phosphofructokinase

from *Pyrococcus horikoshii* (Currie *et al.*, 2009) have been solved. The enzymes present two domains. The larger of them presents an $\alpha\beta\alpha$ fold, where a central β-sheet is flanked by α-helices on both sides (Ito *et al.*, 2001). The small domain is composed of a β-sheet of five strands and some extra α-helices. The active site lies in a crevice between these two domains (Ito *et al.*, 2003). Quite surprisingly, despite the low sequence identity with other known kinases, the ADP-dependent kinases can be classified as members of the ribokinase superfamily.

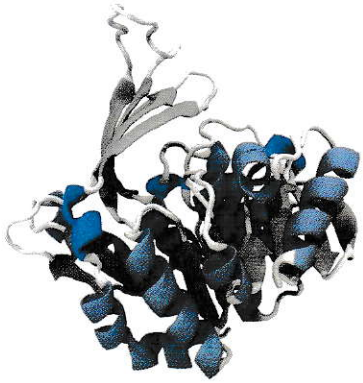
1.3 The ribokinase superfamily

Sugar phosphorylation has been extensively recognized as a key step in many of the known metabolic pathways. Several enzyme families are known to contain kinases which catalyze this type of reactions. Although initially it was proposed that the enzyme families containing the two phosphofructokinases from *E. coli*, PfkA and PfkB, could have a common evolutionary origin (Wu *et al.*, 1991), now we know that they are phylloge-netically unrelated. In the early 1990's Bork *et al.* (1993) noticed that the PfkB group shares a common origin with other sugar kinases such as ribokinase, fructokinases, nucleoside kinases, 6-phosphotagatose kinases, and 1-phosphofructokinases. As the first crystallographic structure was solved in the late 1990's (Sigrell *et al.*, 1998) this analysis was performed only on the basis of sequence data. Later, based on structural data it was possible to add other specificities to the superfamily, such as adenosine kinase[2] (Mathews *et al.*, 1998), 2-keto-3-deoxygluconate kinase (Ohshima *et al.*, 2004), aminoimidazole riboside kinase (Zhang *et al.*, 2004), 4-methyl-5-β-hydroxyethylthiazole kinase (Campobasso *et al.*, 2000), pyridoxal kinase (Li *et al.*, 2002), 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate kinase (Cheng *et al.*, 2002), and the ADP-dependent kinases (Ito *et al.*, 2001). This group is know known as the ribokinase superfamily.

[2]These enzymes have a slightly different fold compared with the other nucleoside kinases from the superfamily mentioned earlier.

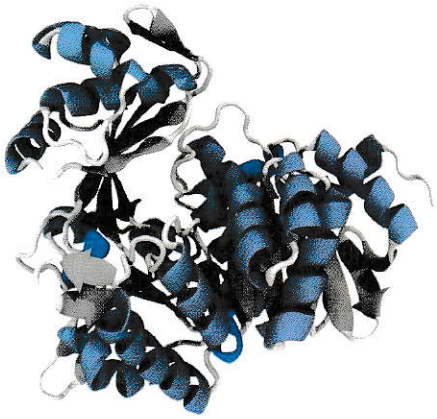
Vitamin kinase like branch

Pyridoxal kinase
Hydroxyethylthiazole kinase
Phosphomethylpyrimidine kinase



PfkB like branch

6-phosphofructokinase
6-phosphotagatose-kinase
1-phosphofructokinase
Nucleoside kinase
Adenosine kinase
Ribokinase
2-keto-3-deoxygluconate kinase
Aminoimidazol riboside kinase



ADP-dependent branch (PfkC)

Glucokinase
6-phosphofructokinase

Figure 1.2: Schematic representation of the three branches of the ribokinase superfamily. For the vitamin kinase like branch the pyridoxal kinase (pdxK) from *E. coli* (PDBID 2DDM) is used as example, for the PfkB like branch the ribokinase from *E. coli* (PDBID 1RKD) is used, and for the ADP-dependent branch the glucokinase from *T. litoralis* (PDBID 1GC5) is shown.

Already based on substrate specificities three major branches can be recognized (Figure 1.2). One of them contains those enzymes that catalyze the transfer of the γ-phosphate of ATP to molecules such as pyridoxal, or pyrimidine derivatives which we know as vitamin kinase like branch. The second contains all the enzymes that catalyze the transfer of the γ-phosphate of ATP to sugar containing molecules, such as fructose-6-phosphate, adenosine, aminoimidazole riboside, etc. We know this as the PfkB like branch. The last of them contains the enzymes that catalyze the transfer of the β-phosphate of ADP to glucose and fructose-6-phosphate which, as was mentioned before is known as PfkC family or ADP-dependent sugar kinase family.

Structurally, the PfkC and PfkB-like branches are enzymes that present two domains. The large domain, which contains the core ribokinase-like fold, is an $\alpha\beta\alpha$ structure where a central β-sheet composed of mainly parallel strands is flanked by α-helices on both sides. Also, they present a smaller β domain which in general is used as a scaffold for dimerization (Sigrell *et al.*, 1998). However, some of the enzymes are monomers. In this case, the hydrophobic core of the small domain is formed by the insertion of some α-helices (Mathews *et al.*, 1998; Ito *et al.*, 2001). On the other hand, the vitamin kinase like branch contains only enzymes which present just the homolog of the large domain.

Zhang *et al.* (2004), based mainly on the presence of the small domain and the monomer complexity, proposed that the most ancient activity of the superfamily should be that catalyzed by the simplest enzyme which is 4-methyl-5-β-hydroxyethylthiazole kinase. In that way, they propose that the increase of complexity in the monomers fold indicates a newer enzyme. By this hypothesis, the ADP-dependent enzymes and the monomeric adenosine kinases should be the newest acquisitions of the superfamily. However, this hypothesis was never tested. Nevertheless, although it could capture the essence of the evolutionary history of this group, considering the linearity of the hypothesis, it is rather unlikely that the true history of the group is entirely represented

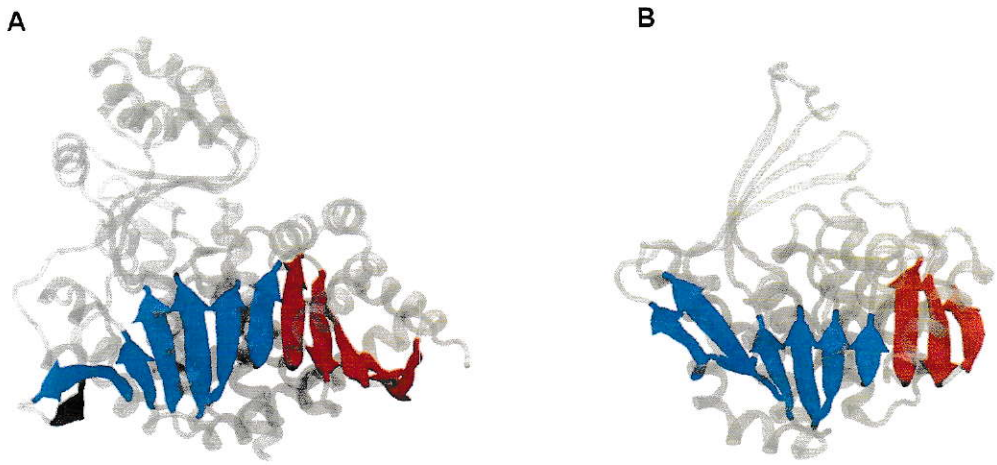


Figure 1.3: Different sectors present in the central β-sheet of the ribokinase-like fold. In blue is shown the rossmann motif while in red is shown the β-meander motif of the ADP-dependent glucokinase from *T. litoralis* (**A**) and the ribokinase from *E. coli* (**B**).

by it.

Based on the topology of the central β-sheet contained in the core ribokinase-like fold a more detailed dissection can be performed. Clearly, the first six to seven strands form a rossmann motif (Figure 1.3) (Cabrera *et al.*, 2010). Interestingly, in this rossmann fold the substrates are not bound to the crevice formed at the C-terminal edge of the sheet, but rather to its side. On the C-terminal end of the protein, a β-meander motif can be found (Figure 1.3) (Cabrera *et al.*, 2010). The phosphoryl acceptor binding site is located on the side of the rossmann region of the protein. Also, some residues present in the acceptor binding site come from the small domain or from another subunit, depending on whether the enzymes present the small domain. On the other hand, the nucleotide binding site is located on the β-meander region of the protein, and thus is always composed of residues from the large domain.

Interestingly, only the ATP-dependent enzymes have a β-meander motif in the C-terminal end of the protein. The ADP-dependent enzymes present a topological re-

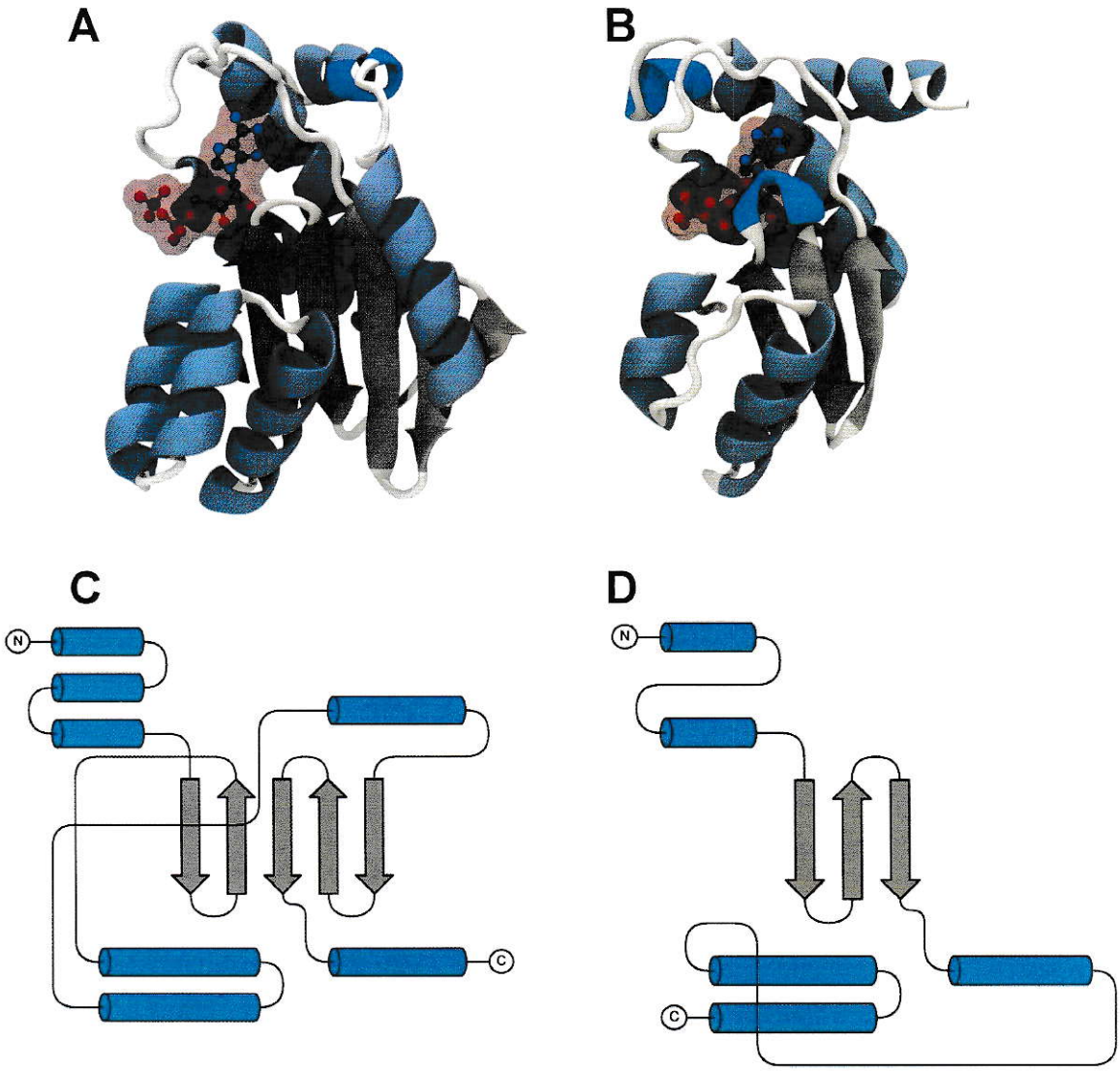


Figure 1.4: Diagram of the β-meander region. The nucleotide binding sites of the ADP-dependent glucokinase from *T. litoralis* (**A**) and the ribokinase from *E. coli* (**B**) are shown. **C** and **D** show topological diagrams of the β-meander region for both enzymes respectively. In both cases the glucokinase from *T. litoralis* is shown from residue 302 to end while the ribokinase from *E. coli* is shown from residue 187 to end.

ordering of the secondary structural elements which produces an equivalent tertiary structure (Figure 2.1). This reordering can be thought as a cyclic permutation of the β-meander region, or equivalently a non-cyclic permutation of the whole protein. As this region of the proteins constitutes almost entirely the nucleotide binding site, and given that the circular permutation is the major structural difference between ADP and ATP-dependent kinases, it could the responsible for the nucleotide specificity. Interestingly, although there is a study of the effect of a non-cyclic permutation on a small protein (Tabtiang *et al.*, 2005), to our knowledge this is the first example of a modification of this kind produced naturally.

1.4 Objectives

1.4.1 Main objective

- The main purpose of this work is to study the structural determinants of the substrate specificity in the ADP-dependent sugar kinase family in an evolutionary context.

1.4.2 Specific objectives

- To study the evolution of the ribokinase superfamily using the bayesian method of phylogenetic inference.
- To study the evolution of the sugar specificity of the ADP-dependent sugar kinase family using the bayesian method of phylogenetic inference.
- To produce and characterize a mutant version of the ADP-dependent glucokinase from *Thermococcus litoralis* containing an ATP-dependent like topology, with emphasis on:
  - The effect of the permutation on the protein folding.
  - The effect of the permutation on the nucleotide specificity of the enzyme.
- To study the structural determinants of the sugar specificity of the ADP-dependent kinases by means of homology modeling and protein-ligand docking using the bifunctional enzyme from *Methanocaldococcus jannaschii* as a model.
- To test the residues found in the objective mentioned above by site-directed mutagenesis using the ADP-dependent phosphofructokinase from *Pyrococcus horikoshii* as a model.

CHAPTER 2

Neither temperature nor topology affect the nucleotide selectivity of the ADP-dependent kinases

Summary

Several archaea of the *Euryarchaeota* present a highly modified version of the Embden-Meyerhof pathway which is composed of only four of the ten canonical enzymes. One of the most interesting modifications here is that the phosphorylation of glucose and fructose-6-phosphate is carried out by two homologous enzymes that use ADP instead of ATP as the phosphoryl donor. While they do not show any significant sequence identity with other known kinases, structurally they can be classified as members of the ribokinase superfamily. The main difference with the other sugar kinases of this superfamily is the presence of a non-cyclic permutation in the C-terminal end of the proteins which constitutes the majority of the nucleotide binding site. In this work we have performed phylogenetic studies of the whole ribokinase superfamily. Also, we tested the effect of the circular permutation by producing a mutant enzyme of the ADP-dependent glucokinase from *Thermococcus litoralis* with ATP-dependent like topology.

We also assayed the effect of temperature on the enzymatic constants of the wild type enzyme. Surprisingly, the results show that the ADP-dependent enzymes are closely related to the other ATP-dependent members of the superfamily which are involved in glycolysis. However, although a very suggestive hypothesis, the topological reordering does not alter the nucleotide selectivity of the enzyme. On the other hand, the circular permutation seems to have a very important effect on the stability, because it produces a protein with an apparent T_m 15 °C lower than the wild type enzyme.

2.1 Introduction

Up until the early 90's it was thought that some archaea of the *Euryarchaeota* used a modified unphosphorylated version of the Entner-Doudoroff pathway to degrade glucose (Mukund and Adams, 1991) which was called pyroglycolysis. However, in 1994 it was possible to demonstrate that, in fact, the flux to pyruvate proceeds through a highly modified version of the Embden-Meyerhof pathway (Kengen *et al.*, 1994). Here, while all the usual reactions take place, just four of the ten textbook enzymes are conserved (Verhees *et al.*, 2003).

One of the most interesting modifications observed is that the phosphorylation of glucose and fructose-6-phosphate is carried out by enzymes that use ADP instead of ATP as the phosphoryl donor (Kengen *et al.*, 1994). Given that these enzymes were discovered in hyperthermophilic organisms, it has been argued in the literature that the main reason for this "ADP-dependence" is the fact that ADP has a higher thermostability than ATP and also that both nucleotides are essentially equivalent since both have a similar ΔG° of hydrolysis. However, these arguments are highly misleading since, (i) as metabolism is a non-equilibrium process the free energy change upon phosphoryl transfers depends on the concentration of the metabolites, (ii) several ATP-dependent enzymes can be found in hyperthermophilic organisms, (iii) the ADP-dependent enzymes are also present in mesophilic organisms, and (iv) the half life of ATP at high temperatures is higher than some other metabolic intermediates present in the Embden-Meyerhof pathway (Dörr *et al.*, 2003).

Sequence studies have shown that both ADP-dependent glucokinases and phosphofructokinases are homologous to each other, but they do not show significant sequence identity to any of the known ATP-dependent enzymes. To date, the crystallographic structures of the ADP-dependent glucokinases from *Pyrococcus furiosus* (*Pf*GK) (Ito *et al.*, 2003), *Pyrococcus horikoshii* (*Ph*GK) (Tsuge *et al.*, 2002), *Thermococcus litoralis* (*Tl*GK) (Ito *et al.*, 2001), and the ADP-dependent phosphofructokinase

from *Pyrococcus horikoshii* (*Ph*PFK) (Currie *et al.*, 2009) have been published. Interestingly, despite the low sequence identity they can be classified as members of the ribokinase superfamily (Ito *et al.*, 2001). Besides the ADP-dependent enzymes, this group contains kinases that can transfer the γ-phosphate of ATP the substrates like ribose, 2-keto-3-deoxyglucose, fructose-6-phosphate, tagatose-6-phosphate, fructose-1-phosphate, pyridoxal, 4-methyl-5-β-hydroxyethyl thiazole, 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate, and several nucleosides.

The ribokinase like fold is composed by an $\alpha\beta\alpha$ sandwich where a central β-sheet is flanked by eight α-helices, three on one side and five on the other (Zhang *et al.*, 2004). Furthermore, the central β-sheet can be divided into a rossmann like sector and a β-meander like sector (Cabrera *et al.*, 2010). Additionally, these proteins can present an extra small domain attached to the rossmann region which is always composed of a four to five strands β-sheet and occasionally some α-helices. The presence of this domain is related to phosphoryl-acceptor binding and to subunit interaction as it serves as interface for some of the dimeric ATP-dependent sugar kinases (Sigrell *et al.*, 1998).

Zhang *et al.* (2004) have proposed an evolutionary hypothesis for the superfamily. On the basis of the structural complexity of the monomers (mainly the presence/absence of the small domain) and the quaternary structure of the enzymes of the superfamily with known structure, they proposed that the more ancient enzymes would be those related with the phosphorylation of coenzymes (like pyridoxal kinase). These enzymes lack the small domain and show the most simple monomer architecture. Then, the tendency in time was to increase the monomer complexity with a concomitant decrease in the aggregation state of the proteins. This puts monomeric enzymes such as adenosine kinases and the ADP-dependent kinases as the most recent acquisitions of the group. However, this hypothesis has been never tested by phylogenetic calculations.

Besides the presence/absence of the small domain the main structural difference between members of the superfamily is the topology of the β-meander region. In fact,

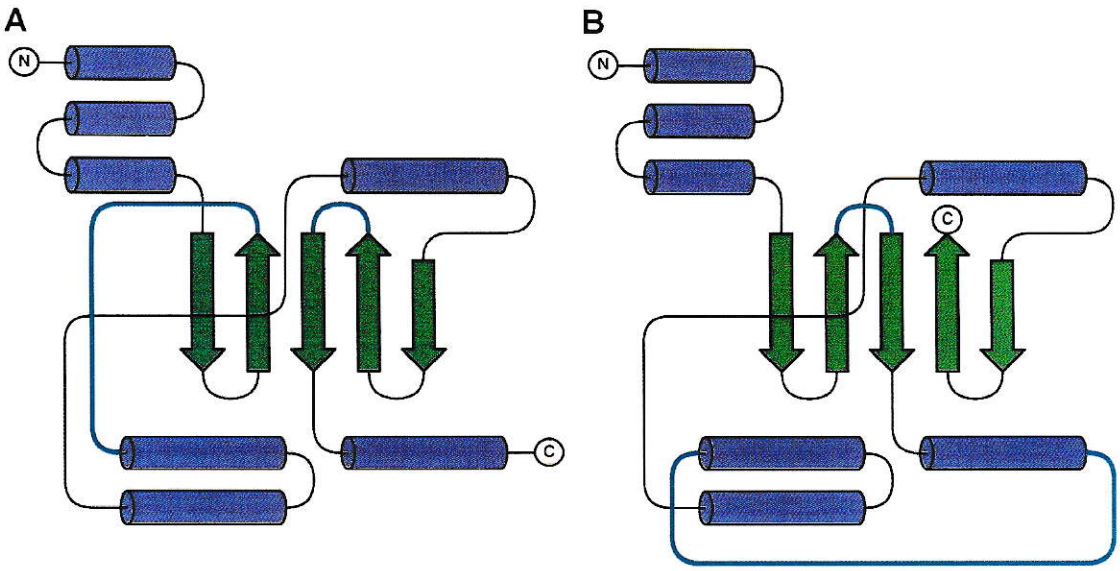


Figure 2.1: Diagram of the perGK enzyme construction. Residues between 302 and the end of the protein are shown **A.** *Tl*GK. **B.** perGK. For simplicity only the C-terminal region of the protein is shown. In blue are shown those connections that are interchanged to produce the new topology. For both enzymes the representation starts at residue 302.

only the ATP-dependent proteins have a true β-meander motif at the C-terminal end. On the other hand, while the secondary structural elements of the C-terminal region of the ADP-dependent enzymes are packed in the same way as the ATP-dependent enzymes, the connectivity of them differs significantly. Using the glucokinase from *T. litoralis* as a guide, the rearrangement can be understood as a cyclic permutation of the region between residue 362 and the end of the protein producing a non-cyclic permutation of the whole structure. Considering that this region constitutes almost entirely the nucleotide binding site, this strongly suggests that the nucleotide specificity is determined by the C-terminal topology.

In this work, we have performed phylogenetic calculations for the ribokinase superfamily. Also, we tested the effect of temperature and the C-terminal circular permutation on the nucleotide specificity and on the folding of the ADP-dependent enzymes using the glucokinase from *T. litoralis* as a model. To this end, we created a mutant

enzyme using the scheme shown in Figure 2.1. Although the C-terminal region of the ADP-dependent enzymes does not possess a true β-meander motif, it will be named β-meander region for consistency with the superfamily. The results show that, as opposed to the Zhang's hypothesis, the ADP-pendent enzymes are closely related to the phosphofructokinases of the superfamily. Surprisingly, both temperature and the circular permutation have almost no effect on the nucleotide specificity. On the other hand, the topology reordering seems to produce a big change on the unfolding kinetics of the protein.

2.2 Experimental procedures

2.2.1 Phylogenetic analysis of the ribokinase superfamily

To study the evolution of the ribokinase superfamily and the appearance of the ADP-dependent activity, the Bayesian method of phylogenetic inference, as implemented in the MrBayes software, was used (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). Initially, 41 structures from the 3 major branches of the ribokinase superfamily (4 ADP-dependent, 12 ATP-dependent vitamin kinase like, and 25 ATP-dependent PfkB like) were structurally aligned. Given the low sequence and low structural conservation at this level of evolutionary relationship, it was not possible to obtain a good dendrogram based just on sequence. To overcome this problem the following approach was used. Besides the few regions that can be structurally aligned directly, the C-terminal helices that are involved in the circular permutation were arranged in the ADP-dependent enzymes to coincide with the ATP topology. Also, as it has been suggested before by Zhang *et al.* (2004) the small domain can be used as a good phylogenetic marker. To exploit this feature, a sequence alignment of the small domain was also included. Finally, a matrix of discrete (standard) data representing the presence/absence of secondary structural motifs in the small domain was included as an extra partition.

For the sequence data, a mixed model of evolution with gamma distribution of mutation rates amongst sites was used. For the morphological partition the standard discrete model available in MrBayes was used.

2.2.2 Cloning of *T. litoralis* glucokinase and perGK genes

The expression vector pET17b containing the gene for the ADP-dependent glucokinase from *Thermococcus litoralis* was a kind gift from Dr. Takayoshi Wakagi (Biotechnology Department, University of Tokyo). *E. coli* BL21(DE3) pLysS cells were transformed

with this construct and stored at -80 °C with 50% glycerol for further usage.

The gene for the perGK mutant was synthesized *de novo* by GenScript[1] and sent to our laboratory inside the pUC57 vector. The codon usage was optimized for protein expression in *E. coli*.

The gene was later subcloned using the Gateway system (Invitrogen). To this end, the gene was extracted from the original vector by PCR, using the *Pfu* DNA polymerase (Fermentas). The resulting product was ligated to the pENTR™/TEV/TOPO® vector, which adds to the N-terminal end of the produced protein a proteolytic site for the tobacco etch virus (TEV) protease.

Finally, the gene was transferred to the pDEST™17 vector using the LR Clonase™ II enzyme mix (Invitrogen). This vector adds, additionally, a coding sequence for a hexa-histidine tag in the the 5' and of the gen. In this way, the histidine tag can be used for purification and can be later removed by proteolysis.

BL21(DE3) pLysS *E. coli* cells were transformed with the resulting plasmid. These bacteria were stored at -80 °C until their use.

2.2.3 Determination of the protein concentration

*Tl*GK and perGK concentrations were determined spectrophotometrically by the sample absorbance at 280 nm using a theoretical extinction coefficient of 50,310 M^{-1} cm^{-1} which was calculated directly from protein sequences using the ExPASy server[2].

2.2.4 *Thermococcus litoralis* glucokinase purification

For the purification of this enzyme, 50 ml of Luria Bertani (LB) broth containing 100 μg/mL ampicillin and 35 μg/mL chloramphenicol were inoculated with 50 μL of the corresponding strain. The culture was grown overnight at 37 °C. Then, these 50

[1] `http://www.genscript.com/`

[2] The server is available at `http://ca.expasy.org/tools/protparam.html`

mL were inoculated in 1 L of LB broth containing 100 μg/mL ampicilin and 35 μg/mL chloramphenicol. The culture was allowed to reach mid-exponential phase ($A_{580} \sim 0.5$). Here, protein overexpression was induced for the next 18 hours by the addition of 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG). The culture was harvested by centrifugation at 3,220 g for 15 minutes.

Cells were resuspended in 50 ml of buffer TrisHCl 50 mM pH 7.8, $MgCl_2$ 5 mM (Buffer A) and they were disrupted by sonication. The debris was removed by centrifugation at 8,000 g for 10 minutes. The supernatant was incubated at 90 °C for 30 minutes and then it was centrifuged for 15 minutes at 3200 g. Then, the soluble part was saturated with 60% ammonium sulfate which was maintained at 4°C with permanent stirring for 1 hour. Precipitated protein were removed by centrifugation at 8,000 g for 10 minutes. The remaining solution was loaded into a Phenyl sepharose HP column (GE Healthcare) pre equilibrated with Buffer A plus 60% ammonium sulfate. Then, the column was washed with 5 resin volumes of the same buffer. The protein was eluted by making a linear gradient of ammonium sulfate from 60% to 0% on buffer A. The fractions with the highest activity were pooled and dialyzed against Buffer A for at least 8 hours at room temperature.

After this, the protein was charged into a HiTrap Q column (GE Healthcare) previously equilibrated with Buffer A. The column was washed with 10 resin volumes of Buffer A and then the protein was eluted using a linear gradient from 0 to 1 M KCl with the same buffer. The fractions with the highest activities were pooled, concentrated, and stored at 4 °C in 20% glycerol.

2.2.5 perGK mutant purification

For this enzyme, bacterial pellets were prepared using the same protocol described above except that protein overexpression was induced using 0.1 mM IPTG and then the culture was grown at 18 °C for 24 hours.

Cells were resuspended in 50 mL of phosphate buffer 20 mM pH 7.4, $MgCl_2$ 5 mM, imidazol 20 mM, and NaCl 0.3 M (Buffer B) and disrupted by sonication. The cell debris was removed centrifuging the solution 10 minutes at 8,000 g. Later, the supernatant was incubated at 70 °C for 20 minutes and then unfolded proteins were removed by 15 minutes of centrifugation at 3,200 g.

The remaining solution was charged into a HisTrap^{TM} column (GE Healthcare) previously equilibrated with Buffer B. The column was washed with 5 resin volumes of Buffer B and the the protein was eluted using a linear gradient between 20 and 500 mM imidazol in the same buffer. The fractions with the highest activities were pooled and immediately mixed with the TEV protease (a ratio 1:20 protease mass:perGK mass was used). The reaction was allowed to proceed overnight at 34 °C.

Later, the solution was dialyzed against buffer TrisHCl 25 mM pH 7.8 for, at least, 4 hours. Then, the protein was charged into a HiTrap Q column (GE Healthcare), washed and eluted essentially in the same way as the native glucokinase. The fractions with the highest activities were pooled, concentrated, and stored at 4°C in 20% glycerol.

2.2.6 Native hydrodynamic radius determination

The native hydrodynamic radius (R_h) of *Tl*GK and perGK were determined using size exclusion chromatography. The experiments were performed using a Bio-Sil SEC-250 column (7,8 mm x 30 cm) (BioRad, Hercules, CA, USA) with a Waters 1525 HPLC system and equilibrated with buffer TrisHCl 50 mM pH 7.4, $MgCl_2$ 5 mM, and 0.1 KCl 0.1 M. All runs were performed using a flux of 0.8 ml/min.

A R_h standard containing Vitamin B-12 (8.5 Å R_h), horse myoglobin (19 Å R_h), chicken ovalbumin (30.5 Å R_h), bovine gamma globulin (41.8 Å R_h), and bovine thyroglobulin (85 Å R_h) was used to calibrate the system.

Sample R_h values were calculated from the elution volume. To this end, the protein elution volumes were converted to R_h values using the relation:

$$R_h = -51.2\sqrt{-\log\left(\frac{V_e - V_o}{V_t - V_o}\right)} - 11.6 \tag{2.1}$$

which was obtained from the run with the standard. Here V_e is the elution volume, V_o is the void volume, and V_t is the total volume of the column (Uversky, 1993).

2.2.7 Circular dichroism measurements

All circular dichroism measurements were performed on a Jasco J-815 spectropolarimeter coupled to a Peltier temperature control system. Typically, protein concentrations between 2 and 3 μM were used with a cuvette of 1 mm pathlength. All experiments were performed in phosphate buffer 25 mM pH 7.5.

For spectral measurements, 16 accumulations were used with a scan velocity of 50 nm/min and a response time of 2 seconds.

For thermal unfolding experiments, both proteins were first mixed with 3 M guanidinium hydrochloride. Only in this way it is possible to observe the unfolding transition at temperatures within the range allowed in the Peltier system. All transitions were monitored by the change in ellipticity at 222 nm. To average out some of the instrumental noise, a response time of 4 seconds was used.

2.2.8 Isothermal titration calorimetry measurements

All isothermal titration calorimetry (ITC) experiments were performed on a VP-ITC (MicroCal). For the determination of the binding constant of the *Tl*GK-MgADP complex the protein was first dialyzed against buffer HEPES 25 mM pH 7.8 (Buffer C), $MgCl_2$ 5 mM over night. For each experiment 30 injections were performed to reach a molar ratio between MgADP and *Tl*GK of ~3. Given the low solubility of this enzyme, experiments were performed with protein concentrations of ~ 140 μM which corresponds to approximately 8.8 times the estimated K_d.

To obtain a direct measurement of the phosphotransfer reaction catalyzed by *Tl*GK and perGK ITC was also used. The apparent molar enthalpy of reaction was measured by injecting the protein to the reaction cell containing 4 mM glucose, 20, 75, 150, or 250 μM ADP, and 5 mM $MgCl_2$ over the nucleotide concentration in Buffer C. In all cases the final protein concentration in the reaction cell was 10 nM.

To measure the kinetic parameters for MgADP, the multiple injection method was used (for a detailed explanation see appendix A). For all experiments a MgADP 20 mM solution was added to the syringe. Three injections of 0.5 μL, followed by four injections of 1 μL and nine injections of 2 μL were used. The reaction cell contained glucose 4 mM, $MgCl_2$ 5 mM and protein concentrations between 3.5 and 1 nM (depending on the temperature of the assay) on Buffer C.

Given that none of the enzymes is active with ATP, inhibition kinetic experiments were used to determine its binding constant. To this end, essentially the same experiment described above was performed, but including fixed concentrations of MgATP. The K_i was later estimated by doing a global fit of the data to a competitive inhibitory mechanism using SigmaPlot 11 (Systat Software, Inc., CA).

2.2.9 Molecular dynamics simulations

To study in more detail the structural dynamics of *Tl*GK and how this is modified by the C-terminal circular permutation NPT Langevin dynamics simulations were carried out using the GROMACS 4.5 (van der Spoel *et al.*, 2005; Hess *et al.*, 2008) program with the CHARMM27 force field (MacKerell Jr *et al.*, 1998).

A model of the perGK enzyme was created from the native enzyme using MODELLER 9 (Sali and Blundell, 1993). Both proteins were solvated with dodecahedral box of ~20,000 TIP3 water molecules. Systems were neutralized with the appropriate number of sodium ions. Periodic boundary condition was used in all directions. Short range interactions were truncated at 1.2 nm, using a switching function starting at 1.0

nm for the van der Waals potential. Long range electrostatics were calculated using the Particle Mesh Ewald method.

Temperature was maintained at 313 K using a damping coefficient of 0.5 ps^{-1} while pressure was maintained at 1 atm using the Parrinelo-Rahman barostat. All bonds involving hydrogen were restrained to their equilibrium length with the LINCS algorithm which enables the use of a 2 fs integration step. 42 ns of simulation were produced for each system. In both cases, the first 2 ns were used as equilibration phase.

To study heat induced denaturation two extra simulations were carried out. Here a water box of ~30,000 molecules was used to account for the increase of the protein volume upon denaturation. First, systems were equilibrated, 1 ns for *Tl*GK or 5 ns for perGK at 298 K using the NVT ensemble. Later, the temperature of the bath was changed to 550 K and the density was allowed to reach the value of liquid water at this temperature using a NPT ensemble. In this way, the system experiences a smooth expansion[3]. Later, the systems were simulated for 30 ns using again the NVT ensemble.

All analysis were performed using either the GROMACS suite (van der Spoel *et al.*, 2005) or the VMD package (Humphrey *et al.*, 1996).

[3]This process is rather fast, usually it does not take longer than 5 ps.

2.3 Results and discussion

2.3.1 Phylogenetic analysis of the ribokinase superfamily

Table 2.1 shows the enzymes used in the phylogenetic inference done for the superfamily. They include 4 ADP-dependent enzymes, 12 vitamin kinase like enzymes, and 25 PfkB like enzymes. Overall, the tree successfully grouped the enzymes on the basis of the presence/absence of the small domain (Figure 2.2).

The blue clade on Figure 2.2 corresponds to those enzymes lacking the small domain. In general, three groups can be seen: A group including just pyridoxal kinases (2DDM, 1TD2, 1LHP, and 2F7K), a group that contains phosphomethyl pyrimidine kinases and the pyridoxal kinase from *Bacillus subtilis* (1UB0, 2I5B, and 1JXH), and a group containing 3 proteins with unknown function and two hydroxyethylthiazole kinases. The fact that the *Bacillus subtilis* pyridoxal kinase does not belong to the pyridoxal kinase clade could be produced by the small amount of information used to construct the tree. Nevertheless, the posterior probability of this association is quite high, suggesting the need of a detailed revision of this association. However, this point goes far beyond the scope of this work.

PfkB like enzymes are shown in red. Several small clades of enzymes can be recognized. A group containing monomeric adenosine kinases (1BX4 and 1LII) and aminoimidazole riboside kinase (AIRK) (1TZ6) close to a clade containing mainly 2-keto-3-deoxygluconate kinases (KDGK) (2QCV, 2VAR, 2DCN, 2AFB, and 1V1A) can be seen on the left side of Figure 2.2.

With the exception of monomeric adenosine kinases, all PfkB like enzymes are, at least, dimers which use the small domain as interface. It seems like the amount of communication between subunits depends on the orientation of the interface. For instance, enzymes like Pfk-2 from *E. coli* shares several key residues for regulation and catalysis between subunits (Cabrera *et al.*, 2011). On the other hand, enzymes like

Table 2.1: Crystal structures used in the phylogenetic analysis of the ribokinase superfamily.

	PDB Code	Organism	Function
PfkC like	1UA4	*Pyrococcus furiosus*	Glucokinase
	1GC5	*Thermococcus litoralis*	Glucokinase
	1L2L	*Pyrococcus horikoshii*	Glucokinase
	1U2X	*Pyrococcus horikoshii*	Fructose-6-phosphate kinase
Vitamin kinase like	1JXH	*Salmonella typhimurium*	4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate kinase
	1EKQ	*Bacillus subtilis*	Hydroxyethylthiazole kinase
	1V8A	*Pyrococcus horikoshii*	Hydroxyethylthiazole kinase
	1UB0	*Thermus thermophilus*	Phosphomethylpyrimidine kinase
	1LHP	*Ovis Aries*	Pyridoxal kinase
	1TD2	*Escherichia coli*	Pyridoxal kinase (PdxY)
	2DDM	*Escherichia coli*	Pyridoxal kinase (PdxK)
	2F7K	*Homo sapiens*	Pyridoxal kinase
	2I5B	*Bacillus subtilis*	Pyridoxal kinase
	1KYH	*Bacillus subtilis*	Unknown function
	2AX3	*Thermotoga maritima*	Unknown function
	2R3B	*Enterococcus faecalis*	Unknown function
PfkB like	2AFB	*Thermotoga maritima*	2-keto-3-deoxygluconate kinase
	2VAR	*Sulfolobus solfataricus*	2-keto-3-deoxygluconate kinase
	2DCN	*Sulfolobus tokodaii*	2-keto-3-deoxygluconate kinase
	1V1A	*Thermus thermophilus*	2-keto-3-deoxygluconate kinase
	2QCV	*Bacillus halodurans*	5-dehydro-2-deoxygluconate kinase
	1TZ6	*Salmonella enterica*	Aminoimidazol riboside kinase
	1BX4	*Homo sapiens*	Adenosine kinase
	1LII	*Toxoplasma gondii*	Adenosine kinase
	2PKN	*Mycobacterium tuberculosis*	Adenosine kinase
	2C49	*Methanocaldococcus jannaschii*	Nucleoside kinase
	1RKD	*Escherichia coli*	Ribokinase
	1VM7	*Thermotoga maritima*	Ribokinase
	2FV7	*Homo sapiens*	Ribokinase
	2QHP	*Bacteroides thetaiotaomicron*	Fructokinase
	2HW1	*Homo sapiens*	Ketohexokinase
	2ABQ	*Bacillus halodurans*	Fructose-1-phosphate kinase
	2F02	*Enterococcus Faecalis*	Tagatose-6-phosphate kinase
	2JG1	*Staphylococcus aureus*	Tagatose-6-phosphate kinase
	3CQD	*Escherichia coli*	Fructose-6-phosphate kinase
	3BF5	*Thermoplasma acidophilum*	Unknown function
	1VK4	*Thermotoga maritima*	Unknown function
	2NWH	*Agrobacterium tumefaciens*	Unknown function
	2RBC	*Agrobacterium tumefaciens*	Unknown function
	2AJR	*Thermotoga maritima*	Unknown function
	2JG5	*Staphylococcus aureus*	Unknown function

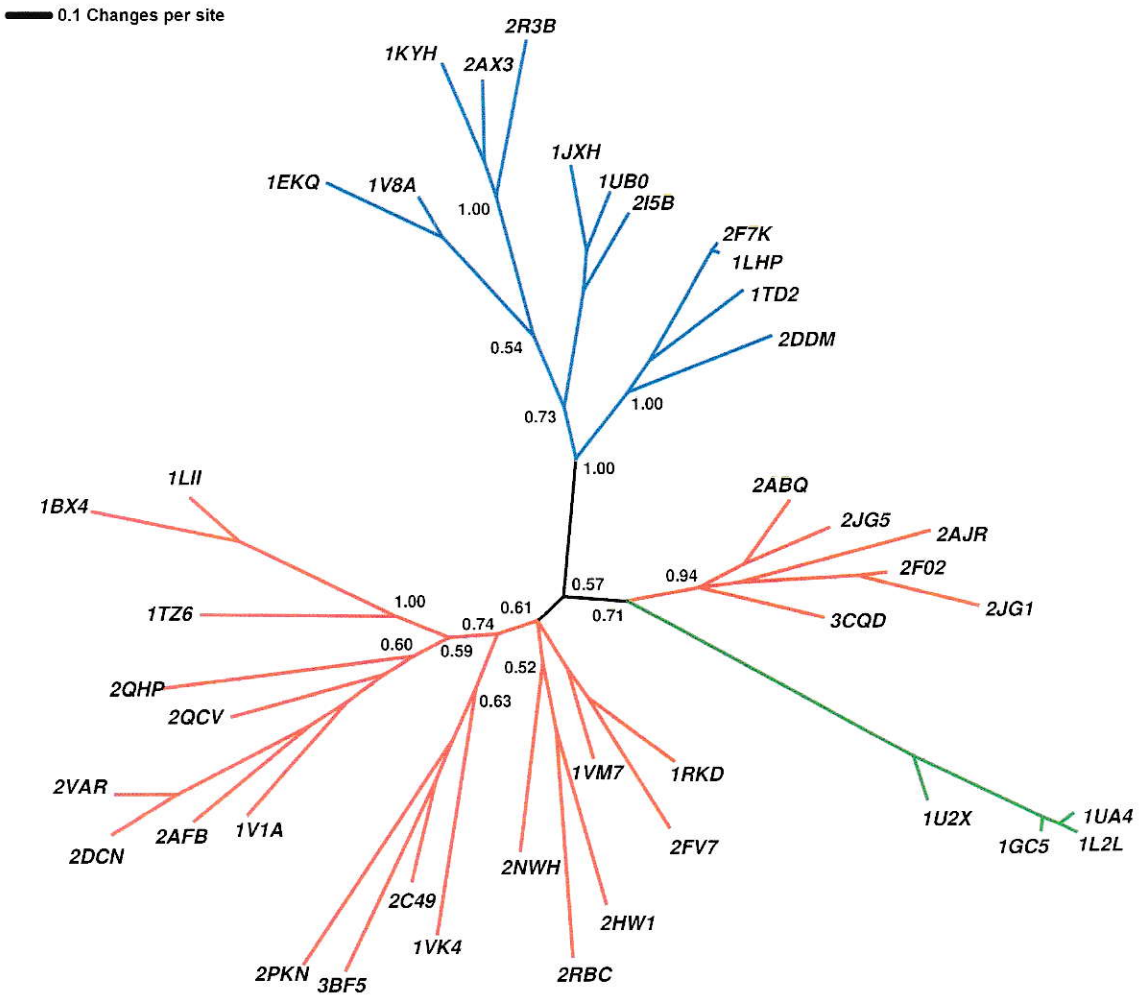


Figure 2.2: Dendrogram for the whole ribokinase superfamily. In green are shown the ADP-dependent enzymes (PfkC like), in red are shown the ATP-dependent sugar and nucleoside kinases (PfkB like), and in blue are shown those enzymes that lacks the small domain (Vitamin kinase like). For display constraints, only the PDB code of each leave is shown in the tree. To know their meaning please see Table 2.1. The posterior probability of some key nodes is also shown. This figure was prepared with Dendroscope (Huson *et al.*, 2007).

KDGK's or AIRK have their subunits rotated in comparison to Pfk-2, which diminishes the communication between subunits and the area of the interface (Figure 2.3). This increase in the independence of the subunits seems to produce ultimately monomeric enzymes like adenosine kinases as suggested by Figure 2.2.

Indeed, similar to the above mentioned observation, Zhang *et al.* (2004) proposed an increase in monomer complexity and a decrease in aggregation state for the modern enzyme versions of this superfamily, which has been proven to be a general trend of protein superfamilies (Fong *et al.*, 2007). However, in their hypothesis ADP-dependent enzymes should group close to adenosine kinases which is not seen in Figure 2.2.

Ribokinases (1VM7, 1RKD, and 2FV7) are grouped with two proteins with unknown function (possible also ribokinases: 2RBC and 2NWH) and an enzyme just described as ketohexokinase (2HW1). Unfortunately, given the amount of information used for this tree, it was not possible to get rid of the polytomy on the origin of this clade. Yet, it is not important for the analysis made in this study.

Curiously, the adenosine kinase (or more general the nucleoside kinase) activity appears twice in this superfamily. There is a small clade grouping some dimeric nucleoside kinases (2C49 and 2PKN) with two enzymes with unknown function (3BF5 and 1VK4). Contrary to monomeric adenosine kinases, these enzymes are characterized by a wide specificity (Hansen *et al.*, 2007) which not only includes nucleosides, but also sugar-phosphates (Hansen and Schönheit, 2001).

The final clade is formed by enzymes that catalyze the phosphorylation of a sugar-phosphate substrate (2ABQ, 2JG5, 2JG1, 2F02, 2AJR, and 3CQD). These enzymes are mainly related with central pathways of sugar degradation such as glycolysis. Interestingly, the green clade shows the ADP-dependent sugar kinase family, which, as opposed to Zhang's hypothesis (Zhang *et al.*, 2004), is more related to sugar kinases that to adenosine kinases. In this fashion, considering that most likely the root of the dendrogram should be between the 0.57 and 1.0 posterior probability nodes, monomeric

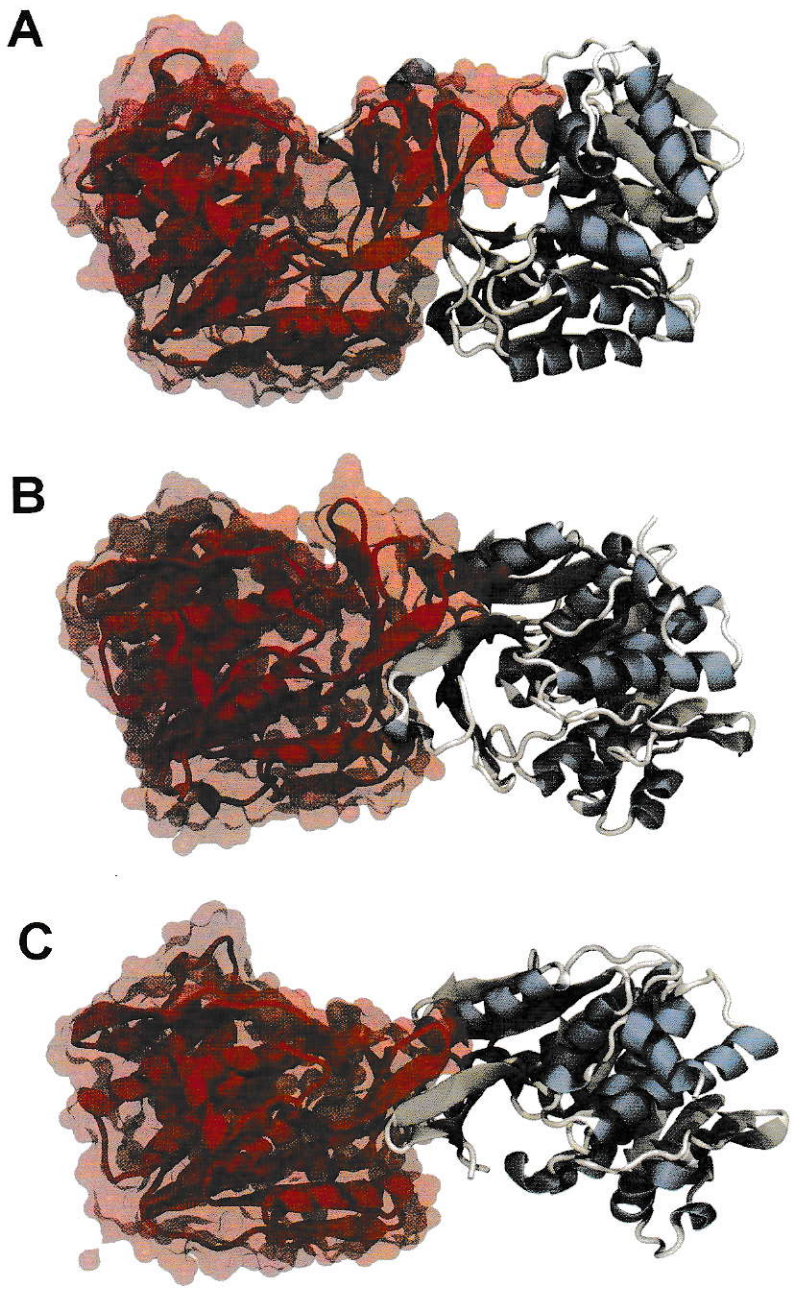


Figure 2.3: Dimeric association for different members of the PfkB like group. **A.** Phosphofructokinase-2 from *E. coli*. This enzyme hides 1429.2 Å^2 of surface per monomer upon dimerization. **B.** 2-keto-3-deoxygluconate kinase from *T. thermophilus*. This enzyme hides 1169.9 Å^2 of surface per monomer upon dimerization. **C.** aminoimidazole riboside kinase from *S. enterica*. This enzyme hides 750.0 Å^2 of surface per monomer upon dimerization. Interface surfaces were calculated using the PISA server (Krissinel and Henrick, 2007)

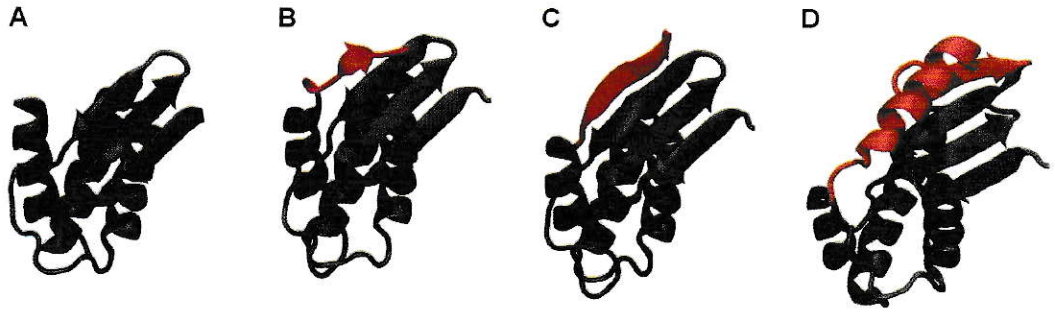


Figure 2.4: β-meander region of several members of the ribokinase superfamily such as Pfk-2 from *E. coli* (**A**), a fructose-1-phosphate kinase from *B. halodurans* (**B**), a putative phosphofructokinase from *S. aureus* (**C**), and the ADP-dependent glucokinase from *P. furiosus* (**D**). In red is shown the C-terminal extension thought to be needed for the circular permutation.

enzymes are first generated in the PfkC family, and later in adenosine kinases through the slow loss of interface complexity.

Besides the aggregation state, the main difference between ADP-dependent enzymes and ATP-dependent enzymes is the ordering of the C-terminal secondary structure elements. This results in a localized circular permutation at the β-meander part of the protein (For a detailed description of the circular permutation see Chapter 1) which maintains the overall tertiary structure. In order to do this, enzymes should extend the C-terminal end. Figure 2.4 shows that indeed, some of the members of the sugar-phosphate clade present an extra β-strand a the carboxilic end which is compatible with this rearrangement.

2.3.2 Circular permutation of the glucokinase from *T. litoralis*

Given that the C-terminal circular permutation is localized exactly in the nucleotide binding site, and given that this is the major difference between ATP and ADP dependent enzymes, we decided to study the effect of this rearrangement using the glucokinase from *T. litoralis* as a model.

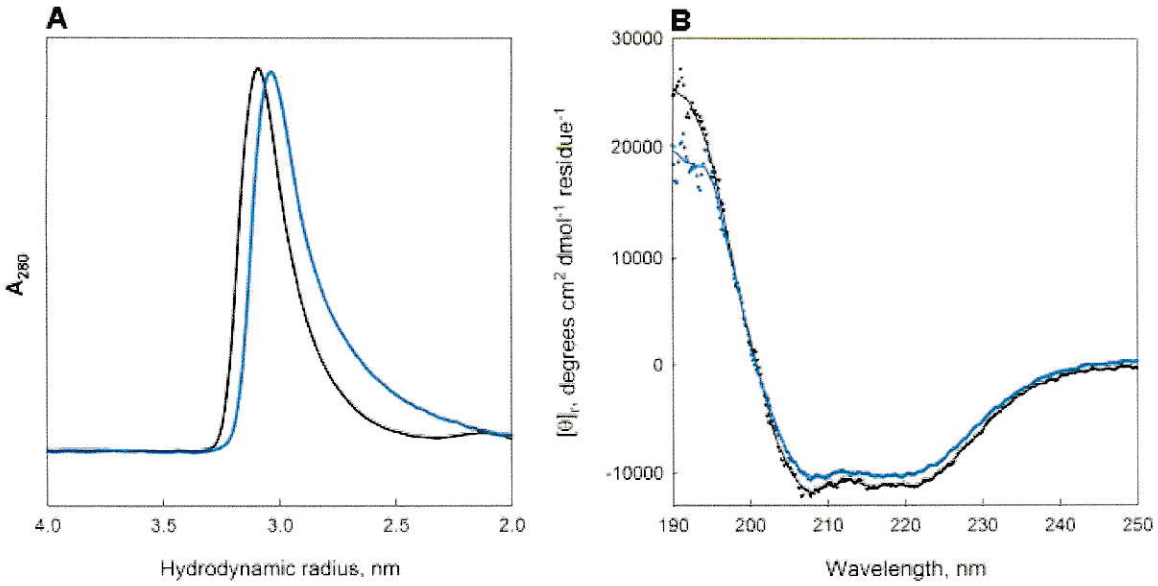


Figure 2.5: Structural properties of *Tl*GK and perGK. **A.** Size exclusion chromatogram. **B.** Circular dichroism spectra. In both cases, the black line corresponds to *Tl*GK while the blue line corresponds to perGK.

Even when the gene for perGK was optimized for expression in *E. coli* the produced mutant enzyme is highly toxic and most of it produces inclusion bodies. However, a very small fraction of the protein remains soluble yielding approximately 125 μg of pure protein per liter of culture.

Figure 2.5 shows some general structural properties of the protein. The size exclusion chromatography showed that the wild type enzyme has a R_h of 30.9 Å while perGK has a R_h of 30.3 Å. Besides this small difference, the peak of perGK has a bigger tail which suggests the presence of some non-specific interaction with the matrix. Nevertheless, both proteins are monomers with similar hydrodynamic behavior. Figure 2.5B shows the circular dichroism spectra of both protein. perGK shows a little less signal than the wild type enzyme. However, both spectra are essentially the same, which suggests that both protein have the same secondary structure content.

Given that both enzymes appears to be equivalent on the secondary and quaternary structural level we performed molecular dynamics simulations to study in more detail

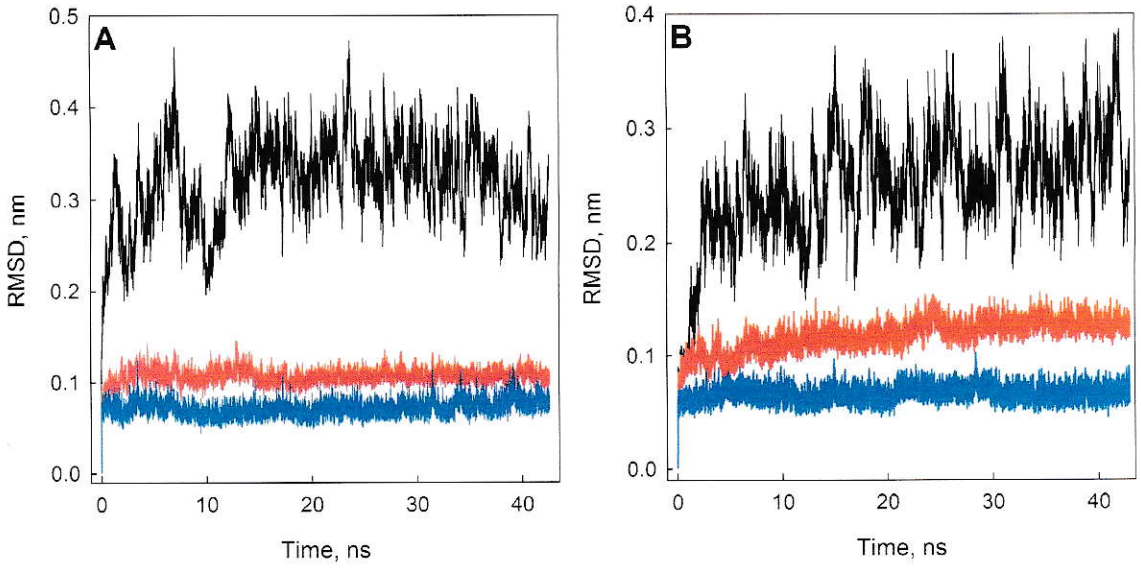


Figure 2.6: Evolution of the α-carbons root mean square deviation through the simulation of *Tl*GK (**A**) and perGK (**B**).

the effect of the change in topology introduced by the permutation.

Figure 2.6 shows the root mean square deviation (RMSD) of the α-carbons respect to the initial frame through the simulation. Both protein have strong fluctuations in the simulation time, as judged by the oscillations of the RMSD for the whole structure (approximately 0.2 nm wide). This seems to be produced mainly by the change in the angle between the domains. Indeed, if we plot the RMSD of each domain respect to the first frame it appears very stable over the entire simulation. This shows that both domains behave essentially as rigid bodies.

Interestingly, the angle between both domains explored by both proteins is always higher that the observed for the x-ray structure of the glucokinase from *T. litoralis*. This shows that, as suggested before (Tsuge *et al.*, 2002), the protein in the absence of the ligands can explore conformations with a higher angle between the domains, and it suggests that, at least, MgADP is needed to visit the conformations with the lowest angles. However, we cannot rule out the possibility that the intermediate conformation

attributed to the binding of ADP in *Tl*GK is caused by crystal packing.

As the oscillation between domains has been previously suggested as one of the key aspects of the catalytic mechanism of these enzymes (Ito *et al.*, 2003) we decided to analyze in more detail the dynamics of *Tl*GK and perGK by means of the principal component analysis module contained in their GROMACS package (van der Spoel *et al.*, 2005; Hess *et al.*, 2008).

For both enzymes the first three eigenvectors account for approximately 50% of the total positional variance of the mainchain. In both cases, the first eigenvector ($\sim$ 35% of the variance for *Tl*GK and 25% for perGK) describes the bending of the domains (Figure 2.7). This is in good agreement with the observation made above, suggesting that this type of fluctuation dominates the dynamics of both proteins.

The second and third eigenvectors are related mainly to the twisting of the domains. First, let us consider the hinge axis as a vector going through the center of mass of the hinge between the domains, which can be used as a reference rotation axis for the domain bending. Then, while the first eigenvector describe a rotation parallel to the hinge axis, the second and third principal components of the dynamics describe motions that are perpendicular to the hinge axis. For both proteins, these are additionally necessary to close the proteins (Figure 2.7). It is important to note that while a fully closed protein is never visited during the simulation time, the principal component analysis suggests that it could be visited, but with a very low likelihood.

For the principal component analysis it is necessary to build a variance/covariance matrix first. The diagonal of this matrix contains the variance in the position (relative to their mean) of each atom. Then the trace of this matrix (or equivalently, the sum of all eigenvalues) can be used as a measurement of the flexibility of the molecule. However, to be meaningful, it is possible to compare matrices build for the same number of atoms which is (almost) the case between *Tl*GK and perGK. For *Tl*GK the trace of the covariance matrix is 17.8419 nm^2 while for perGK is 19.0806 nm^2. This shows that,

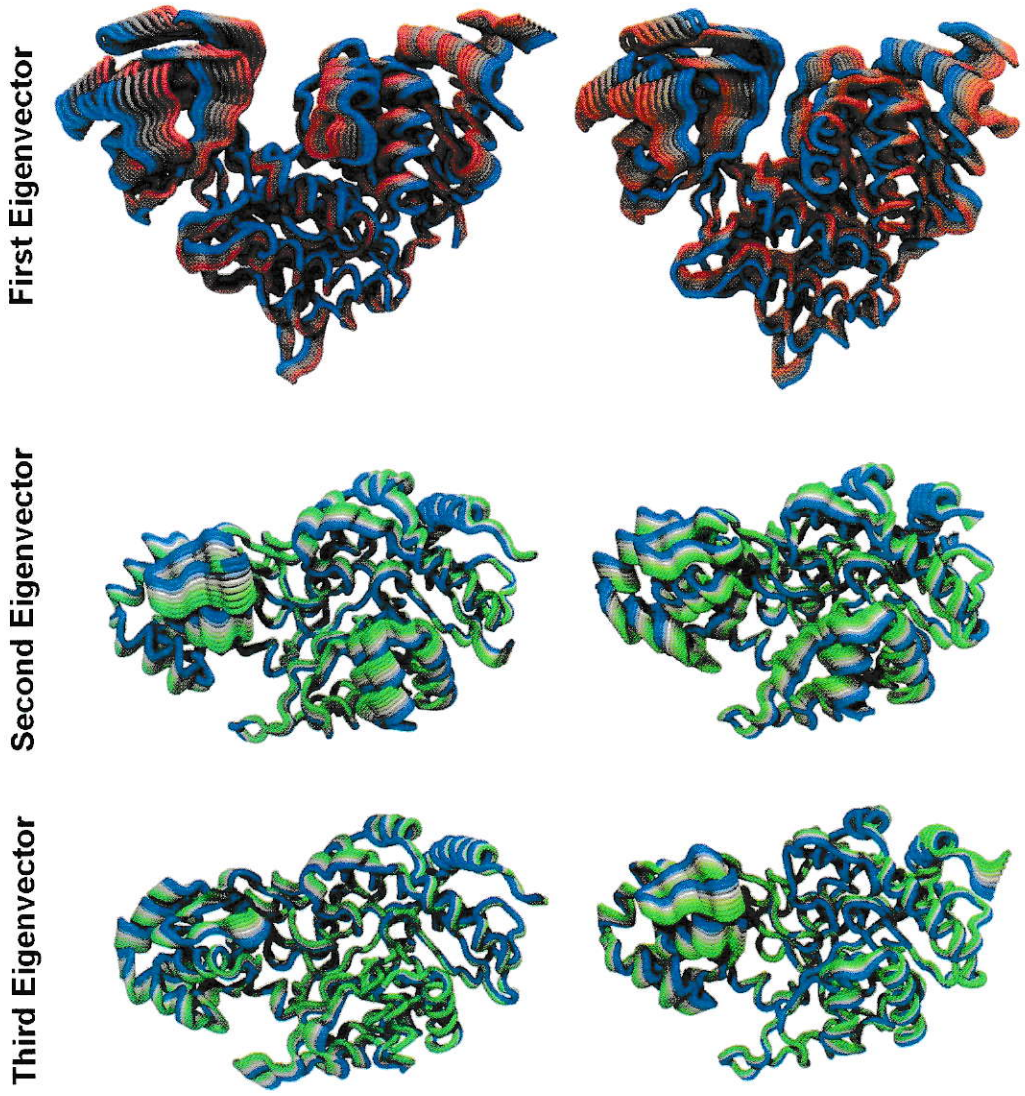


Figure 2.7: First three principal components describing the structural fluctuations of *Tl*GK (left) and perGK (right). For clarity, structures are colored according to their position through the eigenvector.

while the dynamics of both proteins are very similar, perGK is slightly more flexible than *Tl*GK.

Recently, Hub and de Groot (2009) have developed a method called functional mode analysis (FMA) which use the eigenvectors obtained by means of principal component analysis, normal mode calculations, full correlation analysis (Lange and Grubmüller, 2008), etc, as a basis to describe a functionally relevant structural property of a protein. In their formulation they either optimize the Pearson correlation coefficient between the model and the data or the mutual information between the two. While the first method relies on linear correlations the second one can take care of any type of correlation. For both methods, a small fraction of the data is left out for cross validation. As in this work we are only using the Pearson version, the coefficient obtained for both data sets are called R_m and R_c, which stands for model-building set and cross-validation set respectively.

Figure 2.8 shows the FMA of the domain bending seen in *Tl*GK and perGK. As it was already mentioned, the first eigenvector can describe most of the change in the angle between domains. Indeed, a model built using just the first 10 eigenvectors successfully captures almost the whole functional mode as can be seen in Figure 2.8C and D. Interestingly, while for *Tl*GK the motion is mostly captured in the first principal component, for perGK it is a little more distributed within the first few eigenvectors (Figure 2.8A and B).

To study in more detail the dynamics of the nucleotide binding site in these enzymes we used the root mean square distance of the residues in this crevice[4] as a functional mode. Contrary to the domain angle case, here the models were very poor. However, including the first 50 eigenvectors it is possible to construct a model with a R_m of 0.64 and a R_c of 0.52 for *Tl*GK (Figure 2.9A). On the other hand, it is not possible to build

[4]All the heavy atoms of the residues which have at least one atom inside a 7 Å radius sphere from the ADP in the crystal structure of *Tl*GK (or the equivalent residues in perGK) were used in the analysis.

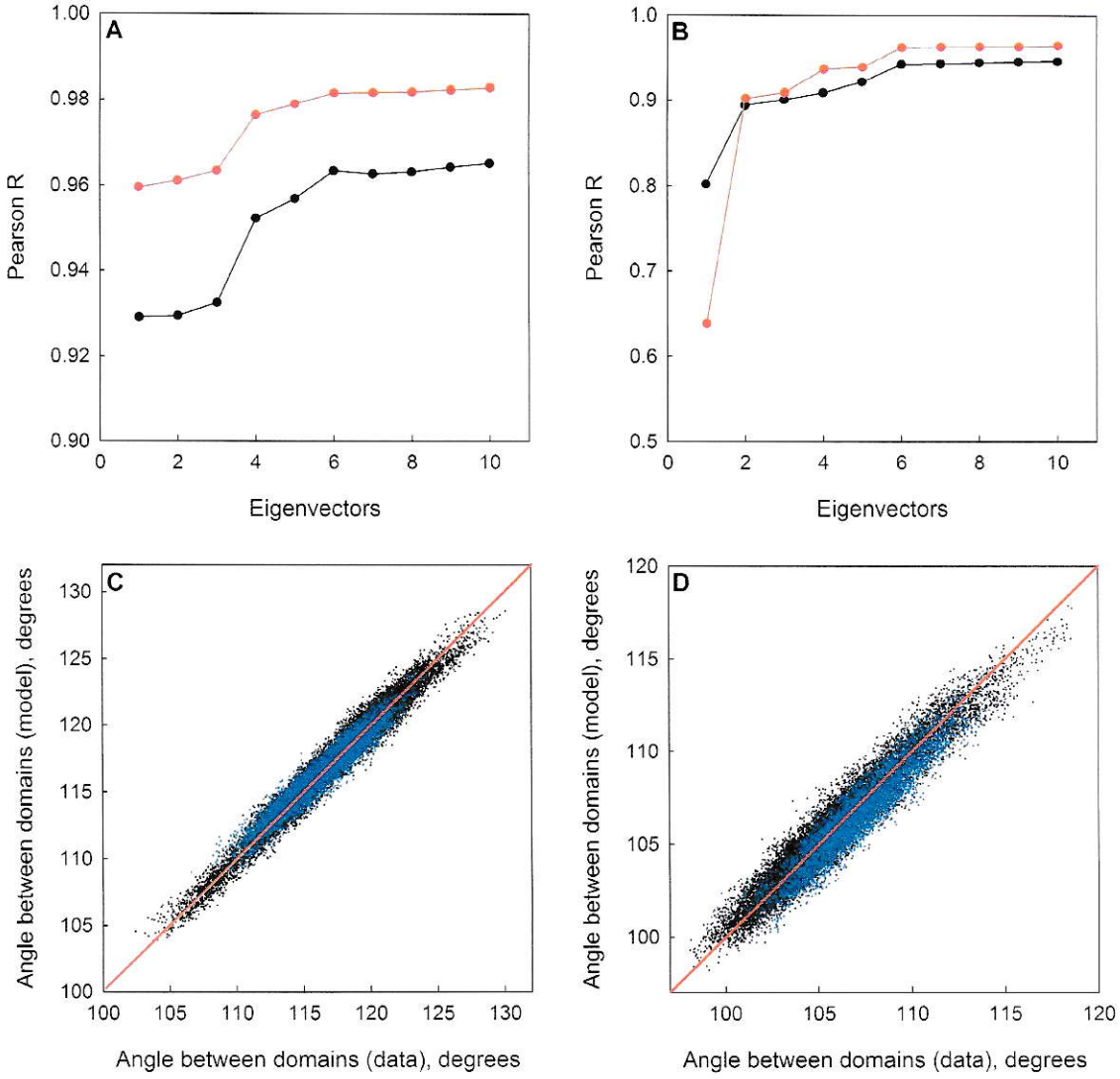


Figure 2.8: FMA of the angle between the domains. Models were constructed using the first 10 eigenvectors. **A** and **B** show R_m (red) and R_c (black) for *Tl*GK and perGK respectively as a function of the number of eigenvectors used in the model building procedure. **C** and **D** compare the models with the data for *Tl*GK and perGK respectively. In black are shown the data used for model construction and in blue the data left for cross validation. See text for further details.

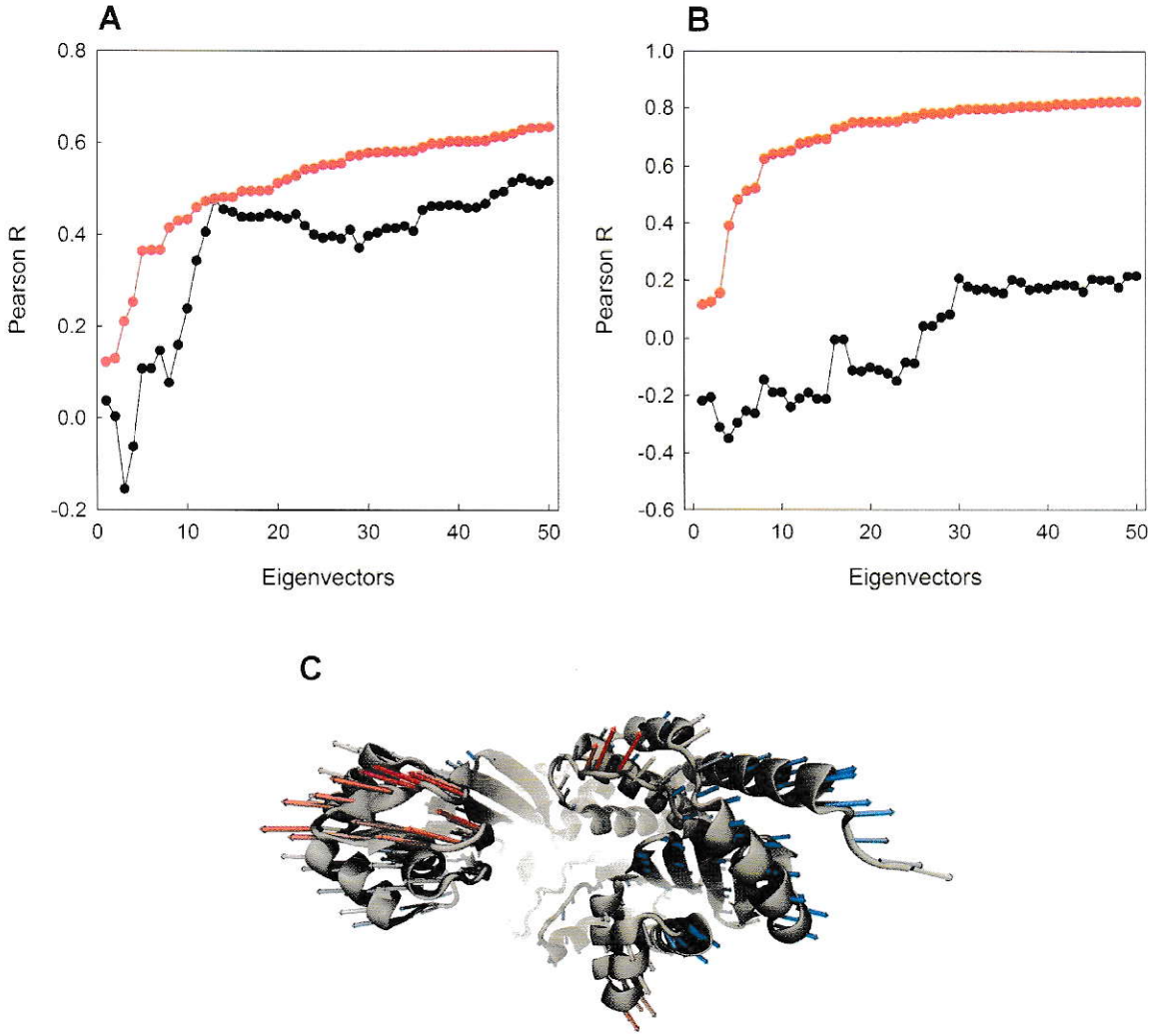


Figure 2.9: FMA of the fluctuation of the ADP binding site. Models were constructed using the first 50 eigenvectors. **A** and **B** show R_m (red) and R_c (black) for *Tl*GK and perGK respectively as a function of the number of eigenvectors used in the model building procedure. **C** shows the ensemble weighted maximally correlated motion that explains the fluctuation of the ADP binding site in *Tl*GK. The arrows indicates the direction of the motion of the atoms. Both arrow size and color is proportional to the displacement. The color scale goes from blue → white → red.

a model for this motion in perGK as it can be seen in Figure 2.9B. Here, while R_m reaches high values, R_c remains always close to 0 which shows that the model is not able to capture the movement.

It has been pointed out by Hub and de Groot (2009) that while it is possible to find a combination of eigenvectors that produces a motion that is maximally correlated with a functional quantity, the motion parallel to this new vector must not be representing a true motion of the protein given the restrictions imposed by the energy landscape of the molecule. In this case, the protein should follow a path through a vector with a small angle respect to the originally found, but which is not hindered by energetic frustration. This kind of vector was called ensemble weighted maximally correlated motion (ewMCM) in the original article (Hub and de Groot, 2009).

Figure 2.9C shows the ewMCM for the motion associated with the ADP binding site of *Tl*GK. As expected, it involves mostly nucleotide binding site in a way that could be thought as a "breathing" of the crevice.

2.3.3 Heat induced denaturation of *Tl*GK and perGK

While the circular permutation has little effect on the structural properties of the enzyme, the topology rearrangement could have a big effect on the folding of the protein. To study this aspect we performed heat induced denaturation experiments using the ellipticity at 222 nm as a structural probe.

*Tl*GK has a extraordinary heat tolerance. Previous differential scanning calorimetry experiments performed with this enzyme have shown that the melting temperature (T_m) is close to 105 °C (not shown). As the peltier device which controls the temperature in the spectropolarimeter is limited to 90 °C, all unfolding experiments were performed in the presence of 3 M guanidinium chloride. This concentration is high enough to ensure that both proteins are unfolded below 90 °C, but on the other hand they still maintain their native structure at moderated temperatures.

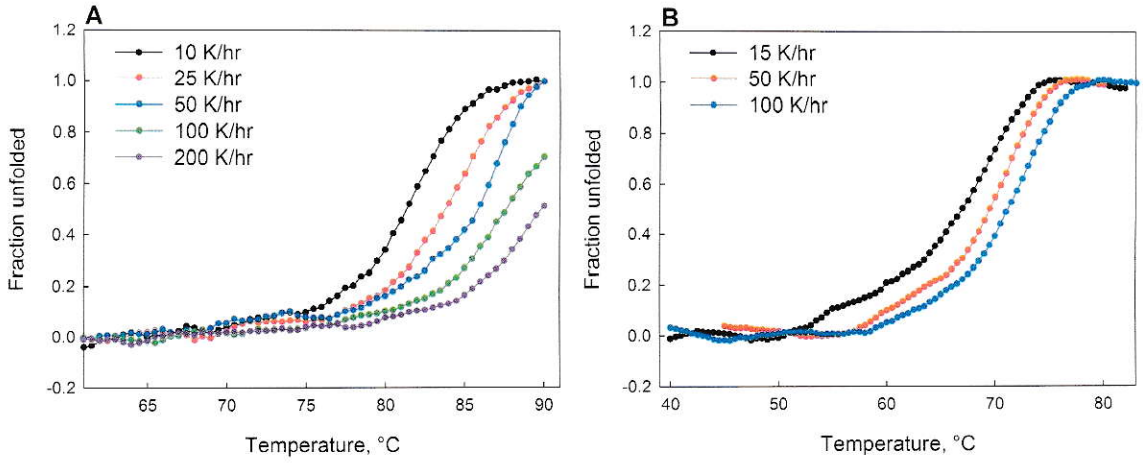


Figure 2.10: Heat induced unfolding at different scan rates. **A.** *Tl*GK. **B.** perGK.

Both proteins unfold irreversibly, irrespective of the scan rate used (not shown). As a non-equilibrium process, the observed unfolding transition should depend on the scan rate. Interestingly, perGK has always a lower apparent T_m when compared to *Tl*GK. (Figure 2.10). For instance, the apparent T_m for *Tl*GK at a scan rate of 50 K hr^{-1} is 85.4 °C while the apparent T_m for perGK at the same scan rate is 69.2 °C (Figure 2.10). This decrease in thermophilicity is in good agreement with the small increase in flexibility seen by molecular dynamics since, thermotolerance has been related before to protein rigidity (Mamonova *et al.*, 2010).

The dependence of T_m with the scan rate is much higher in *Tl*GK. It has been already recognized before that one of the key characteristics of hyperthermophilic enzymes is a remarkably slow unfolding process (Luke *et al.*, 2007). Considering the the circular permutation produces a high decrease in thermostability and a change in the way that T_m depends on the scan rate (Figure 2.10), the results suggest that this topological modification has a strong impact on the unfolding kinetics of the enzyme.

A more detailed comparison of both plots in Figure 2.10 suggests that while for *Tl*GK the unfolding seems to be a two state process, perGK seems to accumulate an intermediate (for example, between 55 and 60 °C in the 15 K hr^{-1} curve). The origin

of this is still unclear and currently this feature is under investigation.

To gain a better understanding of the process we performed molecular dynamics simulations of the early stages of the heat induced unfolding. After 30 ns simulation at 550 K, both protein have lost some structure, although much of the secondary structure is still present (Figure 2.11). This approach for the simulation of heat induced unfolding has been used extensively by other (Li and Daggett, 1994; Duan and Nilsson, 2005; de Bakker *et al.*, 1999; Lazaridis *et al.*, 1997; Lazaridis and Karplus, 1997; Fulton *et al.*, 1999). Interestingly, the time needed to see the unfolding event is much lower than the observed here, suggesting that *Tl*GK and perGK have a very slow unfolding transition which is in good agreement with their thermostability. However, the studies mentioned above have been performed using, generally, small proteins which could also be the reason of the faster process.

Most of the loss of secondary structure is due to the breaking of the strands (Figure 2.11). perGK shows the highest structural loss in the C-terminal region which is not the case of *Tl*GK (Figure 2.11). This is due to the topological rearrangement which has enabled the carboxilic extension shown in Figure 2.4 to unfold almost independently.

For both proteins the unfolding process seems to be highly modular. As the small domain appears as an extension of the rossmann part of the protein the unfolding of these two sectors are highly synchronized in *Tl*GK (Figure 2.12). In the first 3 ns, both the rossmann and the small domain experience a fast, but discrete loss of secondary structure, which is then maintained for the next 12 ns. Then both regions of the protein start to loss structure until they reach a fraction of secondary structure between 0.3 and 0.4 at 30 ns. On the other hand, the β-meander region of the protein maintains almost all its structure during the simulation. The main loss of structure here occurs as a steady and small decrease in the first 15 ns.

In the case of perGK the secondary structure content of the small domain only decrease about 20% in the first 3 ns and then it remains stable for the rest of the

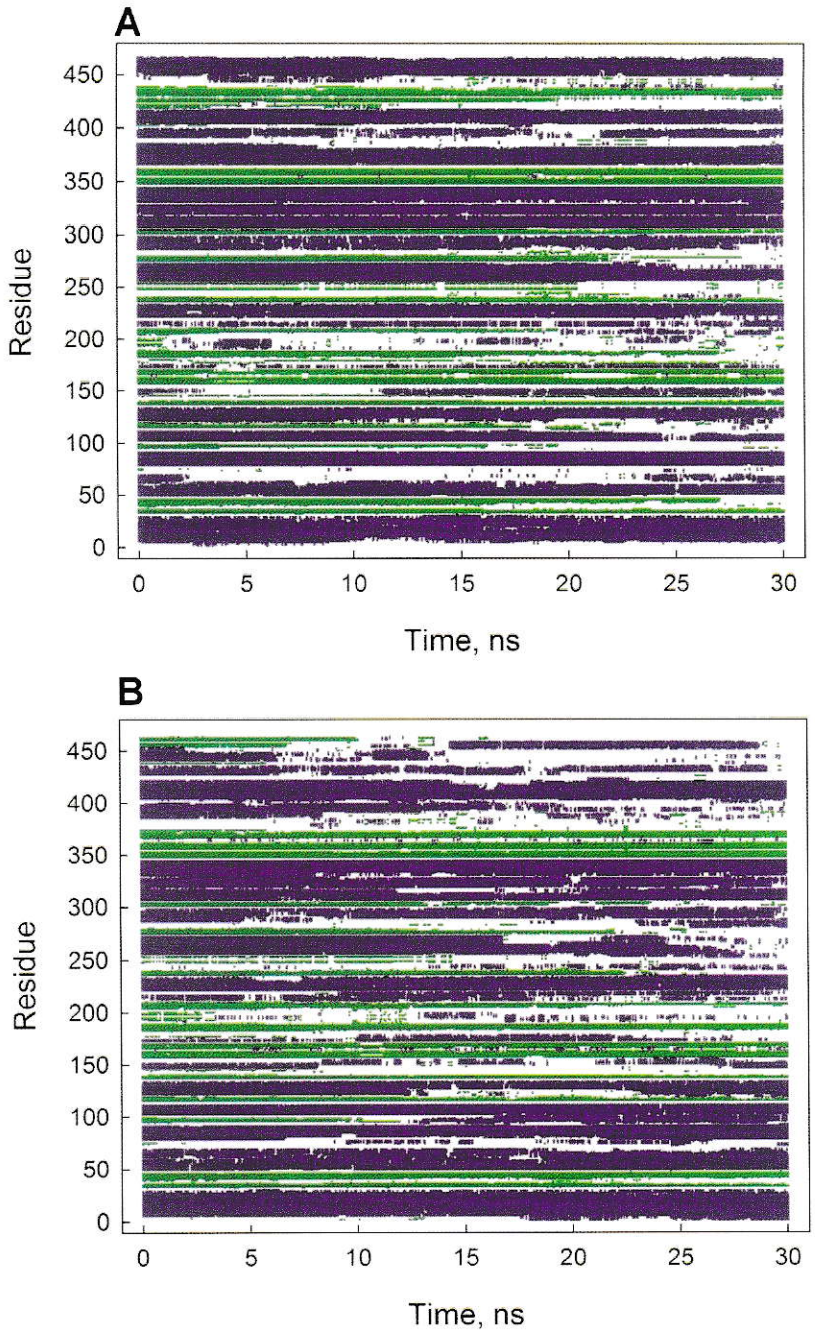


Figure 2.11: Secondary structure evolution in the unfolding molecular dynamics simulations. **A.** *Tl*GK. **B.** perGK. Helices are shown in violet and sheets are shown in green.

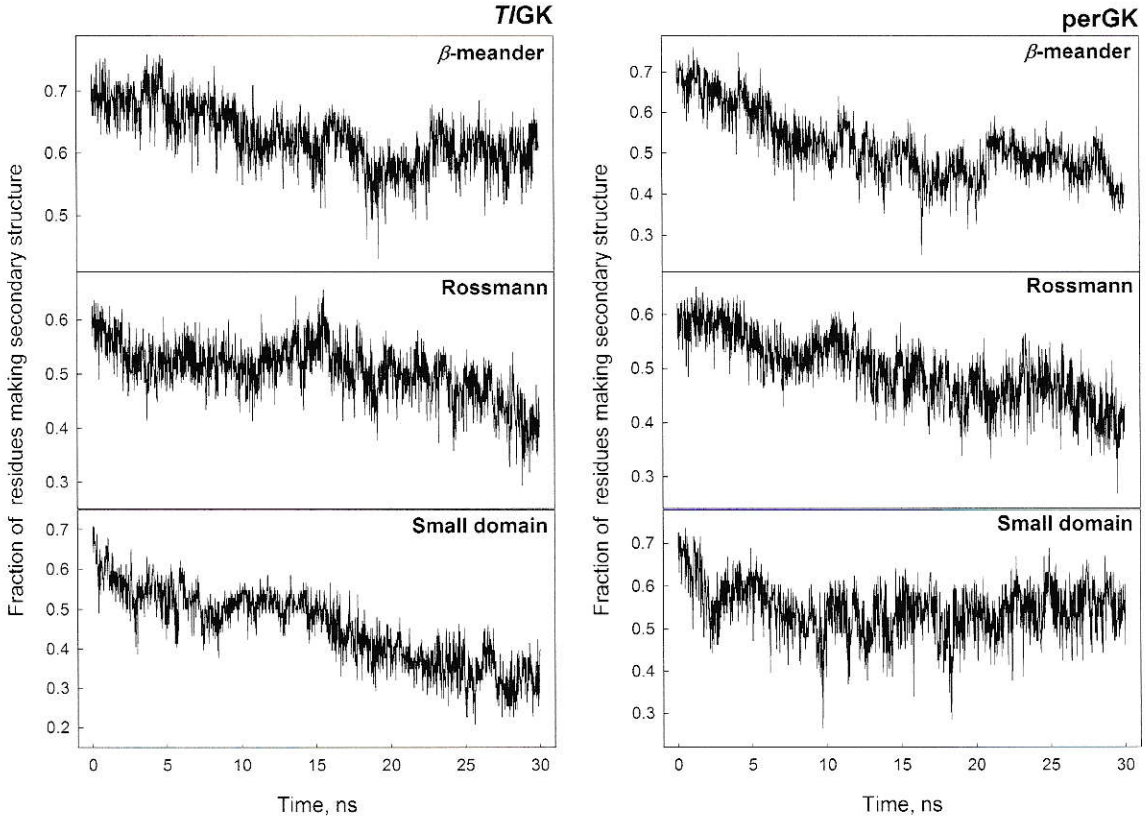


Figure 2.12: Sectorized secondary structure evolution in the unfolding molecular dynamics simulations. **A.** *Tl*GK. **B.** perGK.

simulation. Now, the β-meander and the rossmann part of the protein are those whose unfolding transitions are synchronized. In this case, both of them show a steady decrease of secondary structure through the 30 ns of the simulation.

The results suggest that the stability at the rossmann part of the proteins is strongly dependent on the other sectors of the protein fold. As in the *Tl*GK case small domain is the most labile, then upon its denaturation the rossmann part collapses. On the other hand, given that the circular permutation apparently results in a destabilization of the β-meander region, now the whole large domains unfolds as a unit.

However, it has to be considered that the degree of opening between the domains could be a factor for the unfolding pathway sampled. Indeed, as in this microscopic level the process is driven stochastically by the thermal fluctuations, more simulations are needed. This is being currently performed in our laboratory.

2.3.4 Nucleotide specificity of *Tl*GK and perGK

To test whether the C-terminal circular permutation is involved in the ATP/ADP discrimination in members of the ribokinase superfamily, we assayed the ability of *Tl*GK and perGK to phosphorylate glucose using either ADP or ATP.

Traditionally, coupled enzymatic assays are used to measure enzyme kinetics with these enzymes. For glucokinases the production of glucose-6-phosphate is coupled to the reduction of $NAD(P)^+$ by using a glucose-6-phosphate dehydrogenase (Koga *et al.*, 2000). For phosphofructokinases the production of fructose-1,6-bisP is coupled to the reduction of NADH by using a cocktail of aldolase, triose-phosphate isomerase, and glycerophosphate dehydrogenase (Currie *et al.*, 2009). In both cases, the assays recognize specifically the sugar-phosphate produced irrespective of the phosphoryl source. As it has already been discussed by us elsewhere (Guixé and Merino, 2009) this could produce serious artifacts since given the high concentrations of ATP used in most of the nucleotide specificity test published (Ronimus *et al.*, 1999; Tuininga *et al.*, 1999)

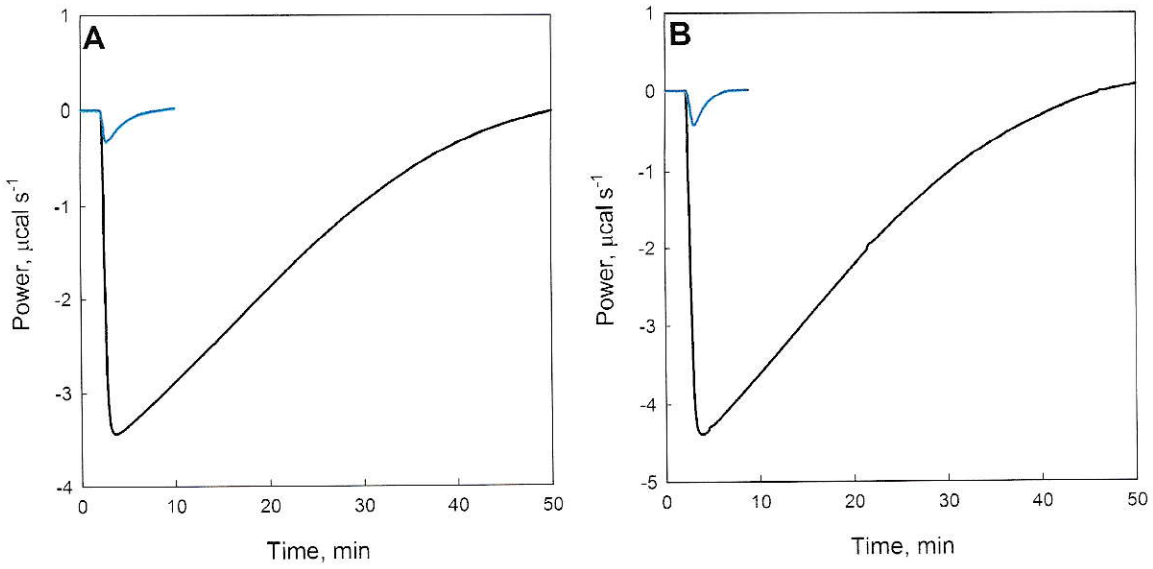


Figure 2.13: Nucleotide specificity of *Tl*GK and perGK. The black line corresponds to the kinetic assay performed with 250 μM ADP while the blue line corresponds to the assay using 250 μM ATP with either *Tl*GK (**A**) or perGK (**B**). All experiments were performed at 30 °C

even trace amount of ADP (from impurities or hydrolysis in the cuvette) in the order of 1% could produce velocities in the order of $0.5k_{cat}$.

The isothermal titration calorimeter is able to measure the heat produced (or absorbed) from an enzymatic reaction which gives a direct measurement of the catalytic process. Of course, this gets rid of the problem mentioned above. Figure 2.13 shows an experiment where a single injection of protein was performed to a cell with 250 μM of either ADP or ATP, 4 mM glucose, and 5.25 mM Mg^{2+} at 30 °C. A final concentration of 10 nM and 15 nM were used for *Tl*GK and perGK respectively. The apparent molar enthalpy measured for the ADP-dependent reaction is -11.4 ± 0.47 kcal/mol for *Tl*GK and -13.3 kcal/mol for perGK[5]. The cause of this difference is most likely related with experimental error. Nevertheless, both values are very similar.

In both cases, the peak observed for ATP is only ~3% of the total heat observed

[5]This value is the average of two measurements.

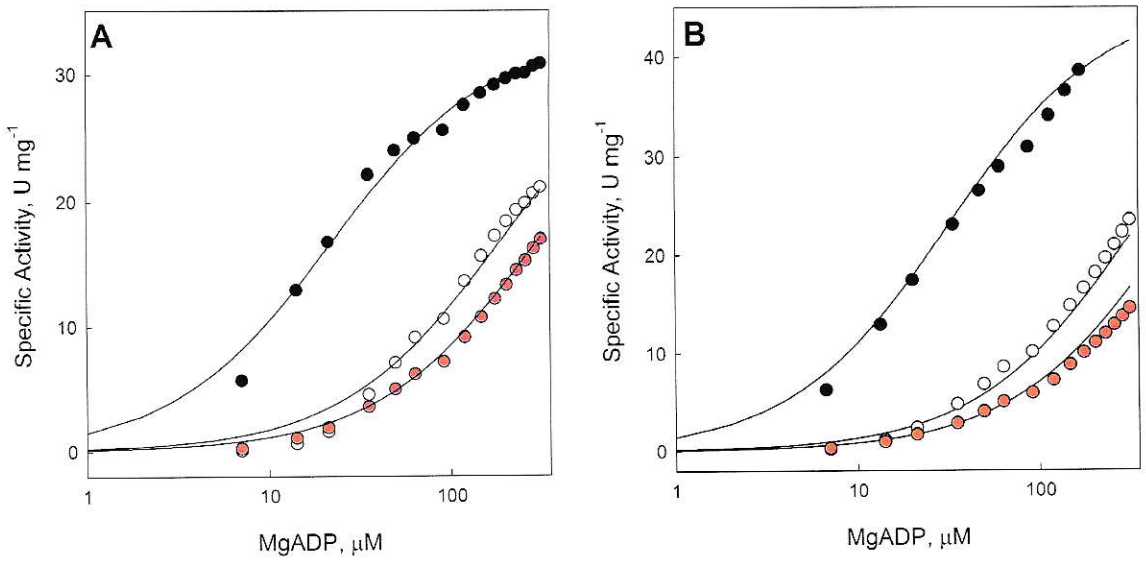


Figure 2.14: ADP/ATP competitive inhibition in *Tl*GK and perGK. Saturation curves with filled circles were performed without ATP, empty circles have 3 mM ATP and red circles have 5 mM ATP. **A.** *Tl*GK. **B.** perGK. All experiments were performed at 30 °C.

in the ADP case. Surprisingly, this shows that both enzymes are not able to catalyze the transfer of the γ-phosphate of ATP to glucose. Unfortunately, the kinetic constants for MgADP estimated directly from Figure 2.13 differ, for both enzymes, from those measured by the coupled assay, being K_M almost an order of magnitude higher. In fact, the K_M value measured is directly proportional to the initial MgADP concentration (not shown) which suggests the the products of the reaction are acting as strong competitive inhibitors, rendering this method useless for this particular enzyme.

In order to study the kinetic behavior of both enzymes the multiple injection method was used (for an explanation of ITC methods for enzyme kinetics see Appendix A). The enzymatic constants obtained by this method ato 30 °C were k_{cat} 29.7 ± 0.4, 41.2 ± 1.1 s^{-1} and K_M 21.5 ± 1.4, 31.4 ± 2.6 μM for *TlGK* and perGK respectively (Figure 2.14). A typical result of this kind of experiment can also be seen in Figure A.4 (Appendix A).

While ATP cannot be used as phosphoryl donor it could still behave as a competitive inhibitor. To study this, the enzyme saturation as a function of MgADP was assayed at different fixed concentrations of MgATP. The resulting curves where globally fitted to the a competitive inhibitory model of the form

$$v = \frac{k_{cat}\epsilon_0[S]}{K_M^{app} + [S]} \tag{2.2}$$

where

$$K_M^{app} = K_M\left(1 + \frac{[I]}{K_i}\right) \tag{2.3}$$

and ϵ_0 is the concentration of enzyme, [S] is the initial concentration of substrate, [I] is the concentration of inhibitor, K_M and k_{cat} have their usual meaning, and K_i represents the dissociation constant for the protein-inhibitor complex under the rapid-equilibrium approach.

Clearly, the model successfully describes the data. Then, the K_d value for the protein-MgATP complex can be calculated from the fit. The value obtained for *Tl*GK is $4.0 \cdot 10^{-4} \pm 2.5 \cdot 10^{-5}$ while the value obtained for perGK is $3.0 \cdot 10^{-4} \pm 1.8 \cdot 10^{-5}$. This demonstrates while the microscopic behavior of the nucleotide binding site is altered (Figure 2.9) neither the gross structural features (Figure 2.5) nor the catalytic properties (Figures 2.13 and 2.14) are modified by the C-terminal circular permutation. This is quite interesting since, although it is a very suggestive hypothesis, this topology change has absolutely no effect on the ability of the enzyme to discriminate between ADP and ATP.

Interestingly, using this K_i in equation 2.2 and 2.3 and also considering that from the experiments in Figure 2.13 it is possible to estimate a 3% contamination of ADP in the ATP solution, then an experiment using the glucose-6-phosphate dehydrogenase method with 2 mM ATP should produce an activity of $0.32k_{cat}$ which explains the ATP-

dependent activity observed for these enzymes in some articles (Guixé and Merino, 2009).

As it was already mentioned above, it is not possible to obtain the kinetic constants directly from a single injection experiment due to a dependence between the initial concentration of ADP and the parameters obtained. Interestingly, if this effect is indeed produced by the products of the reaction, then it is possible to obtain the K_d for the AMP-protein complex by analyzing the progress curves obtained from the ITC. To this end, we used the DynaFit 4 program (BioKin Ltd.) (Kuzmic, 1996) which numerically solves a system of differential equations describing a kinetic model provided by the user. As now we are solving a dynamic system, the quantities fitted are the velocity constants for every binding event plus the catalytic constant of the reaction itself, which, of course, increases approximately two fold the degrees of freedom. As the amount of data here does not allow us to fit all the constant together, and considering that we are essentially interested in the binding constants, we fixed the value of k_{cat} to that obtained by the multiple injection method (29.7 ± 0.4 s^{-1}) and, when needed, the value of the association steps (e.g. the kinetic constant for the process $E + S \rightharpoonup ES$) were set[6] to 10^8 M^{-1} s^{-1}.

We compared the ability of a simple Michaelian model to explain the data with a model containing competitive inhibition produced by the product. Given that preliminary data from our laboratory strongly suggest that these enzymes have an ordered sequential mechanism of substrate binding being ADP the first substrate to bind, we also tested a mixed-like model of inhibition where one of the products acts as a competitive inhibitor and the other as an uncompetitive one.

Figure 2.15 shows the fit of the Michaelian model and also the mixed-like model for comparison. Clearly from the residuals of the fit, the most simple model is unable to reproduce the data. Also, the calculated K_M value is 2.2 fold higher than the measured

[6]Given that, in principle, all this association steps should be limited by diffusion, then the choice of 10^8 M^{-1} s^{-1} seems appropriate

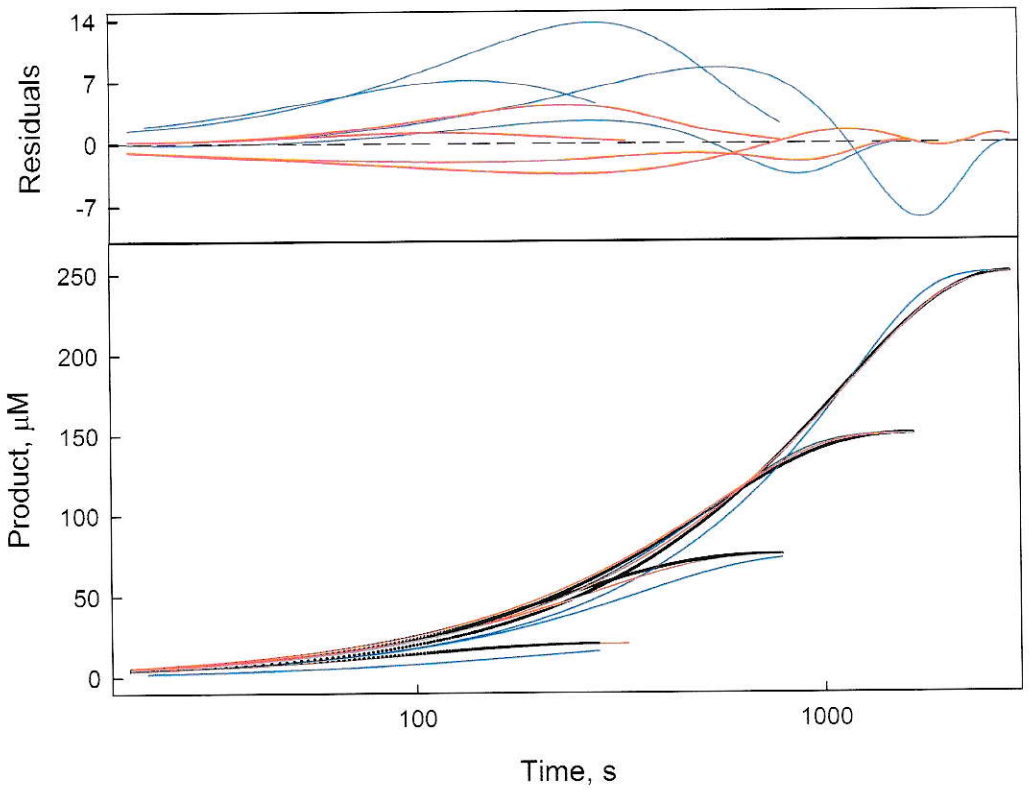


Figure 2.15: Comparison of two kinetic models for *Tl*GK using progress curves. Black curves are the experimental data obtained using initial MgADP concentrations of 20, 75, 150, or 250 μM. The blue curves are the results of the fit to the Michaelian model, while the red curve is the result of the fit to the mixed-like mechanism. All experiments were performed at 30 °C.

by the multiple injection method (Table 2.2). On the other hand, the K_M values calculated by any of the inhibitory models are much closer to that determined by the multiple injection method. Indeed, the mixed-like model is the best ranked by DynaFit.

If we suppose that AMP is the product causing the competitive inhibition, while glucose-6-phosphate is producing the uncompetitive inhibition, the the calculated K_i values for both products are very close to the K_M values of ADP and glucose respectively[7] (Koga *et al.*, 2000). One possible explanation for this is that the increased loss

[7]Initial velocity assays performed at 40 °C using coupled enzymes we have estimated K_M values of 0.42 and 0.02 mM for glucose and ADP respectively which is in good agreement with those reported elsewhere.

Table 2.2: Enzymatic constants obtained from DynaFit. K_i^1 is the inhibition constant for the enzyme-product complex. K_i^2 is the inhibition constant for the enzyme-substrate-product complex.

Model	**Enzymatic constants**		
	K_M	K_i^1	K_i^2
	μM		
Multiple injection method	21.5 ± 1.4		
Michaelian	46.0 ± 0.4		
Competitive	19.7 ± 0.3	$6.4 \cdot 10^{-5} \pm 1.5 \cdot 10^{-6}$	
Mixed-like	13.2 ± 0.2	$5.6 \cdot 10^{-5} \pm 1.2 \cdot 10^{-6}$	$2.7 \cdot 10^{-4} \pm 5.0 \cdot 10^{-6}$

of entropy upon binding due to the extra degrees of freedom of the phosphorylated molecule is compensated by a small gain in enthalpy change upon binding.

2.3.5 Effect of temperature on the activity and the binding properties of *Tl*GK

While most of the archaea that present the modified Embden-Meyerhof pathway are hyperthermophilic organisms the majority of the published studies on the ADP-dependent kinases have been performed at moderated temperatures. This is mainly due to the problems associated with high temperatures studies such as nucleotide hydrolysis, auxiliar enzyme instability, etc.

The measurement of nucleotide binding at high temperatures is one of the most difficult problems to solve since most of the techniques to measure equilibrium binding are usually very slow. On the other hand, one could try some less informative techniques such us measuring the change in fluorescence (being from an intrinsic or an extrinsic probe) with the ligand concentration. We used this approach to measure MgADP binding to *Tl*GK, but, unfortunately, neither the four tryptophan sidechains nor the addition of extrinsic fluorescent probes (such as pyridoxal phosphate or pyrene maleimide) are sensitive to the binding of any ligand.

On the other hand, ITC binding can provide a direct estimate of all the thermody-

namic quantities (ΔG, ΔH, and ΔS) plus the stoichiometry of the complex.

First, let us consider an enzyme with a single set of equivalent sites for a ligand. Then the association constant K is

$$K = \frac{\Theta}{(1-\Theta)[L]} \tag{2.4}$$

where Θ is the fraction of sites occupied by the ligand and $[L]$ is the free concentration of the ligand. Also the total concentration of ligand $[L]_T$ can be expressed as

$$[L]_T = [L] + n\Theta[M]_T \tag{2.5}$$

where n is the number of equivalent sites and $[M]_T$ is the total concentration of enzyme. Combining equations 2.4 and 2.5 we obtain

$$\Theta^2 - \Theta\left(1 + \frac{[L]_T}{n[M]_T} + \frac{1}{nK[M]_T}\right) + \frac{[L]_T}{n[M]_T} = 0 \tag{2.6}$$

On the other hand, the total heat content Q of the solution (determined relative to zero for the free species) at fractional saturation Θ is

$$Q = n\Theta[M]_T\Delta HV \tag{2.7}$$

where ΔH is the molar change of enthalpy upon association and V is the volume of the cell. By solving equation 2.6 for Θ and replacing it in equation 2.7 we obtain the relation

$$Q = \frac{n[M]_T\Delta HV}{2}\left[1 + \frac{[L]_T}{n[M]_T} + \frac{1}{nK[M]_T} - \sqrt{\left(1 + \frac{[L]_T}{n[M]_T} + \frac{1}{nK[M]_T}\right)^2 - \frac{4[L]_T}{n[M]_T}}\right] \tag{2.8}$$

which is used to fit the titration data obtained in the ITC[8]. As K and ΔH are directly

[8]In fact, the real value obtained from the titration experiment is the heat change between two

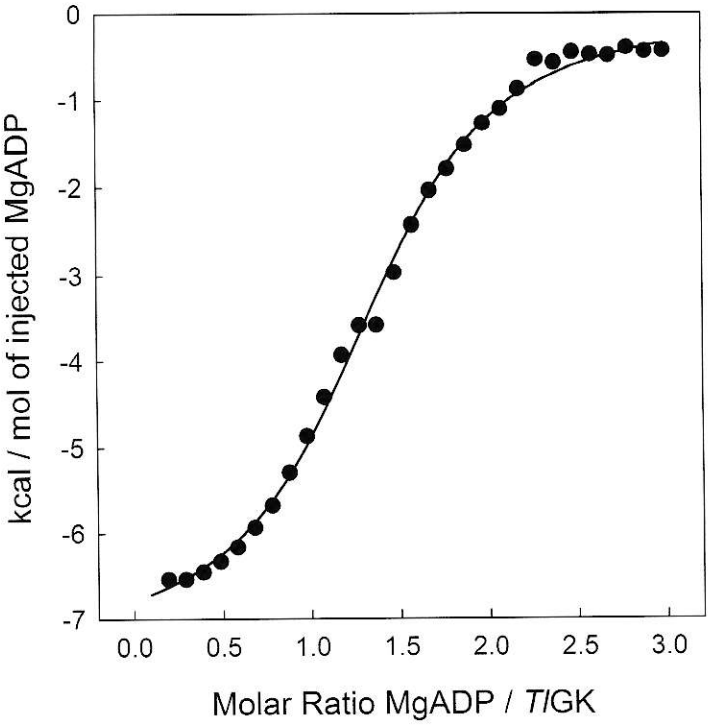


Figure 2.16: Binding of MgADP to *Tl*GK measured by ITC. The experiment was performed at 30 °C.

obtained from the experiment, also ΔG^0 and ΔS can be derived through the relation

$$\Delta G^0 = -RT \ln |K| = \Delta H - T\Delta S \tag{2.9}$$

where R is the gas constant and T is the temperature at which the experiment was performed.

Generally, the change in specific heat upon binding is very small, so the change in enthalpy and entropy are quite temperature independent. In this way, the free energy change can be measured at a small temperature and estimated at a higher temperature from one experiment.

Figure 2.16 shows the result of an ITC binding experiment performed with *Tl*GK

consecutive injections ΔQ. By using equation 2.8 for each pair of consecutive injections n, K, and ΔH can be obtained.

Table 2.3: Dependence of the kinetic properties of *Tl*GK with temperature. The K_M value is for ADP while the K_i value is for ATP.

Temperature	k_{cat}	K_M	K_i
°C	s^{-1}	μM	
30	29.7 ± 0.4	21.5 ± 1.4	$4.0 \cdot 10^{-4} \pm 2.5 \cdot 10^{-5}$
40	47.5 ± 0.7	26.5 ± 1.8	$8.2 \cdot 10^{-4} \pm 5.5 \cdot 10^{-5}$
50	52.1 ± 0.6	19.6 ± 1.1	$6.7 \cdot 10^{-4} \pm 4.2 \cdot 10^{-5}$
60	89.1 ± 2.0	23.9 ± 2.5	N.D.

and MgADP at 30 °C[9]. This gives a ΔG^0 -6.7 kcal/mol (equivalent to a K_d of 1.5 $\cdot 10^{-5}$), a ΔH of -7.4 kcal/mol, and a ΔS -2.26 cal/mol $\cdot$ K. Given the small entropy associated with the binding event, it is possible to estimate a K_d for the *Tl*GK-MgADP complex at the optimal growth temperature of *T. litoralis* (80 °C) of $8.4 \cdot 10^{-5}$.

To gain a better understanding of the effect of temperature on the ability of *Tl*GK to discriminate between ADP and ATP we performed the same experiment shown in Figure 2.14, but at two more temperatures. The results are summarized in Table 2.3.

As in the case of the dissociation constant for the ADP binding both K_M and K_i does not vary much within the temperature range assayed, somehow validating the experiments performed at lower temperatures.

Through Eyring's work (Eyring, 1935) it is possible to obtain the free energy of activation of a given reaction from the dependence of the velocity constant with temperature by using the relation

$$k = \left(\frac{k_B T}{h}\right) e^{\left(\frac{-\Delta G^{\ddagger}}{RT}\right)} \tag{2.10}$$

where k_B is the Boltzmann's constant, h is the Planck's constant, $\Delta G^{\ddagger}$ is the activation free energy, and R is the gas constant. Also considering that $\Delta G^{\ddagger} = \Delta H^{\ddagger} - T\Delta S^{\ddagger}$ (being $\Delta H^{\ddagger}$ the activation enthalpy and $\Delta S^{\ddagger}$ the activation entropy) then the equation above can be rewritten as

[9]All thermodynamic constants are given in the binding direction.

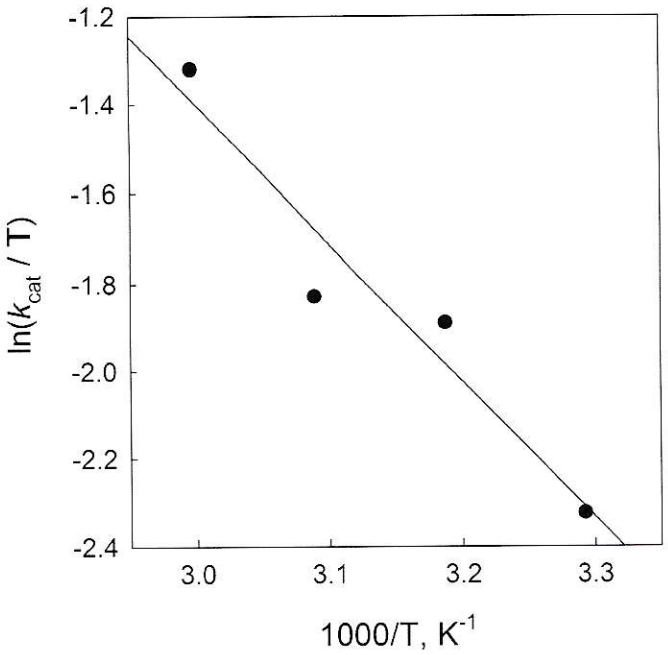


Figure 2.17: Eyring plot for the ADP-dependent glucokinase reaction catalyzed by *Tl*GK.

$$\ln\left|\frac{k}{T}\right| = \ln\left|\frac{k_B}{h}\right| + \frac{\Delta S^{\ddagger}}{R} - \frac{\Delta H^{\ddagger}}{RT} \tag{2.11}$$

Figure 2.17 shows a plot of the $\ln(k_{cat}/T)$ against T^{-1}. With these data and using equation 2.11 it is possible to estimate a $\Delta H^{\ddagger}$ of 6.15 kcal/mol and a $\Delta S^{\ddagger}$ -31.3 cal/mol · K for the transfer reaction. However, it is important to stress out that while the $\Delta H^{\ddagger}$ calculated is independent of the model used to describe the kinetics of the decomposition of the transition state the $\Delta S^{\ddagger}$ is not. Indeed, this value is estimated from the intersection of the $\ln(k_{cat}/T)$ against T^{-1} plot which is the sum of the term $\Delta S^{\ddagger}/R$ and $\ln(K_B/h)$. In this case, the later term comes from the Eyring's theory and it is related with the frequency of the vibrational mode responsible for converting the activated complex to the product and should vary depending on the theory used.

This kind of analysis has been performed on other ADP-dependent kinases before

(Hansen and Schönheit, 2004; Verhees *et al.*, 2002). In fact, a small kink in the Arrhenius like curve can be observed at 60 °C in the glucokinase from *P. furiosus* (Verhees *et al.*, 2002). However, such a behavior cannot be observed in the phosphofructokinase from *Archaeoglobus fulgidus* (Hansen and Schönheit, 2004) nor in our case. Unfortunately, the experiments mentioned above are not very informative since the plots are constructed using activity values measured with a fixed concentration of ligand, and therefore they do not represent the true maximum velocity of the enzyme. In fact, given that the plots are also constructed from specific activity and not k_{cat} it is not possible to obtain any thermodynamic quantity directly from them. As such, they should only be used as a qualitative reference.

2.4 Acknowledgements

We are grateful to Dr. Richard Garratt and all the staff of the Crystallography and Biophysics laboratories who kindly received us in the Instituto de Física de São Carlos (Universidade de São Paulo) where the calorimetric and circular dichroism measurements were done. This work was supported by Grant 1070111 from the Fondo Nacional de Desarrollo Científico y Tecnológico (Fondecyt) Chile, a travel award given by the Vicerrectoria de Asuntos Academicos (Universidad de Chile), and a travel award given by the AMSUD-PASTEUR network.

CHAPTER 3

Specificity evolution of the ADP-dependent sugar kinase family

This work was originally published in the article: Merino, F., and Guixé, V. (2008). Specificity evolution of the ADP-dependent sugar kinase family – *In silico* studies of the glucokinase / phosphofructokinase bifunctional enzyme from *Methanocaldococcus jannaschii*. *FEBS J.* **275**: 4033–4044.

Summary

In several archaea of the *Euryarchaeota* group the glycolytic flux proceeds through a modified version of the Embden-Meyerhof pathway, where the phosphofructokinase and glucokinase enzymes use ADP as the phosphoryl donor. These enzymes are homologous to each other. In the hyperthermophilic methanogenic archaeon *Methanocaldococcus jannaschii*, it has been possible to identify only one homolog for these enzymes, which shows both ADP-dependent glucokinase and phosphofructokinase activity. This enzyme has been proposed as an ancestral form in this family. In this work we studied the evolution of this protein family using the Bayesian method of phylogenetic inference and real value evolutionary trace in order to test the ancestral character of the bifunctional enzyme. Additionally, to search for specificity determinants of these two functions we have modeled the bifunctional protein and its interactions with both sugar substrates

using protein-ligand docking and restricted molecular dynamics. The results show that the evolutionary story of this family is complex. The root of the family is located inside the glucokinase group showing that the bifunctional enzyme is not an ancestral form, but could be a transitional form from glucokinase to phosphofructokinase due to its basal location within the phosphofructokinase group. The evolutionary trace and the molecular modeling experiments showed that the specificity for fructose-6-phosphate is mainly related with the stabilization of a negative charge in the phosphate group, whereas the specificity for glucose is related to the presence of some histidines instead of glutamines/asparagines and with the interaction of this ligand with a glutamic residue corresponding to E82 in the bifunctional enzyme.

3.1 Introduction

Several archaea of the *Euryarchaeota* group present a uniquely modified Embden-Meyerhof pathway which involves only four of the classical enzymes present in the canonical pathway (Verhees *et al.*, 2003). One of the most striking features of this modified glycolysis is the phosphorylation of glucose and fructose-6-phosphate by ADP as the phosphoryl donor, instead of ATP. These ADP-dependent kinases are homologous to each other and they show no sequence similarity to any of the hitherto known ATP-dependent enzymes. Then, it was proposed that they were part of a new family of kinases (Tuininga *et al.*, 1999). The presence of these ADP-dependent enzymes has been reported in several members of the *thermococcales* group (Tuininga *et al.*, 1999; Ronimus *et al.*, 1999; Koga *et al.*, 2000; Jeong *et al.*, 2003) and also, based on kinetic and genomic data, Verhees et al. proposed the operation of this pathway in some methanogenic archaea of the *methanococcales* and *methanosarcinales* groups. (Verhees *et al.*, 2001). Recently, Ronimus and Morgan (Ronimus and Morgan, 2004) reported a close homolog for these enzymes in higher eukaryotic genomes, which has significant ADP-dependent glucokinase activity.

To date, the structure of the ADP-dependent glucokinases (ADP-GK) from *Thermococcus litoralis* (Ito *et al.*, 2001), *Pyrococcus furiosus* (Ito *et al.*, 2003), and *Pyrococcus horikoshii* (Tsuge *et al.*, 2002) and the ADP-dependent phosphofructokinase (ADP-PFK) from *Pyrococcus horikoshii* (Currie *et al.*, 2009) have been solved. Surprisingly, despite the low sequence identity of these enzymes with the hitherto known ATP-dependent kinases, they can be classified, from a structural point of view, as members of the ribokinase superfamily (Ito *et al.*, 2001). The ribokinase-like fold is basically composed of an 8 stranded β–sheet surrounded by 8 α–helices, 3 in one side and 5 in the other. This family was first proposed by Bork *et al.* (1993) based just on sequence alignments, since the first structure for a member of this family was published five years later (Sigrell *et al.*, 1998). Members of this family comprise ATP-dependent

kinases of fructose-6-phosphate, fructose-1-phosphate, tagatose-6-phosphate, fructose, ribose, and nucleosides. Now, with more structural information available it has been possible to recognize that the ribokinase superfamily also contains enzymes that can transfer the *gamma*-phosphate of ATP to some vitamins involved in the B6 synthesis, like pyridoxal kinase (Li *et al.*, 2002) and ADP-dependent kinases. Then, this superfamily can be subdivided in three major groups: the ATP-dependent sugar kinases earlier described by Bork, the ATP-dependent vitamin kinases, and the ADP-dependent sugar kinases. The main structural difference between the vitamin kinase enzymes and the other two groups is that the former presents only the core $\alpha\beta\alpha$ ribokinase-like fold (large domain), whereas the enzymes belonging to the other groups have, in addition, a small domain composed always by a β-sheet and sometimes by some α-helical insertions. This domain acts as a lid in the active site and has been proposed as a good phylogenetic marker for the evolution of this superfamily (Zhang *et al.*, 2004). Although the structure of several members of this superfamily is available, there are few reports that address, from a structural perspective, the evolution of the group (for example, see (Zhang *et al.*, 2004)).

As a result of the biological importance of the identity of the phosphoryl donor, there has been some attention to the structural determinants for the ATP/ADP specificity (Ito *et al.*, 2001). On the other hand, to the best of our knowledge, lacking are studies addressing, from a structural point of view, the issue of the sugar specificity. Probably, this is mainly due to the absence of a structure of an ADP-PFK co-crystallized with fructose-6-phosphate available to date.

The hyperthermophilic methanogenic archeon *Methanocaldococcus jannaschii* has just one ADP-dependent enzyme, which has been kinetically characterized by two different groups (Verhees *et al.*, 2001; Sakuraba *et al.*, 2002). Interestingly, this enzyme has significant ADP-dependent kinase activity with glucose and fructose-6-phosphate. Inhibition studies showed that probably both sugars can bind to the same site in the

enzyme (Sakuraba *et al.*, 2002). Based on this feature it was proposed that this enzyme is an ancestral form that originated the two separated specificities through gene duplication. However, although there are several dendrograms published for the ADP-dependent family, there are not phylogenetic studies available.

In this work we analyze the evolution of the ADP-dependent family of kinases using the Bayesian method of phylogenetic inference and real value evolutionary trace (rvET) (Mihalek *et al.*, 2004) to test the ancestral character of the bifunctional enzyme and to search for specificity determinants of function. Also, we have modeled the ADP-dependent glucokinase/phosphofructokinase enzyme (ADP-GK/PFK) and its interaction with glucose and fructose-6-phosphate. To the best of our knowledge, this is the first time that a protein structural model bound to fructose-6-phosphate is available for a member of the ribokinase superfamily.

3.2 Experimental Procedures

3.2.1 Sequence and structural alignments, evolutionary trace, codon usage and phylogenetic analysis

The structure of the ADP-GK's from *Pyrococcus horikoshii*, *Pyrococcus furiosus*, and *Thermococcus litoralis* (PDB ID's 1L2L, 1UA4, and 1GC5 respectively) and the structure of the ADP-PFK from *Pyrococcus horikoshii* (PDB ID 1U2X) were structurally aligned using the CE-MC program (Guda *et al.*, 2001). Additionally each pdb was split off into small and large domains. A new structural alignment was performed for the separated domains using the DeepView program (Guex and Peitsch, 1997). Using these two structural alignments a final sequence alignment was constructed. All the archaeal sequences for ADP-dependent kinases available in the GeneBank non-redundant database were aligned in ClustalX (Thompson *et al.*, 1997) using the structural alignment previously constructed as a profile and finally, this alignment was corrected manually. Then the sequences of the eukaryotic ADP-GK's were aligned in clustalX to the former alignment. This final alignment was again corrected manually using the secondary structure prediction performed by the Jpred server[1] (Cuff *et al.*, 1998) for the Mus musculus ADP-GK as a guide.

To study possible specificity determinants within this family a real value evolutionary trace (Mihalek *et al.*, 2004) analysis was performed for the archaeal sequences including both specificities as well as for the two specificities separately.

A phylogenetic tree for the archaeal ADP-dependent family of kinases was constructed using the Bayesian method of phylogenetic inference implemented in MrBayes 3.1 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). The eukaryotic sequences were used as the out-group. To test the possibility of horizontal gene transfer in this family, the genes were grouped according to their relative synonymous

[1] available at `http://www.compbio.dundee.ac.uk/~www-jpred`

codon usage with the GCUA 1.2 software (McInerney, 1998). Additionally, the ADP-GK gene from *Methanothrix thermophila* was compared with all the archaeal genomes included in the phylogenetic analysis using the graphical codon usage analysis software (Fuhrmann *et al.*, 2004). Also, ADP-PFK's and ADP-GK's from the *methanosarcinales* group were compared to their own genomes using the same procedure.

3.2.2 Molecular modeling of the bifunctional enzyme

Since it has been shown that binding of the phosphoryl-acceptor ligand can change the angle between the two domains in members of the ribokinase superfmily, the bi-funcional enzyme was modeled in two different conformations. The open conformation was modeled using 1U2X as template. This enzyme has 40% sequence identity with the ADP-GK/PFK. There is a unique structure in the closed conformation (1UA4) which shares only 25% of sequence identity with the ADP-GK/PFK. Because this level of sequence similarity produces models that are not appropriated for the studies performed in this work, 1U2X and 1UA4 were used as templates as follows: 1U2X was split off into its large and small domains and then aligned with 1UA4. Later the closed enzyme and also the two fragments were used as templates for the closed conformation.

15 models were constructed for each conformation with MODELLER 8 (Sali and Blundell, 1993). The quality of the models was evaluated using PROCHECK (Laskowski *et al.*, 1993), Prosa2003 (Sippl, 1993), Verify3D (Lüthy *et al.*, 1992), and ProQ (Wallner and Elofsson, 2003).

3.2.3 Partial charges derivation and docking calculations

To test the results of the rvET analysis, α-D-glucose and β-D-fructose-6-phosphate were docked to the best model of the enzyme in each conformation. The atomic partial charges of each molecule were derived with the RED III program (Pigache *et al.*, 2004) using the rigid body reorientation algorithm and the RESP-A1 model for the charge

fitting. The quantum mechanics software used was PC-GAMESS 7.1[2]. The derivation for fructose-6-phosphate was done using the multi-molecule approach with β-D-fructose and methyl-phosphate as fragments.

The K_M measurements performed by Sakuraba *et al.* (2002) were done at pH 6.5 for fructose-6-phosphate and glucose. To take this effect into account, the protonation state of the ionizable residues was predicted at pH 6.5 in each conformation using the web server H++[3] (Gordon *et al.*, 2005). After this procedure partial charges were added with the CHARMM force field (MacKerell Jr *et al.*, 1998).

The docking calculations were performed with AutoDock 4 (Morris *et al.*, 1998) with the protonation states in accord with the pH's used in the kinetic experiments performed by Sakuraba's group (Sakuraba *et al.*, 2002). Fifty separated runs of the Lamarkian genetic algorithm using 2,500,000 energy evaluations and the other parameters as default were performed for both ligands in the open conformation. The same procedure was used for the closed conformation except that when the ligand was fructose-6-phosphate the number of energy evaluations was raised to 25,000,000.

All the side chains were rigid in the calculations with the open conformation and in the calculation with the closed conformation and glucose. After the modeling procedure the side chain of K170 point far from the binding site in the closed conformation probably due to charge constraints in the absence of the phosphate group of fructose-6-phosphate. For this reason, this side chain was free to move when the ligand was fructose-6-phosphate in this conformation.

In the two docking calculations that involve glucose, just the bond between the C5 and C6 was free to rotate, whereas when the ligand was fructose-6-phosphate, the bonds between C1 and C2, C5 and C6, C6 and O6, and O6 and the phosphate atom were free.

[2]Granovsky AA, `http://classic.chem.msu.su/gran/gamess/index.html`
[3]available at `http://biophysics.cs.vt.edu/H++`

3.2.4 Restricted molecular dynamics calculations

To test whether the interactions predicted by the docking calcuations for both sugars in the closed conformation were stable, restricted molecular dynamics simulations were performed. The system was solvated with the smallest possible sphere of water and all the simulations were performed with NAMD 2.6 using the spherical boundary condition and the SHAKE constraints (Phillips *et al.*, 2005). The systems were minimized with 1,500 iterations of the conjugated gradient algorithm. Later, the system was equilibrated at 300 K for 50 picoseconds and then the simulation was extended for 0.5 nanoseconds using a time step of 2 femtoseconds. All the residues with at least one atom inside a 7 Å radius of the ligand were free to move, while the others were fixed in the equilibration and simulation steps. All the atoms were free in the minimization procedure. Both simulations were performed using the CHARMM force field (MacKerell Jr *et al.*, 1998; Foloppe and MacKerell Jr, 2000; MacKerell Jr and Banavali, 2000). The parameters for glucose were available in CHARMM. For fructose-6-phosphate we used the parameters for the sugar phosphate available for the nucleic acids (Foloppe and MacKerell Jr, 2000; MacKerell Jr and Banavali, 2000). The same partial charges used in the docking calculations were used here.

All the structural figures as well as the trajectory analysis were performed using the VMD software (Humphrey *et al.*, 1996).

3.3 Results and Discussion

3.3.1 Evolutionary analysis of the ADP-dependent kinase family

To study the evolution of the archaeal part of the ADP-dependent family a phylogenetic tree was constructed using the Bayesian method of phylogenetic inference with the eukaryotic part as the out-group. The archaeal group presents ADP-GK's and ADP-PFK's, while the eukaryotic group present only ADP-GK's. For this reason, it seems reasonable to think that the divergence between the archaeal and eukaryotic groups occurred prior to the gene duplication event in the archaeal part. Based on this hypothesis the eukaryotic group was chosen as out-group.

The inferred tree was very robust to the improvements in the alignment used in terms of topology and posterior probability (not shown). Also, it was very similar to the dendrogram proposed by Ronimus and Morgan 2004 using the maximum parsimony and neighbor joining methods. It shows basically five groups: the ADP-GK's from eukaryotic sources (out-group sequences), the ADP-GK's from the *thermococcales* group, the ADP-GK's from the *methanosarcinales* group, the ADP-PFK's from the *thermococcales* and *methanococcales* groups, and the ADP-PFK's from the *methanosarcinales* group (Figure 3.1). Surprisingly, the root of the cladogram was not between the ADP-PFK's and the ADP-GK's. Instead, it was located between the ADP-GK's from *thermococcales* and *methanosarcinales*. This means that there was not a bifunctional origin as was proposed early by Sakuraba *et al.* (2002). The two ADP-PFK's groups appear from the same clade as the *thermococcales* ADP-GK's. Unfortunately, the only characterized gene from a methanogenic source is the ADP-GK/PFK from *M. jannaschii*. Based on the evolutionary trace and phylogenetic results (see below), it is possible that the ADP-PFK's from the *methanosarcinales* group can also use glucose as phosphoryl acceptor. This could be important especially for the operation of the Embden-Meyerhof

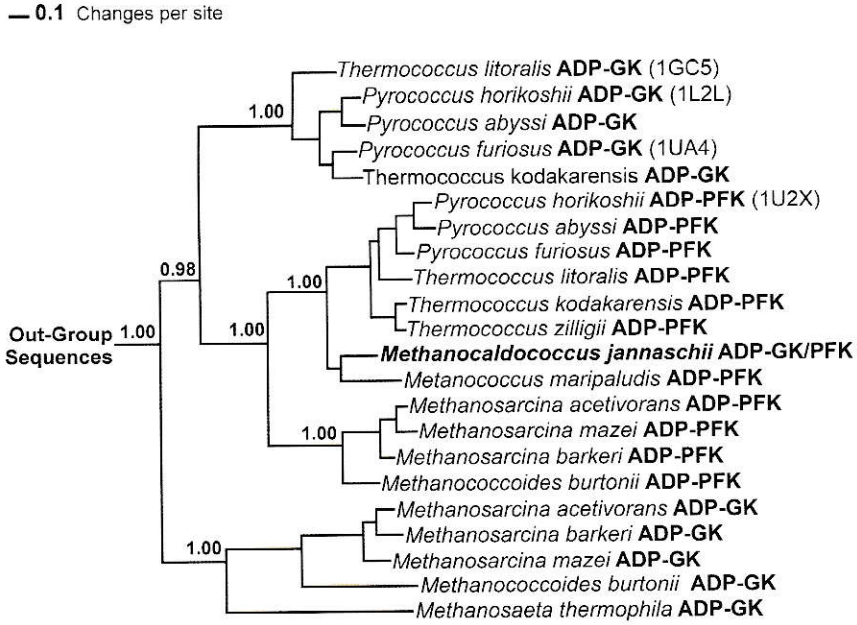


Figure 3.1: Phylogenetic tree of the ADP-dependent sugar kinase family. The eukaryotic group was used as the outgroup. The posterior probability of each split is shown in the nodes. The distance to the outgroup sequences does not represent the distance in the real tree, because they are only used in the figure to point out the place where the root of the archaeal part is located. The position of the bifunctional enzyme is highlighted by bold letters.

pathway in *M. burtonii* since the ADP-GK gene in this organism has a big C-terminal deletion and probably this protein is not functional. Interestingly, the rest of the sequence shows the same conservation pattern of the other glucokinases, suggesting that either the event is recent or that this protein is still performing its function in some way.

To test the robustness of the root's place, a new rooting was performed using other members of the ribokinase superfamily as the out-group. Again, the root of the group appears in the same place with a high posterior probability (not shown).

To explain the topology of the obtained phylogenetic tree more than a gene duplication event is necessary because the ADP-GK's from the *thermococcales* group appear to be more similar to the ADP-PFK's than to other ADP-GK's. One possible expla-

nation could be that a gene duplication event occurred after the divergence between the *thermococcales* and *methanosarcinales* groups and later, an event of lateral gene transfer, added the ADP-PFK activity to the methanogenic group. A similar scenario for the generation of paralogous proteins involving lateral gene transfer has been proposed by Gogarten *et al.* (2002). To test this hypothesis, the archaeal genes included in the former analysis were grouped according to their relative synonymous codon usage (McInerney, 1998). The clustering of the genes shows basically two groups (Figure 3.2). One of them is formed by the genes from methanogenic archaea and the other is formed by the genes from the hyperthermophilic archaea. This shows that the codon usage of these genes is in good agreement with the phylogeny of the archaeal group and does not support the horizontal gene transfer proposal. However, this methodology fails when the lateral movement has occurred deep in the evolution because codon usage is masked through time due to the accumulation of several mutations. This means that it is possible that the event we are searching for is too ancient to be captured by this methodology. However, it has been shown that horizontal gene transfer is continuously modifying the prokaryotic genomes and that the rate of transfer of housekeeping genes is significant (Jain *et al.*, 1999). The only exception in the clustering seen in figure 3.2 is the ADP-GK from *Methanothrix thermophila* which appears inside the hyperthermophilic archaeal group. Then, the codon usage of this gene was compared to the codon usage of the whole genomes of the archaea used in the phylogenetic inference in terms of relative adaptiveness (Fuhrmann *et al.*, 2004). This analysis showed that, in fact, the codon usage of this ADP-GK is more similar to the codon usage of the *Thermococcus* order than to its own genome (Figure 3.2, inset). *Methanothrix thermophila* is a thermophilic archeon that can grow at temperatures between 35 and 75 °C while the archaea from the *thermococcales* group can grow between 65 and 100 °C. This little temperature overlap suggests that *M. thermophila* can live in the same habitat as some *thermococcales* (at least in terms of growth temperature) giving a chance for lateral gene

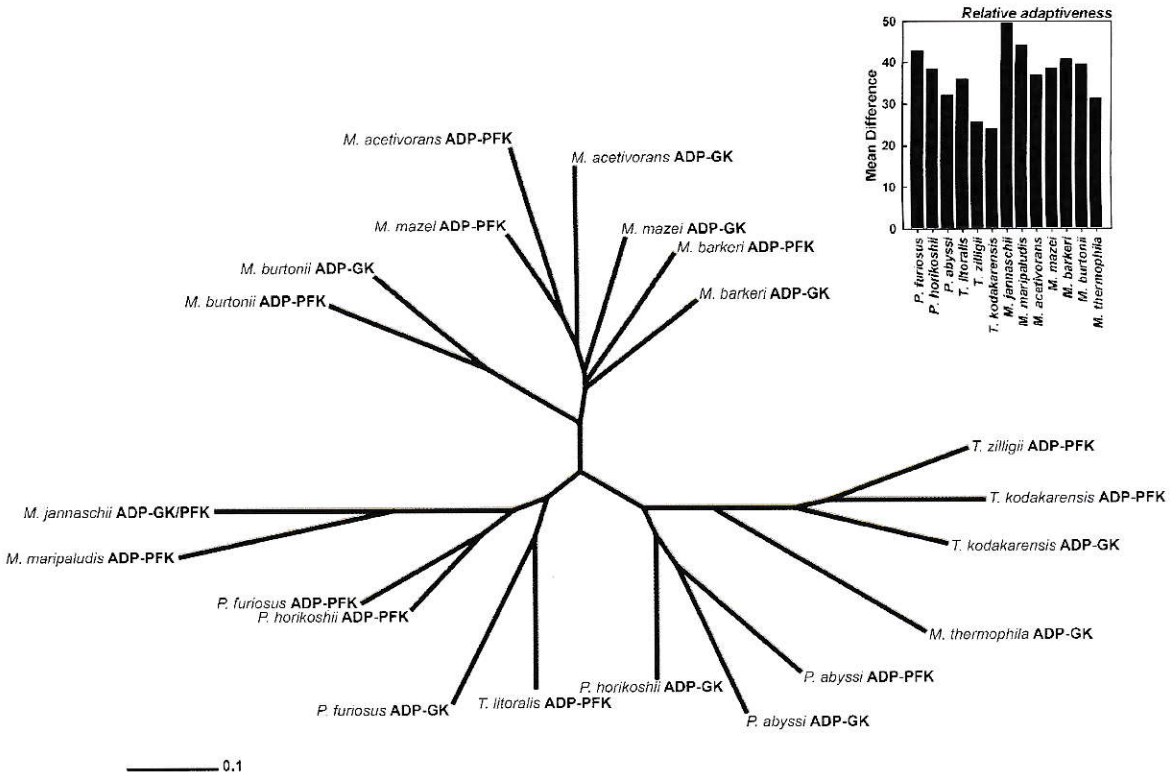


Figure 3.2: Codon usage comparison for the ADP-dependent genes from archaeal sources. Dendrogram grouping the genes according to their average difference in relative synonymous codon usage. Inset: comparison between the codon usage of the ADP-GK from *M. thermophila* and the genomes of the archaea included in the relative synonymous codon usage analysis in terms of relative adaptiveness.

transfer. Although it does not explain the topology of the phylogenetic tree it gives evidence to confirm that lateral gene transfer occurred within this family. However, when the same procedure is used to compare the ADP-GK's and ADP-PFK's from *methanosarcinales* with their genomes it is not possible to see any significant difference (not shown).

3.3.2 Molecular modeling of the bifunctional enzyme

To study in detail the structural determinants of the sugar specificity within the ADP-dependent kinase family, the structure of the ADP-GK/PFK enzyme from *Methanocal-*

dococcus jannaschii was modeled using homology modeling. The active site of these enzymes is located in a cleft between the two domains. It has been previously reported that binding of the phosphoryl acceptor can change the relative orientation between the large and small domains in several kinases of the ribokinase superfamily (Tsuge *et al.*, 2002; Sigrell *et al.*, 1998; Schumacher *et al.*, 2000). Specifically, the two domains approach each other when the ligand is bound changing the enzyme from an open to a closed conformation. For this reason, the bifunctional enzyme was modeled in two different conformations.

The protein quality scores for the 15 models constructed for the open and closed conformations are shown in Table 3.1 and 3.2 respectively. In general, the models show above 92% of their residues within the core residues of the Ramachandran space and a very low percent of them within the disallowed regions. The prosa2003 combined z-score (Sippl, 1993) is very close to that expected for a protein of 462 residues long. Also, ProQ (Wallner and Elofsson, 2003) shows that all the models are classified as good models and the verify3D scores (Lüthy *et al.*, 1992) of the chosen models are always above 0.1.

The models for the open and closed conformations are available in the Protein Model DataBase[4] with the accession id PM0075152 and PM0075151 respectively.

To study in more detail the conformational change induced by the phosphoryl acceptor, models of the two conformations were compared using DynDom (Hayward and Lee, 2002). The angle of closure between the domains is about 47 degrees. The analysis showed that 13 (V27, D28, A29, E102, E103, R104, K170, I171, N172, R173, A201, S202, and R203) of the 462 residues in the protein act as a hinge whereas the others belong to one of the two domains (Figure 3.3). These residues belong to the coils between the domains and to a little part of the β–sheet of the small domain. Of these 13 residues just V27, D28, A29, and R104 are well conserved among the ADP-dependent

[4] `http://mi.caspur.it/PMDB/`

Table 3.1: Quality measurements of the protein models in the open conformation. The model 12 was used for further analysis

	Prosa 2003	**Verify 3D**[a]	**PROCHECK**				**ProQ**	
	Z score		*Core (%)*	*Allowed (%)*	*Generally allowed (%)*	*Disallowed (%)*	*LGScore*	*Maxsub*
1	-12.46	186.28	92.5	6.6	0.7	0.2	5.823	0.529
2	-12.84	197.9	92.7	5.4	1.4	0.5	5.704	0.532
3	-12.86	182.62	93.7	4.9	0.9	0.5	5.583	0.537
4	-12.69	194.67	92.5	5.9	1.4	0.2	5.811	0.554
5	-12.66	189.18	93	5.9	0.9	0.2	5.888	0.526
6	-12.57	191.92	93	5.4	1.2	0.5	5.812	0.536
7	-12.45	188.62	91.3	7	0.9	0.7	5.427	0.517
8	-12.63	192.12	92.3	6.3	0.9	0.5	5.715	0.529
9	-12.57	198.45	92.7	5.6	0.9	0.7	5.629	0.539
10	-11.92	191.31	92.3	6.1	0.7	0.9	5.641	0.541
11	-12.41	190.09	92.5	6.3	0.7	0.5	5.688	0.511
12	-12.85	195.09	92.3	6.6	0.9	0.2	5.716	0.534
13	-12.8	186.64	92	6.3	1.2	0.5	5.749	0.537
14	-12.73	185.17	93	5.4	0.9	0.7	5.607	0.520
15	-12.86	181.64	93	5.4	0.9	0.7	5.942	0.549
1U2X	-13.99	213.9	93.1	6.9	0	0	6.956	0.642
1UA4	-14.32	220.67	93.3	6.2	0.5	0	7.562	0.776

[a]The value shown correspond to the sum over all residues in the protein

Table 3.2: Quality measurements of the protein models in closed conformation. The model 15 was used for further analysis

	Prosa 2003	**Verify 3D**[a]	**PROCHECK**				**ProQ**	
	Z score		*Core (%)*	*Allowed (%)*	*Generally allowed (%)*	*Disallowed (%)*	*LGScore*	*Maxsub*
1	-12.43	194.14	91.6	7	0.7	0.7	6.301	0.567
2	-12.6	166.52	91.6	7.5	0.5	0.5	5.895	0.526
3	-12.44	179.26	91.1	7.7	0.9	0.2	6.417	0.577
4	-12.52	193.44	91.6	7	0.7	0.7	6.195	0.573
5	-12.25	195.31	92	6.8	0.2	0.9	6.209	0.593
6	-12.1	185.9	90.9	7.5	0.5	1.2	6.315	0.595
7	-12.34	186.97	92	6.1	0.9	0.9	6.119	0.580
8	-12.26	191.88	90.9	6.8	1.6	0.7	6.401	0.582
9	-12.37	183	90.6	7.5	1.2	0.7	6.175	0.581
10	-12.26	176.17	91.1	7.5	0.7	0.7	6.276	0.580
11	-12.54	191.37	91.3	7	0.7	0.9	6.240	0.581
12	-12.38	196.33	92	7.3	0.2	0.5	6.445	0.601
13	-12.56	180.13	92	6.8	0.7	0.5	6.697	0.592
14	-12.22	187.56	90.4	8	0.9	0.7	6.390	0.587
15	-12.59	201.53	92	6.8	0.7	0.5	6.724	0.644
1U2X	-13.99	213.9	93.1	6.9	0	0	6.956	0.642
1UA4	-14.32	220.67	93.3	6.2	0.5	0	7.562	0.776

[a]The value shown correspond to the sum over all residues in the protein

kinase family. Of these, D28 is the only one that seems to be involved in sugar binding for both specificities (see below). K170 and R203 are well conserved within the ADP-PFK family and seem to be related with ligand specificity (see below). N172 is an interesting residue; it also seems to be related with sugar discrimination because it forms part of a conserved asparagine/histidine position in ADP-PFK/ADP-GK respectively. E102 and E103 are located in a region where there is at least a glutamic residue and a hydrophobic residue or another glutamic in all the thermophilic enzymes. In the mesophilic enzymes always are two hydrophobic residues. It has been shown that for a thermophilic/mesophilic adenylate kinase pair, the rate of opening of the nucleotide-binding lids is the limiting step for the observed k_{cat} (Wolf-Watz *et al.*, 2004). As a result of the rigidity of the thermophilic enzyme it was not able to do catalysis at low temperatures. A similar mechanism could be operating in the ADP-dependent family and those glutamic residues could be related to the rigidity of the hinge.

When the results of the real value evolutionary trace analysis (Mihalek *et al.*, 2004) are mapped in the structure of the enzymes two obvious patterns emerge (Figure 3.3). First, it is clear that the ADP-PFK group has more conserved residues than the ADP-GK group. This seems reasonable since the former group is evolutionary newer. Second, it can be seen a group of much conserved residues in a cleft between the two domains that forms the active site of these enzymes. It is interesting that when the enzymes are in the closed conformation, the residues form a very dense cluster which is not the case for the open conformation. From this point of view it seems reasonable to think that the closed conformation of these enzymes is necessary for the proper orientation of the ligands to produce catalysis as observed for other members of the superfamily (Ito *et al.*, 2003; Sigrell *et al.*, 1998; Schumacher *et al.*, 2000).

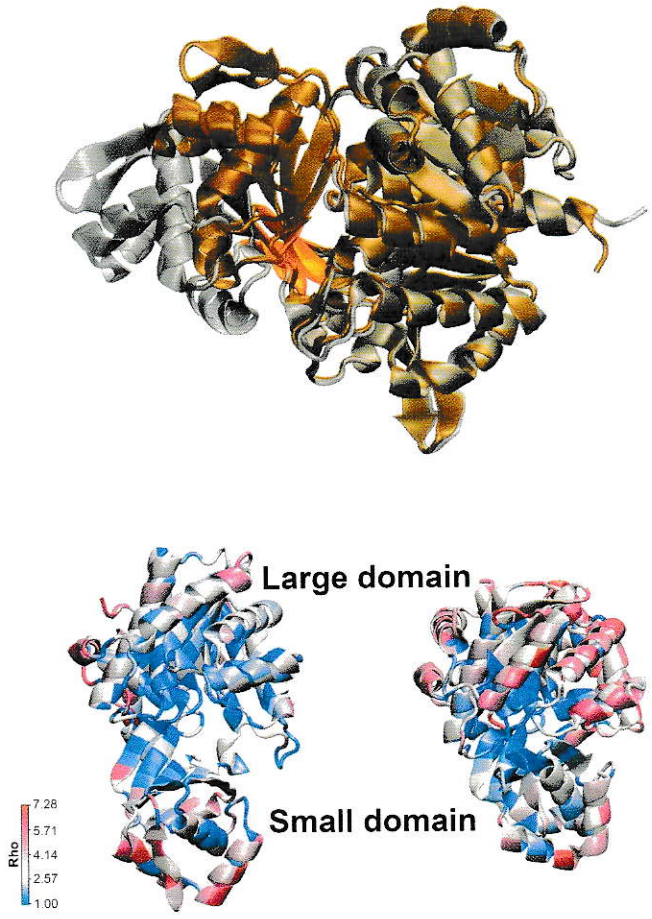


Figure 3.3: Comparison between closed and open conformations of the bifunctional enzyme. Upper: The residues acting as a hinge between both domains are shown in orange. Lower: crystal structure of the ADP-PFK from *P. horikoshii* (left) and the ADP-GK from *P. furiosus* (right). Each residue is colored according to its rho value from the rvET analysis for the corresponding specificity.

3.3.3 Protein-ligand interaction modeling

The ADP-GK/PFK from *M. jannaschii* shows significant activity with glucose and fructose-6-phosphate *in vitro*. However, the relevance of this enzyme in the phosphorylation of these two sugars in vivo depends on the intracellular concentration of the metabolites as well as the specificity of the enzyme for them. It has been previously shown that this enzyme has a specificity constant (k_{cat}/K_M) of $7.5 \cdot 10^5$ s^{-1} M^{-1} for fructose-6-phosphate and $1.2 \cdot 10^4$ s^{-1} M^{-1} for glucose (Sakuraba *et al.*, 2002). For two competing substrates as these, it means that in an equimolar mixture of both sugars the rate at which fructose-6-phosphate is phosphorylated will be approximately 60 times higher than the rate at which glucose does at any substrate concentration (Cornish-Bowden, 1995). This shows that, in fact, the enzyme from *M. jannaschii* is an unspecific phosphofructokinase as can be also deduced by its position in the ADP-PFK branch of the phylogenetic tree (Figure 3.1). Beyond this fact, the high KM value seen for glucose could be a consequence of a high intracellular concentration of this metabolite. If this is true the ADP-GK/PFK could be performing both functions *in vivo*. It has been previously suggested by Verhees *et al.* (2001) that the modified Embeden-Meryerhof pathway could be operative in this archeon based on the presence of the ADP-GK/PFK enzyme (characterized in that publication just as ADP-PFK) and other characteristic enzymes of the pathway such as GAPOR. The occurrence of the dual function *in vivo* then would be good agreement with that hypothesis. Interestingly, it has been shown that this enzyme can use acetyl phosphate as phosphoryl donor with a k_{cat} value similar to the obtained with ADP, but with a higher K_M value (Verhees *et al.*, 2001), demonstrating that there is promiscuity in both binding sites.

Another member of the ribokinase superfamily from *Methanocaldococcus jannaschii* which is very similar to the ATP-dependent phosphofructokinases group has been characterized (Hansen *et al.*, 2007). This enzyme was reported as a nucleoside kinase and, although it has a broad substrate specificity, it cannot use fructose-6-phosphate as

phosphoryl acceptor.

In order to obtain docking results comparable with the kinetic parameters measured for the bifuncional enzyme (Sakuraba *et al.*, 2002) the protonation state of the ionizable residues in the protein was predicted in the H++ server (Gordon *et al.*, 2005) at the pH value used in the kinetic characterization. It has been observed for the members of the ADP dependent kinase family that the optimum pH value is lower for the phosphofructokinase activity than for the glucokinase activity (Tuininga *et al.*, 1999; Ronimus *et al.*, 1999; Koga *et al.*, 2000). To address this point, the protonation state of the ionizable residues in the bifunctional enzyme was also predicted using a pH value of 7.8 that is close the pH optimum for the glucokinase activity. There are several residues with significantly shifted pKa values as could be predicted for a protein with several ionizable residues. Among them, four lysine residues change their protonation state in the pH range studied (6.5 or 7.8). K9, K170, K178, and K238 were protonated at pH 6.5 while they were uncharged at pH 7.8. Of these four residues, K170 is the most interesting related to the sugar specificity problem. In the rvET experiments, this residue appears to be important for ligand discrimination and also, in *E. coli* Pfk-2, a member of the ribokinase superfamily, a structurally equivalent residue is crucial for fructose-6-phosphate binding (Cabrera *et al.*, 2010). This observation could explain the lower pH optimum for the phosphofructokinase than the glucokinase activity in the family.

Beyond the different way that ligands bind to the active site, one would expect that the catalytic mechanism will be conserved among members of a protein superfamily. In the ribokinase superfamily, it has been shown that an aspartic residue acts as the catalytic base in the phosphoryl transfer reaction and that site-directed mutagenesis of this residue causes a dramatic lost of activity. (Ito *et al.*, 2003; Maj *et al.*, 2000; Campobasso *et al.*, 2000). Additionally, a hydrogen bond between this residue and the OH phosphoryl acceptor in the ligand has always been seen by X-ray crystallography

in members of the superfamily (Ito *et al.*, 2003; Sigrell *et al.*, 1998; Schumacher *et al.*, 2000; Ohshima *et al.*, 2004; Safo *et al.*, 2004). In the bifunctional enzyme this aspartic corresponds to the D442 residue. In order to consider a docking conformation "correct" there must be a contact between this residue and the phosphoryl acceptor OH group in the sugar.

The docking calculations described below showed that, as was mentioned before, the open conformation of the enzyme was not appropriated for catalysis. When the docking calculation was performed with glucose as the ligand, 11 clusters of solutions were encountered using a 0.5 Å clustering cut-off. The most populated one contains 18 conformations with dissociation constants estimated at 298 K around 25 mM. D442 makes a hydrogen bond with the C2 hydroxyl group while the C6 hydroxyl group is making two hydrogen bonds with the N26 and D28 side chains. The backbone of G107 also makes a hydrogen bond with the endocyclic oxygen. This suggests that the conformation is not appropriated for catalysis since the hydroxyl acceptor is far away from the catalytic base.

When the ligand was fructose-6-phosphate there were 37 clusters using the same cut-off as above showing that in fact it was not possible to find a low energy binding site. However, all the lower energy clusters show substrate conformations that are bound to the protein mainly through ionic interactions with the phosphate group of the sugar and that are far away from the catalytic D442.

When glucose was docked to the closed conformation of the bifunctional enzyme, just 3 clusters were encountered. The first of them showed 48 conformations and they are very similar to the conformation seen for glucose in the ADP-GK from *P. furiosus* (Ito *et al.*, 2003). The lower estimated free energy of binding in this cluster was -2.51 kcal/mol which corresponds to a K_d of 14.36 mM at 298 K, that is about one order of magnitude higher than the observed K_M of 1.6 mM.

In the case of the docking between the closed conformation and fructose-6-phosphate

20 clusters were encountered. The first cluster showed just 3 conformations with a very broad range of estimated free energy of binding. The second cluster contained 9 conformations. The ring of the sugar in this cluster binds in a very similar position to that of glucose. Additionally, the phosphoryl acceptor OH group makes a hydrogen bond with the lateral chain of D442. For these reasons this cluster was used for further analysis. The lower estimated free energy of binding in this cluster was -5.79 kcal/mol, which corresponds to a K_d of 56 μM at 298 K, that is about 5 times higher than the observed K_M of 10 μM. Although there are differences between the predicted and observed binding free energies, the higher affinity of the enzyme for fructose-6-phosphate than for glucose is consistent with the kinetic data. Interestingly, when the same calculations are performed with the partial charges of the ligands derived using the Gasteiger method (Gasteiger and Marsili, 1980) the results are structurally very similar, but the predicted dissociation constants are 1.20 mM and 2.9 μM, for glucose and fructose-6-phosphate respectively, which are in well agreement with the experimental data. However, the "QM-derived" charges are theoretically more appropriated.

Searching for contacts between the sugars and the protein a 5 Å radius was inspected. Glucose is making contact with N24, N26, D28, E82, G107, Q108, I111, N172, I174, R197, I199, Q235, V438, G439, and D442 (Figure 3.4). Fructose-6-phosphate has the same contacts with the exception of E82 and the inclusion of residues R203 and K170. The last residue is beyond the 5 Å cut-off, but it appears to make a weak ionic pair with the phosphate group of the ligand (Figure 3.4).

Among the several contacts between the proteins and the sugars, only 5 residues seem to be important for the sugar specificity based on the rvET analysis (Figure 3.4 and 3.5). This is a very striking result based on the differences between a pyranose and a furanose-phosphate and it suggests a high plasticity of the sugar binding site within this protein family.

The E82 side chain makes a hydrogen bond with the OH group of C2 in glucose,

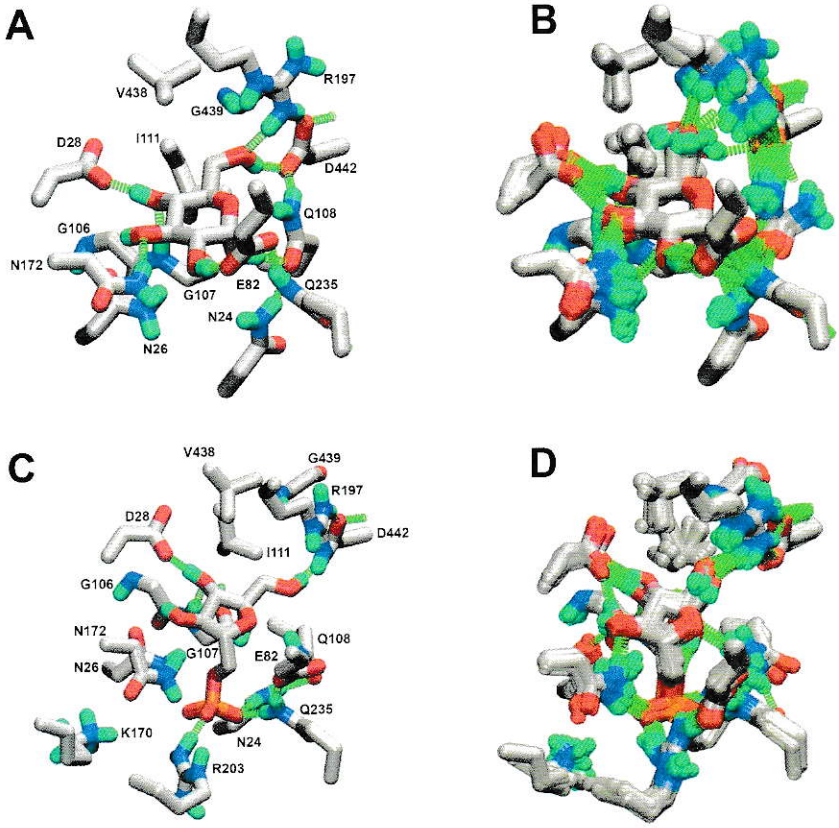


Figure 3.4: Protein-ligand interaction modeling results. **A.** α-D-glucose docked in the closed conformation of the bifunctional enzyme after the energy minimization procedure. **B.** Set of representative frames of the α-D-glucose molecular dynamics simulation. **C.** β-D-Fructose 6-phosphate docked in the closed conformation of the bifunctional enzyme after the energy minimization procedure. **D.** Set of representative frames of the β-D-fructose 6-phosphate molecular dynamics simulation.

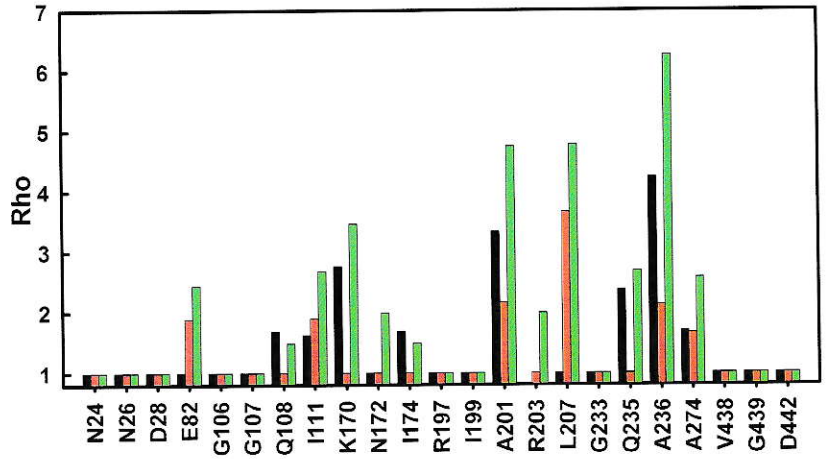


Figure 3.5: Results of the rvET analysis for all the residues within 5 Å from the ligands. The results for the glucokinase specificity are shown in black, those for the phosphofructokinase specificity in red, and those for the whole family in green. Lys170 is beyond the 5 Å cut-off, but it seems to be important for phosphate binding. Residues 201, 207, 236 and 274 are not in contact with the ligand, but they are included in the 5 Å radius. Rho values for Val438, Gly439 and Asp442 were calculated excluding the ADP-GK from *Methanococcoides burtonii*, as the gene encoding this has a C-terminal deletion and the proteinis probably not functional.

but makes no contact with fructose-6-phosphate. The rvET rho value for this position in the alignment was 1 for the glucokinase group, 1.89 for the phosphofructokinase group, and 2.44 for the whole family. This residue has been proposed to be important for the discrimination of glucose over fructose-6-phosphate (Ito *et al.*, 2003; Sakuraba *et al.*, 2002). Interestingly, this residue is present in all the methanogenic ADP-PFK's suggesting the possibility of these enzymes to bind glucose. N172 and Q235 interact with the phosphate group when the ligand is fructose-6-phosphate. When the ligand is glucose, N172 interacts with the C3 hydroxyl group whereas Q235 makes a hydrogen bond with the side chain of E82. The equivalent residue for Q235 in the ADP-GK from *P. furiosus* is making contact with the C1 hydroxyl group of glucose through a water molecule (Ito *et al.*, 2003). This interaction is not seen in our simulation due to the absence of water. N172 and Q235 are strictly conserved as asparagine and glutamine in the phosphofructokinase group. On the other hand, the N172 position is occupied by a highly conserved histidine residue while the Q235 position is occupied by glutamine or histidine residues in the glucokinase group. The explanation for this preference is not clear.

K170 and R203 interact with the phosphate group of fructose-6-phosphate through a weak ionic pair and a strong salt bridge respectively, as judged by their distance. Within the phosphofructokinase group these residues are strictly conserved while in the glucokinase group the K170 position has a rvET rho value of 2.76 and the R203 position is a gap. In the glucokinase group, the two positions that follow R203 in the alignment are negative charged residues that may be repelling the negative charge of the phosphate group in fructose-6-phosphate and thus avoiding the binding of this ligand.

To test the stability of the contacts seen in the docking experiments and in order to refine the interaction with ligands, especially for fructose-6-phosphate, short restricted molecular dynamics simulations were performed with the final docking results. Figure 3.4 shows a representative set of frames for each simulation. The interactions were

conserved in the 0.5 ns of simulation showing that the predicted contacts are stable. However, the hydroxyl acceptor in both ligands turned towards the sugar ring in the simulations far from the catalytic base, to interact with R197 and G439. This interaction occurred through the whole simulation for fructose-6-phosphate and was intermittent with the interaction with D442 in the case of glucose. R197 and G439 have been seen to interact with glucose through a pair of water molecules in the crystal structure of *P. furiosus* (Ito *et al.*, 2003). The absence of these water molecules within the active site in the molecular dynamics simulations could explain the observed behavior. Although the position of the internal water for glucose could be predicted by copying the ones seen in the crystal structure of the ADP-GK from *P. furiosus*, there is no way to predict the internal waters for fructose-6-phosphate. Since in the closed conformation the active site is very deep in the protein structure, it would require an extremely long dynamic simulation to see the waters from the solvatation sphere to place correctly within the sugar binding site.

In conclusion, we have shown that the root of the ADP-dependent kinase family is within the glucokinase group which discards the hypothesis of the ancestral character of the bifunctional enzyme. The results also suggest at least one event of horizontal gene transfer in this family. Finally, it seems that the principal determinant for the fructose-6-phosphate specificity is the presence of residues with the capability to stabilize the negative charge of the phosphate group, whilst the presence of the E82 side chain and the preference of histidine over asparagine/glutamine are key elements for glucose specificity in the active site.

3.4 Acknowledgements

We acknowledge Dr. Ricardo Cabrera for critically reading of the manuscript. This work was supported by Grant 1070111 from the Fondo Nacional de Desarrollo Científico y Tecnológico (Fondecyt) Chile.

CHAPTER 4

Biophysical characterization of the ADP-dependent 6-Phosphofructokinase from *Pyrococcus horikoshii* OT3

This work was originally published in the article[1]: Currie, M.A., Merino, F., Skarina, T., Wong, A.H.Y., Singer, A., Brown, G., Savchenko, A. Caniuguir, A., Guixé, V., Yakunin, A.F., and Jia, Z. (2009). ADP-dependent 6-phosphofructokinase from *Pyrococcus horikoshii* OT3: Structure determination and biochemical characterization of PH1645. *J. Biol. Chem.*, **284**: 22664–22671.

Summary

Some hyperthermophilic archaea use a modified glycolytic pathway that employs an ADP-dependent glucokinase (ADP- GK) and an ADP-dependent phosphofructokinase (ADP-PFK) or, in the case of Methanococcus jannaschii, a bifunctional ADP-dependent glucophosphofructokinase (ADP-GK/PFK). The crystal structures of three ADP-GKs have been determined. However, there is no structural information available for ADP-PFKs or the ADP-GK/PFK. Here, we present the first crystal structure of an ADP-PFK from *Pyrococcus horikoshii* OT3 (*Ph*PFK) in both apo- and AMP-bound forms

[1]The on-line version of this article (available at `http://www.jbc.org`) contains supplemental data, Table S1, and Figs. S1 and S2.

determined to 2.0-Å and 1.9-Å resolution, respectively, along with biochemical characterization of the enzyme. The overall structure of *Ph*PFK maintains a similar large and small α/β domain structure seen in the ADP-GK structures. A large conformational change accompanies binding of phosphoryl donor, acceptor, or both, in all members of the ribokinase superfamily characterized thus far, which is believed to be critical to enzyme function. Surprisingly, no such conformational change was observed in the AMP-bound *Ph*PFK structure compared with the apo structure. Through comprehensive site-directed mutagenesis of the substrate binding pocket we identified residues that were critical for both substrate recognition and the phosphotransfer reaction. The catalytic residues and many of the substrate binding residues are conserved between *Ph*PFK and ADP-GKs; however, four key residues differ in the sugar-binding pocket, which we have shown determine the sugar-binding specificity. Using these results we were able to engineer a mutant *Ph*PFK that mimics the ADP-GK/PFK and is able to phosphorylate both fructose 6-phosphate and glucose.

4.1 Introduction

Glycolysis is a central and highly conserved metabolic pathway in all three domains of life. However, the Embden-Meyerhof glycolytic pathway of some hyperthermophilic archaea displays distinct differences from the classical pathway. Glyceraldehyde-3-phosphate is converted to 3-phospho (3P)-glycerate by glyceraldehyde-3-phosphate ferredoxin oxidoreductase in a single step instead of using glyceraldehyde-3-phosphate dehydrogenase and phospho-glycerate kinase in the classical two-step reaction mechanism (Mukund and Adams, 1995; van der Oost *et al.*, 1998). Also, the classical ATP-dependent glucokinases (ATP-GKs) and phosphofructokinases (ATP-PFKs) are replaced with novel ADP-dependent glucokinases and ADP-dependent phosphofructokinases or, in the case of *Methanococcus jannaschii*, a bifunctional ADP-dependent gluco-/phosphofructokinase (ADP-GK/PFK) (Kengen *et al.*, 1994, 1995; Koga *et al.*, 2000; Sakuraba *et al.*, 2002; Tuininga *et al.*, 1999).

PFKs convert fructose 6-phosphate to fructose 1,6-bisphosphate. This is an early step in the Embden-Meyerhof pathway and therefore represents a critical control point for the entire pathway. Sequence and structural characterization classify PFKs into two convergent protein families (Bateman *et al.*, 2000): the PfkA family and the ribokinase superfamily, which includes the PfkB family and ADP-GK/ADP-PFK family. The PfkA family includes both allosterically regulated ATP-dependent enzymes found in a variety of eukarya and bacteria and non-allosterically regulated PPi-dependent enzymes found in all three domains of life (Hansen *et al.*, 2002; Meyer *et al.*, 1996). The PfkB family is a diverse family of ATP-dependent carbohydrate and pyrimidine kinases that is also present in all three domains of life. In general, ATP-PFKs from the PfkB family can be differentiated from PfkA ATP-PFKs by their lack of allosteric regulation, although the minor ATP-PFK from *Escherichia coli*, Pfk-2, is allosterically regulated by MgATP (Guixé *et al.*, 1998).

Several crystal structures of the PfkA family PFKs have been reported. The best

studied are the tetrameric PFKs from *E. coli* and *Bacillus stearothermophilus* (Evans *et al.*, 1981; Schirmer and Evans, 1990). Both contain subunits that consist of a large and a small 3-layered $\alpha\beta\alpha$ sandwich domain (Evans *et al.*, 1981; Schirmer and Evans, 1990). The large domain binds to ATP and the small domain binds to fructose-6-phosphate (Evans *et al.*, 1981; Schirmer and Evans, 1990). Recently the crystal structure of Pfk-2, a member of the PfkB family, was solved (Cabrera *et al.*, 2008). It is composed of two domains, a large 3-layered $\alpha\beta\alpha$ sandwich domain, similar to that seen in the PfkA family protein structure mentioned above, and a smaller four-stranded β-sheet domain (Cabrera *et al.*, 2008).

To date ADP-PFKs have been reported as non-regulated enzymes with Michaelis-Menten kinetics. This part of glycolysis is particularly interesting because with the presence of PFKs and fructose-1,6-bisphosphatase there is the possibility to produce futile cycling resulting in a net hydrolysis of nucleotide (ADP in the case of *thermococcales*) which, of course is undesirable because it would uncouple the metabolism. *E. coli* has overcome this problem by using a strong MgATP-induced inhibition when fructose-6-phosphate is low in both of its PFKs: Pfk-1 from the PfkA family and Pfk-2 from the ribokinase superfamily (Torres *et al.*, 1997; Zheng and Kemp, 1992). On the other hand, the archaea with the modified Embden-Meyerhof pathway seems to control this issue at the transcriptional level as, depending on the growth medium, either the production of PFK or fructose-1,6-bisphosphatase is turned off (Schut *et al.*, 2003). In this way, regulation of PFK is not needed.

The crystal structures of the ADP-GK from *Pyrococcus horikoshii* (*Ph*GK), the ADP-GK from *Thermococcus litoralis* (*Tl*GK) bound to ADP, and the ADP-GK from *Pyrococcus furiosus* (*Pf*GK) in complex with AMP and glucose have been reported, although there are no crystal structures of ADP-PFK or the bifunctional ADP-GK/PFK (Ito *et al.*, 2001; Tsuge *et al.*, 2002; Ito *et al.*, 2003). Despite low sequence identity between ADP-GKs and other kinases (e.g. *Ph*GK shares 11 and 12% identity with *E.*

coli ribokinase and human adenosine kinase, respectively) the fold is similar to that of ribokinases (Ito *et al.*, 2001). As a result, ADP-GKs have been classified as members of the ribokinase superfamily (Ito *et al.*, 2001). Overall, ADP-GK structures are composed of a large and a small α/β domain (Ito *et al.*, 2001; Tsuge *et al.*, 2002; Ito *et al.*, 2003). The ADP binding pocket is found on the surface of the large domain (Ito *et al.*, 2001; Tsuge *et al.*, 2002; Ito *et al.*, 2003). Residues responsible for glucose binding are found on the surface of both the large and small domains adjacent to the ADP-binding pocket (Ito *et al.*, 2001; Tsuge *et al.*, 2002; Ito *et al.*, 2003). Upon glucose binding, the protein undergoes a conformational change whereby the large and small domains close in creating the active-site pocket (Ito *et al.*, 2003). Interestingly, the unique ADP binding pocket of ADP-GKs recognizes the α- and β-phosphate of the ADP in an almost identical manner to how the β- and γ-phosphate of ATP are recognized in ATP-dependent kinases of the ribokinase superfamily (Ito *et al.*, 2001).

Here, we report the first structure of an ADP-PFK, alone and in complex with AMP, together with the biochemical characterization of the kinetic properties and substrate specificities. Moreover, we carried out site-directed mutagenesis in the sugar-binding pocket and active site of *Ph*PFK and identified residues that are critical for PFK activity and for distinguishing between glucose and fructose-6-phosphate.

4.2 Experimental procedures

4.2.1 Cloning, expression, and protein purification

The open reading frame for *Ph*PFK (PH1645) was amplified by PCR and inserted between the NdeI and BamHI restriction sites of a modified pET-15b expression vector (Novagen) (vector p11) as previously described (Zhang *et al.*, 2001). This construct generated an N-terminal hexahistidine tag joined to the *Ph*PFK protein by the TEV protease recognition site (ENLYFQ↓G). Recombinant wild type and mutant (see below) native and selenomethionine-labeled *Ph*PFK proteins were expressed in BL21(DE3) and DL41(DE3) *E. coli* cells (Novagen), respectively. Protein was purified using nickel-nitrilotriacetic acid affinity and size exclusion chromatographies. For more details see supplemental materials[2].

4.2.2 Site-directed mutagenesis

Site-directed mutagenesis was performed using QuikChange^{TM} site-directed mutagenesis kit (Stratagene) according to the manufacturer's protocol using the *Ph*PFK expression vector described above as the template. See Table 4.1 for the list of mutations made.

4.2.3 Protein crystallization

Protein crystals of apo-*Ph*PFK were generated through hanging drop vapor diffusion at 21 °C by mixing 2 μL of protein solution (10 mg/ml) with 2 μL of well solution consisting of 22% PEG 4000, 0.1 M Tris-HCl, pH 8.5, and 0.2 M $LiSO_4$. Crystals of the *Ph*PFK complex with AMP were obtained by crystallization of the *Ph*PFK D17A protein in the presence of 20% PEG 3350, 0.2 M lithium citrate, 10 mM fructose 6-phosphate, and 5 mM ADP. Prior to data collection, crystals were transferred into

[2]Available at `http://www.jbc.org/cgi/content/full/M109.012401/DC1`

Paratone-N and cryo-cooled in a nitrogen gas steam.

4.2.4 X-ray diffraction, structure determination, and refinement

Apo-*Ph*PFK crystals were placed in a cryoprotectant composed of 15% glycerol added to the crystallization solution and then flash frozen in liquid nitrogen prior to data collection. Multiwavelength anomalous dispersion data were collected at three wavelengths 0.96396, 0.97918, and 0.97943 Å at the Advanced Photon Source (APS, Argonne, IL) beamline 19-ID of the Structural Biology Center-CAT with a SBC-3 CCD detector. The data were processed using HKL2000 (Otwinowski and Minor, 1997). Data collection and processing statistics are shown in Table 4.2.

The structure of *Ph*PFK was solved using the multiwavelength anomalous dispersion method. Selenium sites were located using SOLVE (Terwilliger, 2000; Terwilliger and Berendzen, 1999). Six of the 10 expected selenium sites per asymmetric unit were found. Selenium position refinement, phase calculation, and density modification was performed by SHARP (de La Fortelle and Bricogne, 1997). The structural model was built and refined by XFIT, CNS, and Refmac (Brünger *et al.*, 1998; McRee, 1999). The AMP-bound structure was solved by molecular replacement using Phaser and the model was validated in Coot (Emsley and Cowtan, 2004; McCoy *et al.*, 2007). The final refinement statistics can be found in Table 4.2.

4.2.5 Determination of enzyme activity

PFK activity was assayed spectrophotometrically at 50 °C, by coupling the fructose-1,6-bisphosphate formation to the oxidation of NADH as previously described (Kengen *et al.*, 1994). See supplemental materials for more details.

4.2.6 Substrate specificity

The phosphoryl group donor specificity of *Ph*PFK was determined by measuring the PFK activity as described above substituting ADP for the following phosphoryl group donors, UDP, IDP, GDP, and CDP. The activity is reported as percentage of the maximum activity obtained with ADP. The divalent cation preference was tested similarly, replacing $MgCl_2$ in the assay by $MgSO_4$, $MnCl_2$, $NiSO_4$, $CaCl_2$, or $ZnCl_2$. 2.5, 4, and 7 mM total metal were used, keeping other components as those used in the standard assay. The activity observed using these metals was reported as units/mg.

4.2.7 Kinetic parameters

Kinetic parameters were determined at 50 °C by varying the concentrations of fructose-6-phosphate (0 to 2 mM) with saturating ADP concentrations, and vice versa (Kengen *et al.*, 1994). For some mutants, the range of fructose-6-phosphate concentrations was increased. In all these experiments, $MgCl_2$ was used as the divalent cation. The data were analyzed using SigmaPlot software (Systat Software, Inc., CA), and fitted to the Michaelis-Menten equation.

4.3 Results and discussion

4.3.1 Enzymatic activity and substrate specificity of *Ph*PFK

*Ph*PFK is predicted to be an ADP-PFK based on sequence similarity (Figure 4.1). In addition, the *Ph*PFK structures align well with the available ADP-GK structures (Figures 4.2C and 4.2D). Therefore, to test whether this enzyme is a true ADP-PFK the phosphorylation of F6P was assayed in the presence of several nucleotides. PFK activity was the highest in the presence of ADP or UDP as the phosphoryl group donor (Figure 4.3A). However, *Ph*PFK is also capable of transferring phosphoryl groups from IDP, GDP, and CDP to F6P but at a significantly slower rate (Figure 4.3A). A saturation curve was generated for both ADP and UDP. The calculated k_{cat}/K_M value for UDP was 40 times higher than that of ADP, indicating a strong preference for ADP over UDP as the phosphoryl group donor.

Wild type *Ph*PFK displayed Michaelis-Menten kinetics at 50°C. The following constants were determined using direct fit: K_M values of 15.2 ± 2.5 and 13.2 ± 2.4 μM for F6P and ADP, respectively, and a k_{cat} value of 45.5 ± 4.0 s^{-1}. All kinetic parameters are summarized in Table 4.1 and Figure 4.3B. The temperature of these reactions is much lower than the optimal growth temperature of *P. horikoshii* and since temperature has a significant effect on enzyme kinetics, the constants calculated here may differ from the values at the optimal growth temperature. Unfortunately, the commercially available auxiliar enzymes for this reaction are from mesophilic sources hindering the use of high temperatures in the assay. On the other hand, due to the large amount of mutants characterized in this article the study would be prohibitively longer using a discontinuous assay to measure the kinetic parameters at higher temperatures. Nevertheless, most of the work published for other enzymes from this family uses the same strategy, making the data here directly comparable. The calculated K_M values for F6P and ADP for wild type *Ph*PFK are significantly lower than those of the ADP-PFK

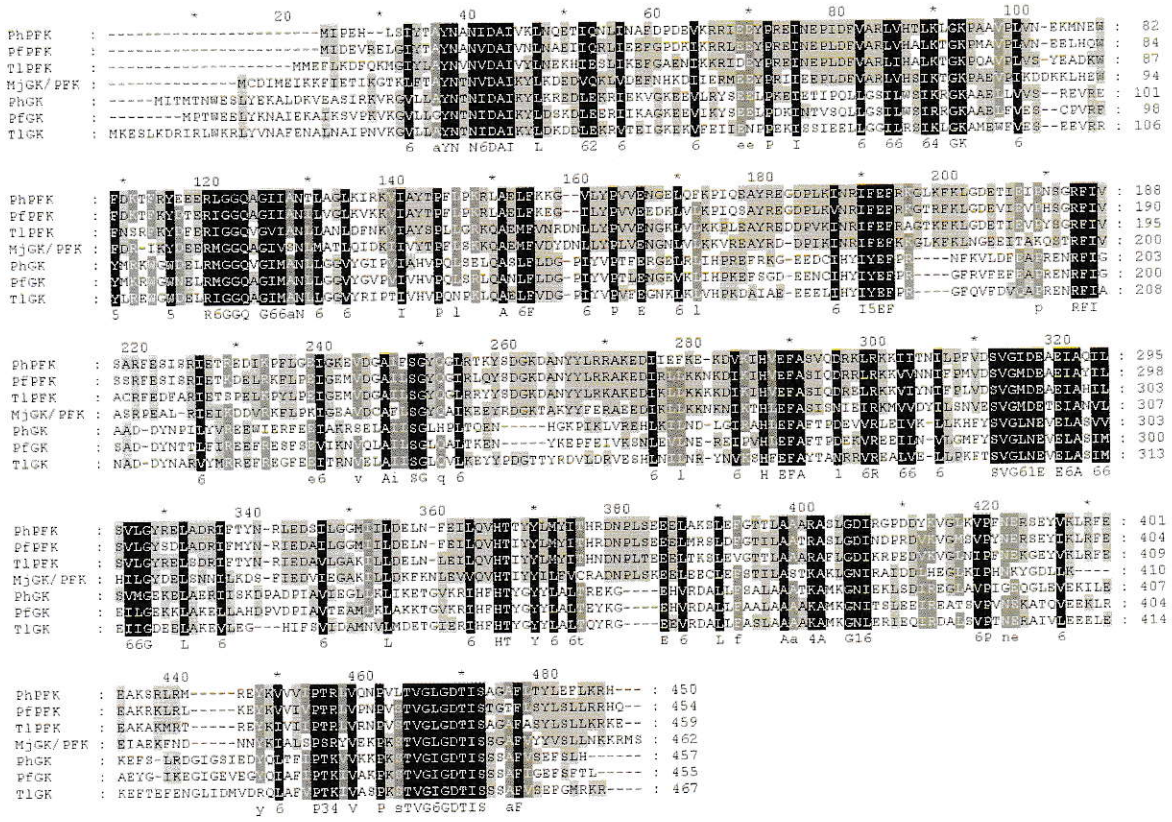


Figure 4.1: Multiple Sequence Alignment of *Ph*PFK with other ADP-PFKs, ADP-GKs, and the bifunctional ADP-GK/PFK. (Pf, *Pyrococcus furiosus*; Ph, *Pyrococcus horikoshii*; Tl, *Thermococcus litoralis*; and Mj, *Methanocaldococcus jannaschii*). This figure was prepared using GeneDoc (Nicholas *et al.*, 1997).

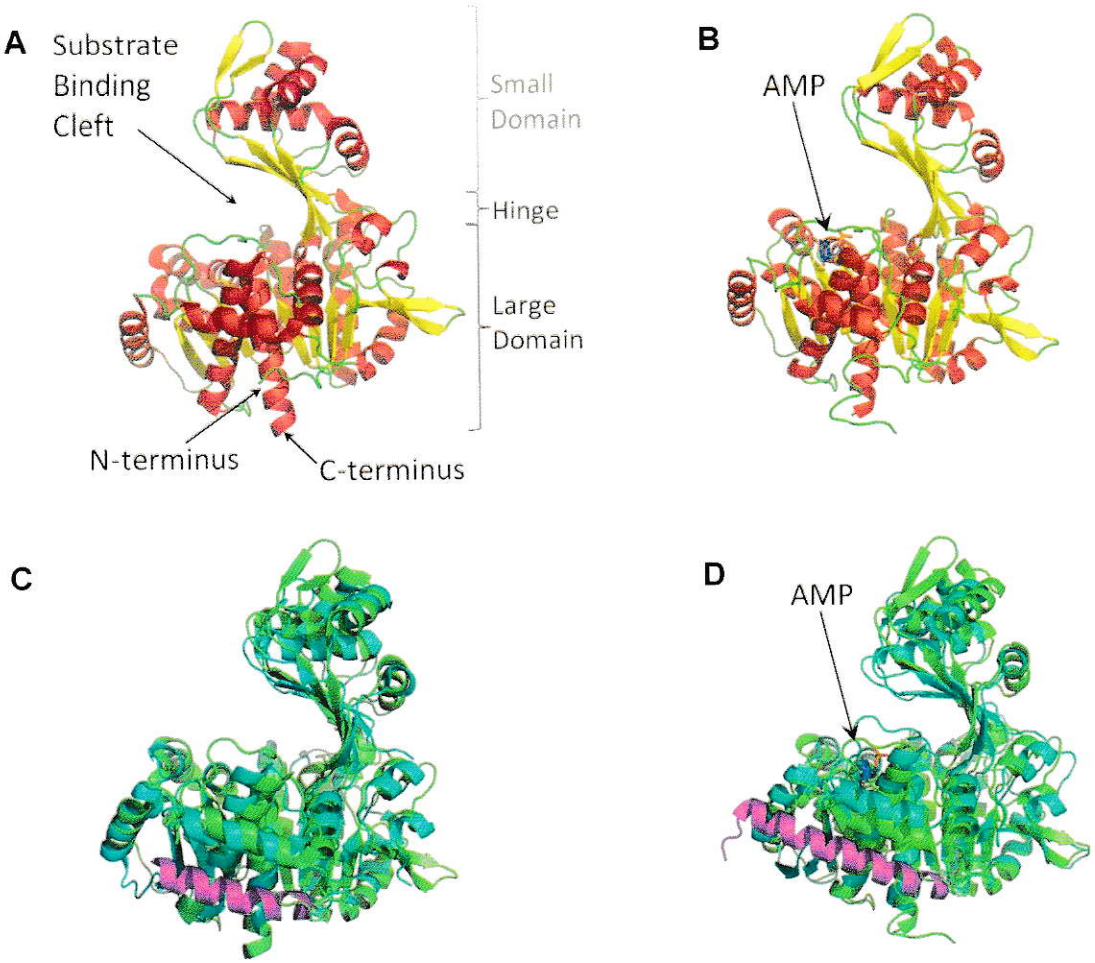


Figure 4.2: Structure of apo and AMP-bound *Ph*PFK and comparisons with ADP-GKs. **A.** Apo-*Ph*PFK. **B.** AMP-bound *Ph*PFK. **C.** Alignment of apo-*Ph*PFK structure (green) with *Ph*GK structure (cyan). The additional N-terminal α-helix of *Ph*GK not found in *Ph*PFK is shown in magenta. **D.** Alignment of AMP bound *Ph*PFK structure (green) and ADP bound *Tl*GK structure (cyan). The additional N-terminal α-helix of *Tl*GK not found in *Ph*PFK is shown in magenta. This figure was generated using PyMOL (DeLano, 2002).

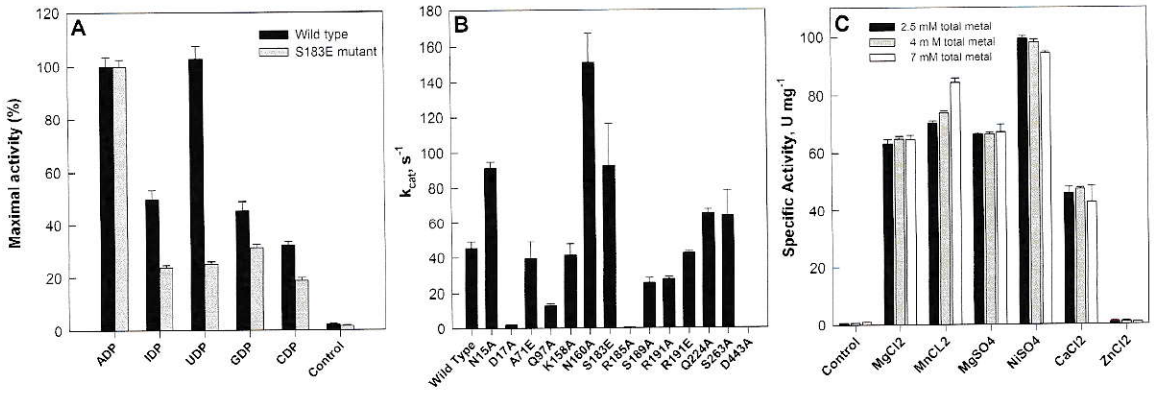


Figure 4.3: *Ph*PFK phosphoryl donor, kinetics, and divalent cation specificity. **A.** Wild type and S183E *Ph*PFK were assayed with different nucleotides as phosphoryl donors. The activity is shown as the percentage of the activity measured in presence of ADP (100%). **B.** Phosphofructokinase activity measurements were made for wild and mutant *Ph*PFKs. **C.** *Ph*PFK activity measured in presence of 2 mM F6P, 2 mM ADP, and the indicated divalent cation concentration. "Control" corresponds to the activity in the absence of nucleotide (**A**) and metal (**C**). Results are given as the mean ±S.E. of three experiments.

from *P. furiosus*, 2.3 and 0.11 mM, respectively (Tuininga *et al.*, 1999). However, the K_M for F6P is similar to the K_M obtained for the bifunctional ADP-GK/PFK from *M. jannaschii*, 9.6 μM, but the K_M for ADP is still much lower, 0.49 mM (Sakuraba *et al.*, 2002; Verhees *et al.*, 2001).

Divalent cations were required for *Ph*PFK activity. Five divalent metal cations were tested (Magnesium was tested with two counter ions to discard any effect of the anion) using 2.5, 4 and 7 mM of total metal with 2 mM total ADP and 2 mM total F6P (Figure 4.3C). In these conditions, the highest *Ph*PFK activity was obtained with $NiSO_4$ (~97 U/mg). The enzyme was also active in the presence of $MgCl_2$, $MnCl_2$, $MgSO_4$, and $CaCl_2$, but no significant activity was detected when the assay was performed in the presence of $ZnCl_2$. An increase in the enzyme activity, concomitantly with the increase in $MnCl_2$ concentration was observed, while the opposite effect is observed in the presence of $NiSO_4$. For the other divalent metals, the activity did not change

Table 4.1: Kinetic parameters of wild type and mutant *Ph*PFKs. For the D17A mutant the kinetic parameters for MgADP were notdetermined since the elevated K_M value for F6P hinders us to reach saturating conditions.

Enzyme	k_{cat}	K_M F6P	k_{cat}/K_M F6P	K_M ADP	k_{cat}/K_M ADP
	s^{-1}	μM	$M^{-1}s^{-1}$	μM	$M^{-1}s^{-1}$
Wild Type	45.5 ± 4.0	15.2 ± 2.5	$2.98 \cdot 10^6$	13.2 ± 2.4	$3.44 \cdot 10^6$
N15A	91.1 ± 3.9	103 ± 1.4	$8.86 \cdot 10^5$	23.8 ± 2.6	$3.82 \cdot 10^6$
D17A	1.98 ± 0.0051	22600 ± 840	$9.57 \cdot 10^1$	N.D.	N.D.
A71E	39.3 ± 9.9	22.2 ± 2.6	$1.77 \cdot 10^6$	13.6 ± 0.6	$2.82 \cdot 10^6$
Q97A	13.0 ± 1.0	7.08 ± 0.74	$1.83 \cdot 10^6$	7.47 ± 0.70	$1.74 \cdot 10^6$
K158A	41.0 ± 6.7	6500 ± 1300	$6.30 \cdot 10^3$	37.9 ± 6.7	$1.08 \cdot 10^6$
N160A	151 ± 16	415 ± 13	$3.65 \cdot 10^5$	72.4 ± 4.9	$2.09 \cdot 10^6$
S183E	91.7 ± 23.6	86.5 ± 9	$1.96 \cdot 10^6$	29.6 ± 2.2	$3.10 \cdot 10^6$
R185A	0.351 ± 0.016	318 ± 9.5	$1.11 \cdot 10^3$	473 ± 24	$7.44 \cdot 10^2$
S189A	25.5 ± 2.6	9.02 ± 2.3	$2.83 \cdot 10^6$	16.0 ± 1.2	$1.60 \cdot 10^6$
R191A	27.4 ± 1.5	254.4 ± 26.1	$1.07 \cdot 10^5$	14.5 ± 1.2	$1.89 \cdot 10^6$
R191E	42.5 ± 1	4871 ± 173	$8.73 \cdot 10^3$	60.5 ± 3.2	$7.02 \cdot 10^5$
Q224A	65.3 ± 2.6	38.6 ± 1.3	$1.69 \cdot 10^6$	29.5 ± 5.0	$2.21 \cdot 10^6$
S263A	64.1 ± 14	39.5 ± 7.8	$1.62 \cdot 10^6$	20.8 ± 3.4	$3.08 \cdot 10^6$
D443A	0.00480 ± 0.00040	12.7 ± 4.2	$3.79 \cdot 10^2$	13.9 ± 3.3	$3.46 \cdot 10^2$

with the cation concentration in the range tested. $CaCl_2$ showed the lowest activity with ~75% of the activity measured in the presence of $MgCl_2$. Although it is generally accepted that magnesium is the in vivo preferred metal, the ability of Ni^{2+}, Mn^{2+}, and to a lower extent Ca^{2+} to support *Ph*PFK activity with a high catalytic rate, suggests that other metals can substitute Mg^{2+}.

4.3.2 Overall structure

The apo-PhPFK structure (Figure 4.2A) was refined to 2.0 Å resolution with R = 0.202 and R_{free} = 0.245 (Table 4.2). The structure contained 450 of the expected 474 residues from the expressed protein construct (Table 4.2). In the final model, 97.9% of the residues were in favored regions of the Ramachandran plot and none were in the disallowed region (Table 4.2). *Ph*PFK was crystallized alone and in complex with ADP. The apo structure was solved by multiwavelength anomalous dispersion method using a selenomethionine derivative and the complex structure was solved by molecular

Table 4.2: Crystallographic data and refinement statistics. Values in parentheses are for the highest resolution shell. PDB codes are 1U2X and 3DRW for apo-*Ph*PFK and its complex with AMP, respectively.

	Apo-*Ph*PFK	**Complex with AMP**
Space group	$P2_1$	$P2_1$
Unit cell dimensions	a = 68.07 Å	a = 68.5 Å
	b = 99.93 Å	b = 104.2 Å
	v = 82.58 Å	c = 70.8 Å
	β = 110.38°	β = 105.1°
Resolution range (Å)	40.0 – 2.0	41.5 – 1.9
Unique/free reflections (5%)	69,682/3,508	71,005/3,770
Completeness (%)	99.2 (98.6)	98.7 (90.7)
R_{merge}	0.08 (0.455)	0.08 (0.353)
$I/\sigma(I)$	21.0 (3.57)	16.7 (3.25)
Redundancy	6.6(5.9)	4.5 (3.9)
R_{cryst}/R_{free} (%)	20.2/24.5	17.7/22.8
RMSD bond lenghts (Å)	0.004	0.005
RMSD bond angles (°)	0.609	0.705
Mean temperature factor (Å^2)	27.65	29.51

replacement using the apo structure as a model. AMP, instead of ADP, was found in the active site. The AMP-bound structure (Figure 4.2B) was refined to a resolution of 1.9 Å with $R = 0.177$ and $R_{free} = 0.228$ (Table 4.2). Both the apo and AMP-bound *Ph*PFK crystal structures contained two molecules in the asymmetric unit. For a more detailed discussion of oligomeric state see Supplemental Materials.

The crystal structure of *Ph*PFK was aligned with PhGK using the DaliLite server (Figure 4.2C) (Holm and Park, 2000). *Ph*PFK and *Ph*GK share 31% identity (Figure 4.1). Overall, 407 residues were aligned, resulting in a Z-Score andRMSD between α-carbon positions of 39.9 and 2.5 Å respectively. The AMP bound *Ph*PFK structure was aligned with the ADP bound *Tl*GK structure using the DaliLite server (Figure 4.2D) (Holm and Park, 2000). These two proteins also shared 31% identity with one another (Figure 4.1). A total of 434 residues were aligned with a Z-Score of 41.1 and a RMSD between α-carbon positions of 3.3 Å. In both cases the structures were

similar, sharing the same large and small α/β domain structures suggesting the common evolutionary origin of ADP-PFKs and ADP-GKs. However, the ADP-GK structures have an additional α-helix at their N-termini and in the ADP bound *Tl*GK structure, the small domain is closer to the large domain compared to AMP bound *Ph*PFK. This is discussed further below.

4.3.3 Nucleotide and Mg^{2+} binding site

Overall, the apo and AMP bound *Ph*PFK structures show very little difference (RMSD between α-carbon positions = 0.7 Å; Z-Score = 61.9). However, AMP binding results in slight changes in the large nucleotide binding loop as well as the main chain positions of the β-sheet and loop that lie directly N-terminal to the large nucleotide binding loop (Figure 4.4). A conformational change of the large nucleotide binding loop was also reported upon ADP binding to *Tl*GK. There is good conservation between the nucleotide binding site of *Ph*PFK and ADP-GKs, however some differences exist in the sugar bound, closed conformation (threading model, see below). In the AMP and glucose bound PfGK structure, Y344 of the large domain hydrogen bonds with E195 from the small domain. However, in *Ph*PFK the equivalent small domain residue, S183, is too short to hydrogen bond with Y338 of the large domain. This difference may account for the difference in phosphoryl group donor specificity in ADP-GKs and ADP-PFKs. ADP-PFKs are able to use GDP, to some extent, whereas ADP-GKs cannot. We made a S183E mutation and tested its ability to use different phosphoryl donors compared to the wild type enzyme (Figure 4.3A). There was no difference in activity between the mutant and wild type using ADP as a phosphoryl donor. However, the S183E mutant displayed significantly less activity when the phosphoryl donor was replaced with IDP, UDP, GDP, or CDP, compared to the wild type.

One of the most intriguing features of these enzymes is the position in that they can bind ADP in a way that both α- and β-phosphates are recognized mainly in the same

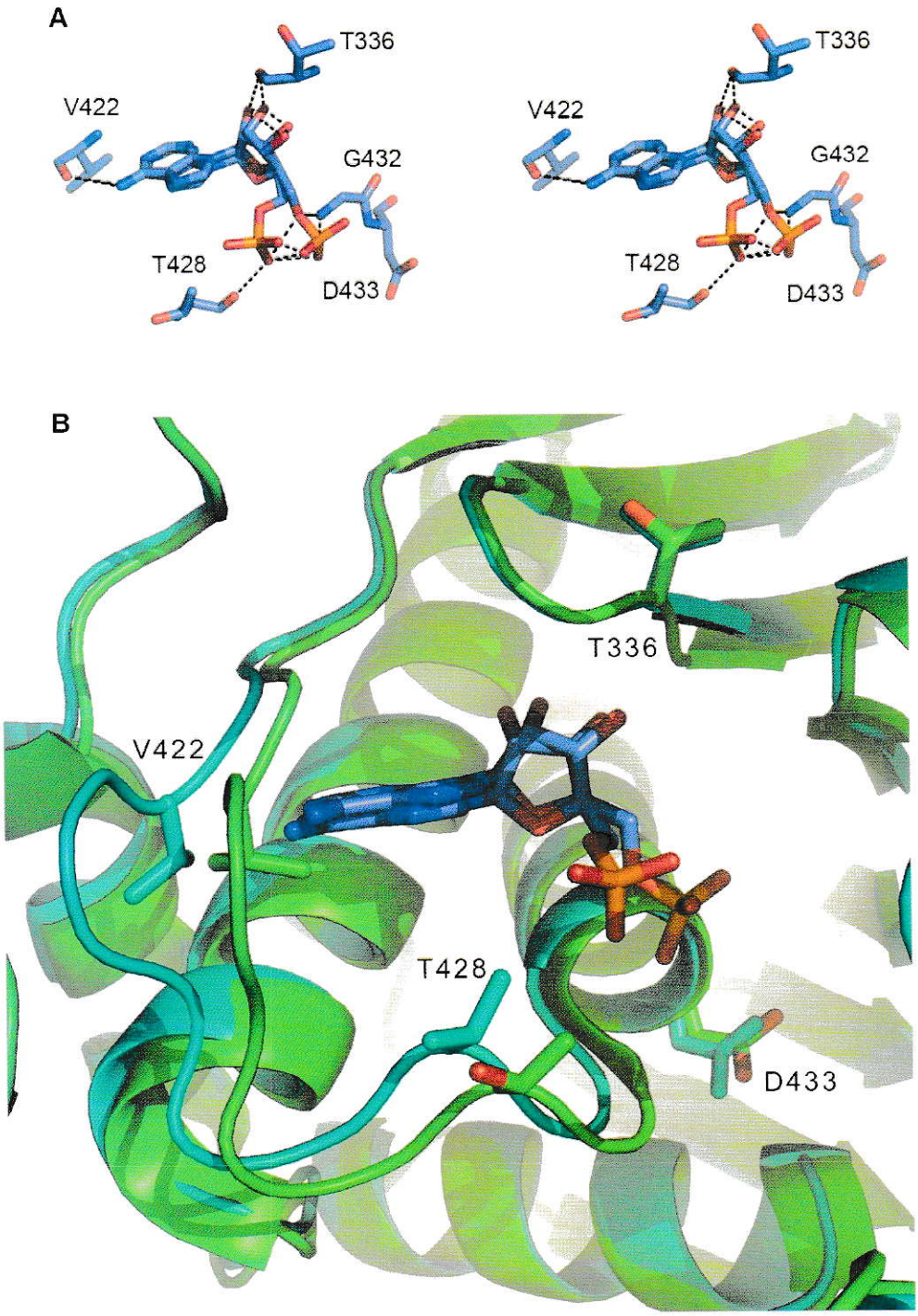


Figure 4.4: AMP binding site of *Ph*PFK. **A.** Stereo view of AMP-bound *Ph*PFK nucleotide binding site. **B.** Comparison of the nucleotide binding site of apo (green) and AMP-bound (cyan) *Ph*PFK. These figures were generated in PyMOL (DeLano, 2002).

way as the β- and γ-phosphates in the ATP dependent enzymes. In this respect, it has been proposed that the presence of a bulky side chain in the bottom of the nucleotide binding crevice (Y357 in the case of *T. litoralis* ADP-GK), below the sugar part of the molecule, could account for the change in specificity (Ito *et al.*, 2003). Moreover, it was proposed that since ADP-PFKs have leucine or isoleucine in this position they can marginally use ATP as phosphoryl donor as it was seen for the enzyme from *P. furiosus* (Tuininga *et al.*, 1999). Since the corresponding lateral chain of *Ph*PFK (L340) is placed in the corresponding position we tested directly the ability of this enzyme to use ATP using the Pyruvate kinase/Lactate dehydrogenase coupled assay, which specifically consumes the ADP produced. Surprisingly, no activity was detected with this nucleotide.

Almost all phosphate transferring enzymes have Mg^{2+} in their active site thereby assisting with the phosphate transfer through interactions with the β- and γ-phosphate groups of ATP or the β-phosphate of ADP, when ADP is the phosphoryl group donor. However, Mg2+ was not observed in the *Ph*PFK structure or any of the ADP-GK structures. Although, 17 water molecules were trapped in the AMP and glucose bound *Pf*GK structure, which offers sufficient space for both a magnesium ion and the β-phosphate of ADP. Despite the prerequisite for Mg^{2+}, this cation has never been observed in the structure of a member of this superfamily, including *Ph*PFK.

4.3.4 Reaction mechanism

A conserved aspartate residue in ribokinases, adenosine kinases, ADP-GKs, ADP-PFKs, and ADP-GK/PFK is believed to be the catalytic base involved in the phosphotransfer reaction of these kinases (Figure 4.1). Mutations of this residue in *Tl*GK resulted in a significant reduction of glucokinase activity (Ito *et al.*, 2003). Similarly, mutation of the corresponding aspartate in *Ph*PFK (D433) to alanine resulted in a loss of PFK activity (Table 4.1).

Furthermore, a conserved arginine residue that approaches the active site during the domain closing event associated with sugar binding is believed to attract the terminal phosphate group of the phosphoryl donor, stimulating cleavage of the phosphodiester bond and transfer of the phosphate group. Mutation of R205 in *Tl*GK to alanine resulted in less than 0.1% the activity of the wild type enzyme (Ito *et al.*, 2003). We observed the same effect upon mutation of the corresponding arginine residue in *Ph*PFK (R185) to alanine (Table 4.1). These two findings support a conserved mechanism amongst ADP-GKs, ADP-PFKs, and ADP-GK/PFKs.

4.3.5 Substrate induced fit

A large conformational change occurs in ADP-GKs as well as ATP-dependent hexokinases/glucokinases and ribokinases upon binding their respective substrates (Figure 4.5) (Ito *et al.*, 2003; Anderson *et al.*, 1978; Mulichak *et al.*, 1998; Sigrell *et al.*, 1998; St Charles *et al.*, 1994). The flexible loops located between the large and small domains form a hinge that folds, closing the cleft between the two domains upon substrate binding (Ito *et al.*, 2003; Anderson *et al.*, 1978; Mulichak *et al.*, 1998; Sigrell *et al.*, 1998; St Charles *et al.*, 1994). This conformational change appears to be essential for positioning the catalytic residues of ADP-GKs and ribokinases (Ito *et al.*, 2003; Sigrell *et al.*, 1998). The three ADP-GKs structures all display varying amounts of domain closure (Figure 4.5) (Ito *et al.*, 2001; Tsuge *et al.*, 2002; Ito *et al.*, 2003). The apo-*Ph*GK is the most open conformation (Figure 4.5)(Tsuge *et al.*, 2002). There is a rotation of about 20° between the large and small domain in the ADP bound *Tl*GK structure compared to the *Ph*GK structure and an additional 20° rotation between the two domains in the AMP and glucose bound *Pf*GK structure compared to the *Tl*GK structure (Figure 4.5)(Ito *et al.*, 2001; Tsuge *et al.*, 2002; Ito *et al.*, 2003). It was initially reported that the *Tl*GK structure was the most open conformation because they were able to successfully soak ADP out of the crystal, which indicates the absence of a large confor-

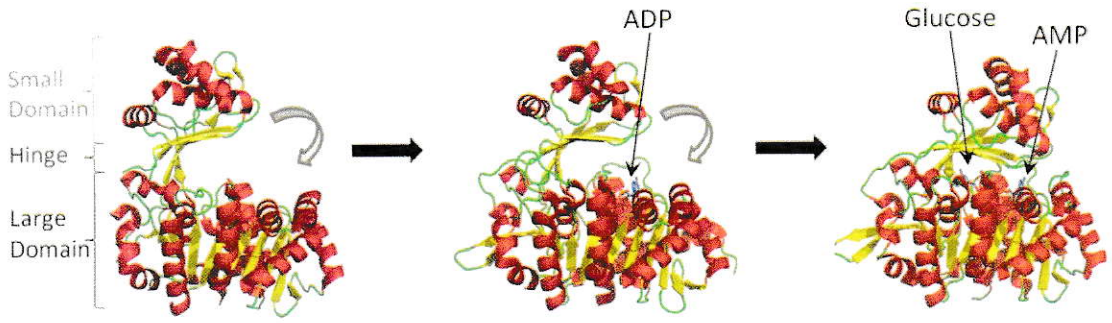


Figure 4.5: Substrate induced conformational changes of ADP-GKs. Left, apo structure from *Pyrococcus horikoshii* (Tsuge *et al.*, 2002); middle, ADP-bound structure from *Thermococcus litoralis* (Ito *et al.*, 2001); and right, AMP and glucose bound structure from *Pyrococcus furiosus* (Ito *et al.*, 2003).

mational change (Ito *et al.*, 2001). The *Ph*GK structure clearly presented a more open conformation, which led to the suggestion that *Tl*GK had been trapped in the ADP bound conformation following the removal of the ADP from the crystal (Tsuge *et al.*, 2002). However, ADP could not be soaked into the *Ph*GK crystal, which is thought to be due to an ADP-dependent domain closing (Tsuge *et al.*, 2002). Interestingly, no such conformational change was observed between the apo and AMP-bound *Ph*PFK structures. It could be that the β-phosphate of ADP is required to fully induce this conformational change. The amine nitrogen of R205 from *Tl*GK is 3.8 Å away from the β-phosphate of ADP and may form important interactions that help to stabilize a more closed conformation (Ito *et al.*, 2001). One other possibility is that for *Ph*PFK both co-factor and substrate are required for the conformational change observed as seen in other structures. In any case, the lack of conformational change in *Ph*PFK in the presence of AMP is somewhat surprising given the structure is from co-crystallization instead of soaking.

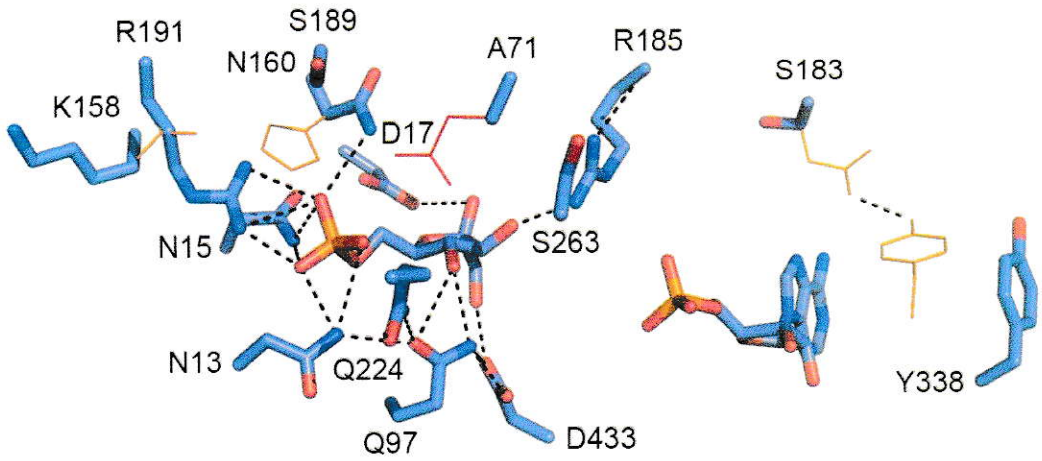


Figure 4.6: Model of *Ph*PFK bound to F6P and AMP. Docking model of the *Ph*PFK-F6P complex. Where different, the positions of residues conserved in ADP-GKs are shown in orange. The red residue corresponds to the glutamate residue conserved in the bifunctional ADP-GK/PFK (E82) and in ADP-GKs (E88). Black hashed lines represent hydrogen bonding.

4.3.6 Sugar binding site

Attempts to co-crystallize or soak crystals with F6P or fructose-1, 6-bisphosphate both in the presence or absence of ADP or nonhydrolyzable ADP analogues were unsuccessful. Due to the overall similarity between the *Ph*PFK structures and ADP-GKs structures, we modeled *Ph*PFK in he closed conformation bound to F6P and AMP based on the AMP and glucose bound *Pf*GK structure (Figure 4.6). The primary sequence of *Ph*PFK was threaded onto the *Pf*GK structure bound to both AMP and glucose, and the resulting model was energy minimized. F6P was docked with the structure using AutoDock (Morris *et al.*, 1998). All of the ten lowest energy docking results placed the phosphate in the same position with only slight differences in the placement of the sugar moiety.

To validate our model and gain further insights into substrate specificity, relevant residues from the F6P binding pocket of *Ph*PFK, the highly conserved residues N15,

D17, Q97, K158, N160, R185, S189, R191, Q224, S263, and D443, were mutated to alanine in order to establish their importance to substrate binding and catalysis. Of these, N15A, D17A, K158A, N160A, R185A, and R191A show the most significant increase in their K_M value for F6P (Table 4.1). N160, N15, and R191 form one, two, and three hydrogen bonds with the phosphate group of F6P, respectively as judged by the docked conformation (Figure 4.6). R185 and D17 form hydrogen bonds with the sugar moiety of F6P (Figures 4.6). As a result, loss of these interactions weakens PhPFK binding to its substrate. K158 is too far to hydrogen bond with the phosphate of F6P in our model (Figure 4.6). However, it is strictly conserved in PFKs and mutation of this residue has the most significant effect on F6P binding after the D17A mutation (Table 4.1). There are several possible explanations for this observation; however, it is most likely a result of the threading procedure itself, since there is no negative charge to stabilize the positive lateral chain of K158 inside the binding pocket and therefore repelling it in the opposite direction. In fact, simple rotation of the side chain of K158 in our model would result in a distance of less than 3.5 Å to the phosphate group of F6P. This was confirmed using a molecular dynamics simulation (see Supplemental Materials).In this conformation, the lateral chains of K158 and R191 could form a salt bridge with the phosphate group of F6P, which would explain the behavior of the mutant enzyme.

Interestingly, the half of the sugar ring opposite to the endocyclic oxygen is recognized mainly in the same way as the comparable part of glucose by ADP-GKs. As expected the residues responsible of this like N15, D17, and G96 are conserved in both specificities. As can be seen below the sugar discrimination comes mainly by the interactions with the other half of the ring. This suggests that the sugar binding site in the family is strongly plastic to accommodate a furanose-P and a pyranose with mainly the same residues, but still very selective to precisely discriminate between the two of them. Q224A and S263A contribute little to the sugar binding: both mutations result

in small increases in K_M for F6P of about 2- to 3-fold (Table 4.1). Surprisingly, Q97A, S189A, and D433A mutations lead to a small decrease in K_M for F6P despite the loss of a hydrogen bond in the case of Q97A and D433A (Table 4.1). Also, R185 and D433 make contact withthe C1 hydroxyl group, which supports the proposed catalytic mechanism. However, due to the lack of the phosphate group from ADP, the position of R185 could be somewhat distorted. Nevertheless, our model is well supported by the mutational data.

4.3.7 Differences between ADP-PFKs and ADP-GKs

The conserved residues N13, N15, D17, G96, Q97 and R185 form hydrogen bonds with F6P, as shown in figure 4.6A. K158, N160, and R191 also form hydrogen bonds with F6P (Figure 4.6). These residues are present in the bifunctional ADP-GK/PFK from *M. jannaschii*, but the equivalent residues in ADP-GKs are a non-conserved small residue, histidine and aspartate, respectively. A similar scenario was seen in a previous *in silico* study using the bifunctional enzyme from *M. jannaschii* (Merino and Guixé, 2008). The positive charge on R191 of *Ph*PFK attracts the phosphate group of F6P, and is still compatible with glucose binding. However, when replaced by aspartate in ADP-GKs, the negatively charged side chain likely repels the phosphate group of F6P, which prevents ADP-GKs from binding F6P. In order to assess the role of R191 in F6P binding, we made two mutations, R191A and R191E. The R191A mutation removes three hydrogen bonds to the phosphate group on F6P and causes a ~17-fold increase in the K_M value for F6P (Table 4.1). However, as predicted, the charge reversal of the R191E mutation has a much more pronounced effect on F6P binding, a ~325-fold increase in K_M (Table 4.1). Even so, the R191E mutant enzyme is not able to phosphorylate glucose, suggesting that the presence of this charged residue in the bottom of the sugar binding site in all ADP-GKs contribute to sugar discrimination by hindering the protein from binding F6P and not by increasing the affinity for glucose.

The key difference between kinases that are capable of phosphorylating glucose and those that cannot is a conserved glutamate residue found in ADP-GKs (E88 from *Pf*GK) and ADP-GK/PFK (E82) that is replaced with alanine in PhPFK (A71). In the *Pf*GK-glucose-AMP complex structure, E88 lateral chain forms a hydrogen bond with the C2 hydroxyl group of the bound glucose molecule. This interaction is thought to be critical for orienting the glucose molecule in the active site and it has also been highlighted as crucial for the sugar specificity in the past (Merino and Guixé, 2008; Sakuraba *et al.*, 2002; Ito *et al.*, 2003). As a result, ADP-PFKs cannot phosphorylate glucose. We tested this hypothesis by mutating A71 to glutamate. As expected, this mutation has little effect on F6P binding compared to the wild type (Table 4.1). However, the A71E mutant enzyme can also bind glucose (K_M = 3.95 ± 0.2 mM) and is capable of phosphorylating glucose (k_{cat} of 2.68 ± 0.05 s^{-1}) (Table 4.1). In contrast, the wild type enzyme displays no activity for glucose. This result demonstrates that A71 is a key player in determining substrate specificity. Interestingly, the catalytic mechanism described before for the *Ph*PFK seems to be even more general since comparison of these features with the non-homologous ATP dependent PFK (Pfk-1) from *E. coli* pointed out striking similarities. In terms of catalysis, in Pfk-1, Asp127 plays a critical role as a general base, increasing the nucleophilicity of the 1-hydroxyl of F6P by abstracting its proton and permitting attack on the γ-phosphate of the substrate ATP in a similar way as D433 in *Ph*PFK (Hellinga and Evans, 1987). In addition, Arg162 and Arg243 interact with the 6-phosphate of F6P. Truncation of these residues to serine results in enzymes with decreased F6P binding ability and reduced cooperativity, but little change in catalytic ability in a very similar way to K158 and R191 (Berger and Evans, 1990). Then, it would appear that some structural features related to catalysis in ADP dependent PFK have analogous counterparts in the ATP PFKs.

4.4 Conclusion

In summary, we determined the first crystal structure of an ADP-dependent phosphofructokinase both alone and in complex with AMP and through comprehensive mutagenesis we identified residues that were critical for both the phosphotransfer reaction and substrate binding. Our results demonstrate that the overall structure of ADP-PFK is similar to that of ADP-GK and that the two share a common mechanism of action. However, unlike the available ADP-GK structures, no conformational change was observed upon nucleotide binding. This was unexpected and may indicate that the nucleotide-dependent conformational change is a non-essential part of ADP-dependent sugar kinase mechanism. Moreover, we identified four key residues responsible for the sugar binding specificity. Using this knowledge, we were able to generate a mutant *Ph*PFK that is capable of phosphorylating both F6P and glucose.

4.5 Acknowledgments

We thank all members of the Ontario Centre for Structural Proteomics (Structural Proteomics in Toronto, SPiT) for help in conducting experiments. We also thank Dr. Richard C. Garratt from the Instituto de Física de São Carlos, Universidade de São Paulo, Brasil, where the Dynamic Light Scattering experiments were performed. The work was supported by Genome Canada (through the Ontario Genomics Initiative), by National Institutes of Health grant GM074942, by grant from Fondo Nacional de Desarrollo Científico y Tecnológico (Fondecyt) 1070111, by the U.S. Department of Energy, Office of Biological Research, under contract DE-AC02-06CH11357, and by Canadian Institutes of Health Research. ZJ is a Canada Research Chair in Structural Biology.

Authors contribution: Protein crystallization, crystallography, and protein-ligand docking were performed by Mark Currie, Andrew Wong, and Zongchao Jia (Queen's University). All molecular biology was performed by Tatiana Skarina, Alexander Singer, Greg Brown, Alexei Savchenko, and Alexander Yakunin (Midwest Center for Structural Genomics). Kinetic characterizatons and molecular dynamics were performed by Felipe Merino, Andres Caniuguir, and Victoria Guixe (Universidad de Chile).

CHAPTER 5

General discussion and concluding remarks

5.1 Evolution of the substrate specificity

As it has been discussed in the previous chapters, the ribokinase superfamily contains enzymes that catalyze the transfer of the terminal phosphate of a nucleotide phosphate to the methyl alcohol end of a big number of small molecules which includes pyridoxal, pyrimidine derivatives, nucleosides, and several sugars. While there are many phosphoryl acceptor substrates in this superfamily, just two nucleotides, ADP and ATP, are described as the primary phosphoryl donors. Of course, the specificity is not strict, and some other nucleotides can replace them. For instance, it has been shown for several ADP-dependent enzymes that other purines (such as GDP) and even pyrimidines (such as UDP, see Chapter 4) can be used (Guixé and Merino, 2009). Also, GTP can also be used by the phosphofructokinase-2 from *E. coli* and even produce substrate inhibition (unpublished results). Yet, it is important to remember that only those nucleotides with the right number of phosphates (either two for the ADP dependent enzymes or three for the ATP dependent) can be used, as it has been demonstrated in Chapter 2 and reviewed by us elsewhere (Guixé and Merino, 2009).

Interestingly, although there is a clear correlation between the topology of the C-terminal end of these proteins and their nucleotide specificity, the circular permutated

version of the ADP-dependent glucokinase from *T. litoralis* is unable to use ATP as the phosphoryl source. Even more, the enzymatic constants are very similar to that of the wild type enzyme, which suggests that the topological reordering has no effect on the nucleotide binding process. It has been proposed before that the bulky side chain of Y357 of *Tl*GK, which is located below the sugar when ADP is docked at the binding site, is responsible for the inability of this enzyme to use ATP (Ito *et al.*, 2001). However, we indirectly demonstrated that this is not the case since for *Ph*PFK the presence of a significantly less bulky side chain does not produce an enzyme with even marginal ATP-dependent activity. It is clear then, that neither the topological rearrangement nor evident point mutations are able to explain the specificity switch. Interestingly, it has been shown before through ancestral gene reconstruction that in order to attain a new function some proteins need to accumulate at least two mutations on a epistatic fashion. The first of them acts as a permissive mutation enabling the second one to do the true functional shift (Bridgham *et al.*, 2009). Of course, the restrictive kind of mutations can also occur (Harms and Thornton, 2010). In this way, it is possible that the C-terminal circular permutation constitutes a permissive mutation that now needs a further step to produce an ATP-dependent enzyme from an ADP-dependent one through some simple point mutations. As there is no clear candidate for this, we are now addressing this issue through directed evolution.

The adaptive value of the appearance of the ADP-dependent enzymes has been a matter of great debate. As we have argued before (see for instance Chapter 2), it is most likely unrelated to the temperature at which most of the *thermococcales* grow. The most intriguing question arising here is what happens with the adenylate charge inside these archaea. As they present a glyceraldehyde-3-phosphate ferredoxin oxidoreductase (Mukund and Adams, 1995) which does not produce ATP and both glucose and fructose-6-phosphate are phosphorylated using ADP as phosphoryl donor, it was though that this modified glycolysis had a net ATP production of zero. However,

it has been demonstrated by Sakuraba *et al.* (2004) that the pyruvate kinase from *P. furiosus* catalyze the synthesis of ATP from AMP, phosphoenolpyruvate, and Pi. In this way, the pathway from glucose to pyruvate produces two ATP molecules from every glucose degraded. On the other hand, the metabolic role of the eukaryotic ADP-dependent glucokinases is unclear, but they have been suggested to be used in ischemic conditions (Ronimus and Morgan, 2004).

Opposed to the phosphoryl donor case, as we have already mentioned above, the ribokinase superfamily present a rich variety of possible phosphoryl acceptor. Now, the structural relation is quite clear. The presence of the small domain appears to be the key element in the acceptor specificity problem. Indeed, those enzymes which lacks the small domain (the vitamin kinase like branch) are able to phosphorylate just pyrimidine derivatives and some structurally similar molecules. On the other hand, those enzymes that do present the small domain phosphorylate sugar containing (usually a furanose) molecules. A key aspartic residue that comes from the first strand of the small domain is highly conserved. Crystallographic structures show that this side chain makes two very short h-bonds with two hydroxyl groups of the substrate (e.g. O4 and O3 in fructose-6-phosphate, O2 and O3 in ribose, O3 and O4 in glucose). The removal of this aspartic acid in *Ph*PFK (D17) increases the K_M value of the enzyme more than three orders of magnitude. The mutation of the homologous side chain of the phosphofructokinase-2 from *E. coli* (D14) also produces an enzyme with a highly reduced stability and affinity for fructose-6-phosphate (unpublished results).

As it has been explored in Chapters 3 and 4 the structural determinants of the sugar specificity of the ADP-dependent enzymes is mainly related with the endocyclic oxygen half of the sugars mainly through side chains that comes from the small domain. It is clear that the ability of the ADP-dependent enzymes to bind fructose-6-phosphate depends on the presence of two positively charged residues that stabilize the negative charge of the substrate while the glucose binding is mainly determined by the presence

of a key glutamic residue that interacts with the hydroxyl group attached to the carbon two of glucose. Indeed, the introduction of this glutamic residue in the corresponding position of *Ph*PFK (the A71E mutation) produces a dual enzyme.

Recently, we used a more elegant method known as explicit likelihood of subset covariation (ELCS) (Dekker *et al.*, 2004) to explore the correlation between mutations to search for the structural specificity determinants.

Figure 5.1 shows the group of side chains with the highest ELCS score. Surprisingly, the group contain a side chain that belongs to the NXXE motif (see below) and some residues that we proposed before as specificity related. The role of the R48/D65[1], R65/S76, P73/F90 mutations is not clear, but seem to be related to the dynamics of the small domain. K158/C174, N160/H176, and R191/D203 are clearly interacting with the sugars. Interestingly, when the position R191/D203 presents an arginine, this positive side chain coordinates the phosphate group present in the fructose-6-phosphate molecule. On the other hand, when it presents an aspartic acid, this side chain interacts with the histidine in the N160/H176 position, allowing the histidine to be correctly positioned to make an h-bond with the O2 hydroxyl group of glucose. Curiously, the position equivalent to A71 of *Ph*PFK does not appear to be correlated with other positions by the ELCS method. However this could be due to the small amount of sequence information used for the analysis. Surprisingly, besides the A71E mutation we have recently produced a new bifunctional enzyme by changing the N160 side chain of *Ph*PFK into a histidine. This mutation dramatically increases the K_M value for fructose-6-phosphate to 6.3 $\pm$ 0.72 mM and decreases the k_{cat} almost four-fold. However, for this mutant no clear saturation is seen even for 25 mM glucose. Based on a Lineweaver-Burk plot, it is possible to estimate a k_{cat} of 2.42 s^{-1} and a K_M value of 25.3 mM. Strikingly, both mutations point to an interaction with the O2 hydroxyl group of glucose.

[1]We use the numbering of *Ph*PFK/*Pf*GK for the correlated mutations

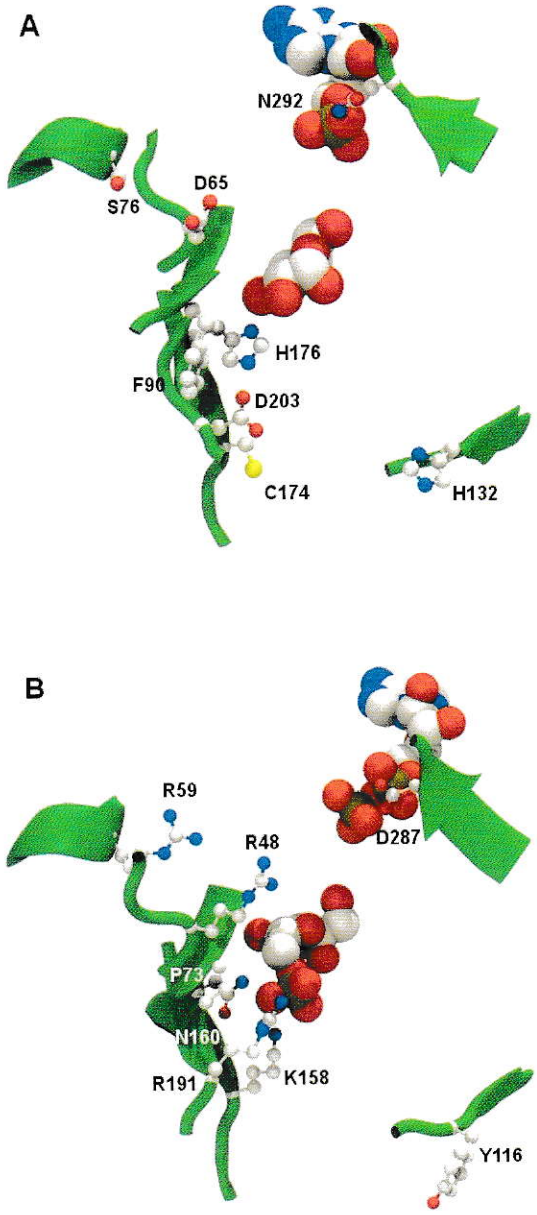


Figure 5.1: First cluster of correlated mutations in the PfkC family identified by the ELCS method. **A.** Crystal structure of the glucokinase from *P. furiosus*. Glucose and AMP are shown. **B.** Structural model for the ternary complex between the phosphofructokinase from *P. horikoshii*, ADP and fructose-6-phosphate. The coordinates were derived from the molecular dynamics simulation performed in Chapter 4.

5.2 Core catalytic mechanism

Amongst the different reactions catalyzed by the enzymes that belong to a particular superfamily, one would expect that a similar catalytic mechanism would be present in all those proteins. On the basis of structural and sequence comparison of several members of the ribokinase superfamily an aspartic acid residue inside a highly conserved motif called GXGD has been proposed to act as a general base that removes the proton from the hydroxyl acceptor group and thus activates it to attack the terminal phosphate of ADP/ATP (Figure 5.2). The removal of this side chain in *Ph*PFK decreases the enzyme activity by four orders of magnitude. The same mutation in *Tl*GK produces also a significant decrease in the catalytic constant (Ito *et al.*, 2003). Indeed, the mutation of this side chain in several members of the ribokinase superfamily shows a very similar effect (Mathews *et al.*, 1998; Maj *et al.*, 2000; Datta *et al.*, 2005). Interestingly, some members of the vitamin kinases like branch (hydroxyethylthiazole kinase and 4-amino-5-hydroxymethyl-2-methylpyrimidine kinase) have a cysteine residue at the position corresponding to the GXGD aspartic acid (Figure 5.2). Mutation of this residue in the hydroxyethylthiazole kinase from *Bacillus subtilis* (C198) into an aspartic acid produces an enzyme with a 9-fold higher activity, while mutation of this cysteine by serine or alanine produced enzymes with specific activities of 20% and 40% compared to the wild type (Campobasso *et al.*, 2000). Based on this Campobasso *et al.* (2000) proposed that in this case the γ-phosphate of ATP and not the cysteine residue may function as the alcohol activating base. This could indicate that, similar to the specificity case, the presence of the small domain could modify the catalytic mechanism of this superfamily. Yet, to prove this high level calculations are needed.

To date, all of the characterized enzymes of the superfamily have shown a strict dependence on divalent metal cations to catalyze the phosphoryl transfer. Moreover, it has been demonstrated that the binding of a second metal to these enzymes modulates the activity. However, depending on the specificity of the enzyme it acts as an inhibitor

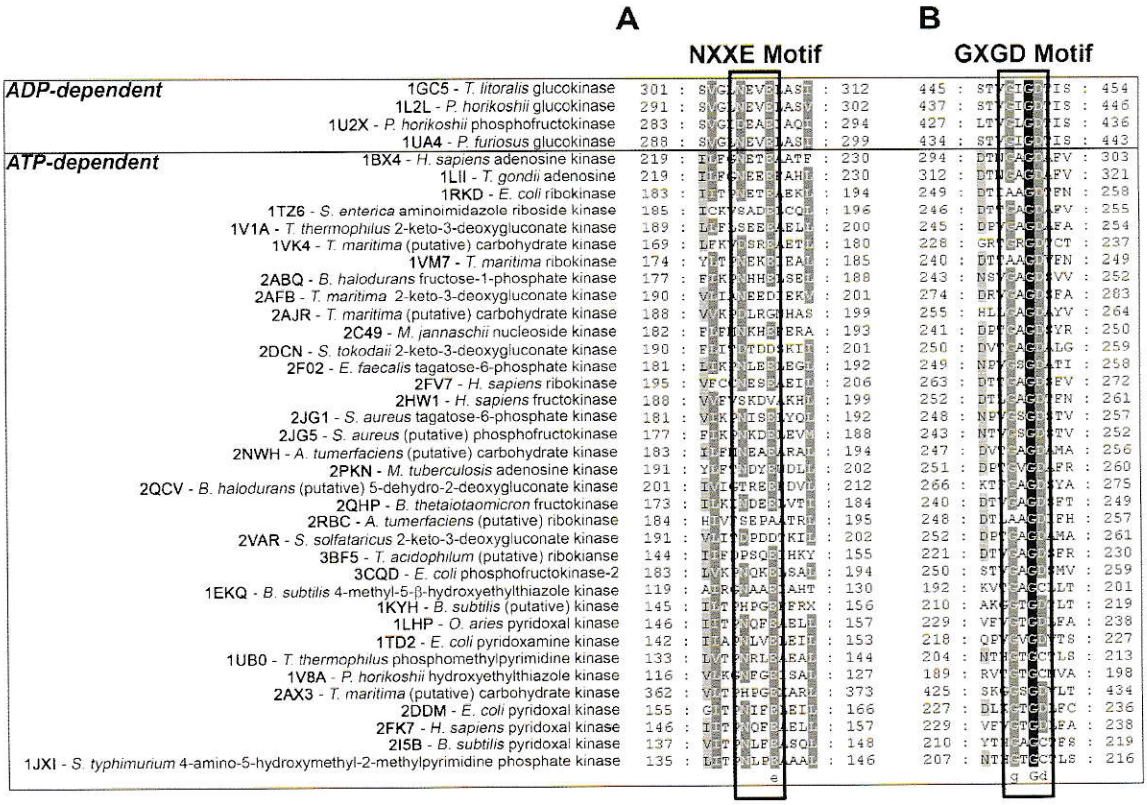


Figure 5.2: Sequence alignment of the motifs related with the core catalytic mechanism of the ribokinase superfamily. **A.** NXXE motif. **B.** GXGD motif. The PDB code of every enzyme is shown in bold letters.

or an activator. For instance, the phosphofructokinase-2 from *E. coli* is activated by magnesium (Parducci *et al.*, 2006; Rivas-Pardo *et al.*, 2011) while the adenosine kinase from different source are inhibited by it (Maj *et al.*, 2002). On the other hand, the ADP-dependent glucokinases are inhibited by manganese while divalent metal cations have almost no regulatory effect on the ADP-dependent phosphofructokinases (Merino *et al.*, In editorial revision). It has been demonstrated that another highly conserved motif called NXXE is related with the metal coordination in this superfamily (Figure 5.2) (Parducci *et al.*, 2006; Guixé and Merino, 2009; Maj *et al.*, 2002; Merino *et al.*, In editorial revision). Mutation of residues in this motif affects both k_{cat} and regulation. Interestingly, on the ADP dependent enzymes only glucokinases present an asparagine inside the NXXE motif. The phosphofructokinases have an aspartic acid

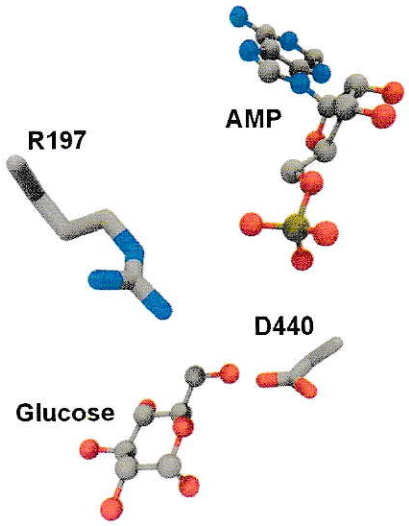


Figure 5.3: Residues related with the catalytic mechanism of the ADP-dependent kinases. The crystallographic structure of *Pf*GK is used. Glucose, AMP, D440 (from the GXGD motif) and R197 are shown.

residue in the equivalent position (Figure 5.2). This substitution is correlated with the specificity determining residues as shown by the ELCS method. This changes the electrostatic environment of the metal binding site, which could explain the different regulatory properties between ADP-dependent glucokinases and phosphofructokinases (Merino *et al.*, In editorial revision). As the substitution is correlated with specificity, it suggests that in the modified Embden-Meyerhof pathway glucokinases need to be more regulated than phosphofructokinases.

Besides the aspartic acid inside the GXGD motif and the NXXE motif residues, the enzymes of the ribokinase superfamily always present a basic residue (such as lysine or arginine) in the active site which is related to the transfer reaction. However, as these residues are not conserved through the different specificities they have been much less studied. For instance, the phosphofructokinase-2 from *E. coli* has a lysine residue (K27) which protrudes to the other monomer in the dimer, and it is proposed to have a dual purpose: it stabilizes the transition state by transferring a proton to it and also it acts as a switch in the competitive binding of fructose-6-phosphate and the regulatory

ATP (Cabrera *et al.*, 2011). On the other hand, the ADP-dependent enzymes have an arginine residue that is thought to be involved in the reaction (Figure 5.3). The removal of this side chain in *PhPFK* (R197) (See Chapter 2) or *Tl*GK (Ito *et al.*, 2003) produces enzymes with extremely reduced catalytic constants. Some molecular dynamics simulations of the catalytic ensembles of *Pf*GK and *Ph*PFK have shown an interesting geometry of the interaction between this arginine and the β-phosphate of ADP. Here the phosphate group is coordinated by an ε-nitrogen hydrogen and a η-nitrogen hydrogen (Merino *et al.*, In editorial revision). Given its position (Figure 5.3) and considering that the ε-nitrogen has the lowest pK_a inside the guanidinium head, we believe the this arginine residue also serves two functions: on the one hand it stabilizes the transition state by transferring a proton to it and, on the other hand, it pulls the phosphate towards the sugar, maybe assisted by the domain bending motion of the enzyme.

5.3 Concluding remarks

The discovery of the ADP-dependent kinases in hyperthemophilic archaea led to some authors to suggest that this feature is an adaptation to high temperatures. However, several facts demonstrated that the presence of these proteins is not related to the hyperthermophilic life. On the other hand, a highly suggestive topological reordering of the nucleotide binding site appeared to be related with the nucleotide specificity. However, we have proven through this work that it is also unrelated with the ATP/ADP discrimination. As we still do not understand the structural determinants of this functional switch, the future work of our group should be pointed at that objective.

Studies about the evolution of the ADP-dependent sugar kinase family showed that the root of the family is located inside the glucokinase group, demonstrating that the bifunctional (glucokinase/phosphofructokinase) enzyme is not an ancestral form, but

could be a transitional form from glucokinase to phosphofructokinase. Also, we have shown that the ADP-dependent enzymes are closely related to the PfkB group of kinases. Unfortunately, to date it has not been possible to obtain the crystal structure of any ADP-dependent phosphofructokinase in the presence of fructose-6-phosphate. However, based on structural modeling we have been able to understand partially the structure/specificity relation up to the point where we can produce bifunctional enzymes from specific ones. Strikingly, the sugar discrimination is somehow concentrated in very few hotspots in the structure. Indeed, the introduction of just one h-bond or some salt bridges seems to modulate the affinity for glucose or fructose-6-phosphate respectively. Unfortunately, to date, we have been unable to absolutely switch the specificity of these enzymes.

References

Allers, T. and Mevarech, M. (2005). Archaeal genetics - the third way. *Nat. Rev. Genet.* **6**(1): 58–73.

Anderson, C.M., Stenkamp, R.E., and Steitz, T.A. (1978). Sequencing a protein by x-ray crystallography. II. Refinement of yeast hexokinase B co-ordinates and sequence at 2.1 Å resolution. *J. Mol. Biol.* **123**(1): 15–33.

Atkinson, D.E. and Walton, G.M. (1965). Kinetics of regulatory enzymes. *Escherichia coli* phosphofructokinase. *J. Biol. Chem.* **240**: 757–763.

de Bakker, P.I., Hünenberger, P.H., and McCammon, J.A. (1999). Molecular dynamics simulations of the hyperthermophilic protein sac7d from *Sulfolobus acidocaldarius*: contribution of salt bridges to thermostability. *J. Mol. Biol.* **285**(4): 1811–1830.

Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Howe, K.L., and Sonnhammer, E.L. (2000). The Pfam protein families database. *Nucleic Acids Res.* **28**(1): 263–266.

Berger, S.A. and Evans, P.R. (1990). Active-site mutants altering the cooperativity of *E. coli* phosphofructokinase. *Nature* **343**(6258): 575–576.

Bork, P., Sander, C., and Valencia, A. (1993). Convergent evolution of similar enzymatic function on different protein folds: the hexokinase, ribokinase, and galactokinase families of sugar kinases. *Protein Sci.* **2**(1): 31–40.

Bridgham, J.T., Ortlund, E.A., and Thornton, J.W. (2009). An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* **461**(7263): 515–519.

Brünger, A., Adams, P., Clore, G., DeLano, W., Gros, P., Grosse-Kunstleve, R., Jiang, J., Kuszewski, J., Nilges, M., Pannu, N., *et al.* (1998). Crystallography & NMR system: a new software suite for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **54**: 905–921.

Cabrera, R., Ambrosio, A.L.B., Garratt, R.C., Guixé, V., and Babul, J. (2008). Crystallographic structure of phosphofructokinase-2 from *Escherichia coli* in complex with two ATP molecules. Implications for substrate inhibition. *J. Mol. Biol.* **383**(3): 588–602.

Cabrera, R., Babul, J., and Guixé, V. (2010). Ribokinase family evolution and the role of conserved residues at the active site of the PfkB subfamily representative, Pfk-2 from *Escherichia coli*. *Arch. Biochem. Biophys.* **502**(1): 23–30.

Cabrera, R., Baez, M., Pereira, H.M., Caniuguir, A., Garratt, R.C., and Babul, J. (2011). The crystal complex of phosphofructokinase-2 of *Escherichia coli* with fructose-6-phosphate: kinetic and structural analysis of the allosteric ATP inhibition. *J. Biol. Chem.* **286**(7): 5774–5783.

Campobasso, N., Mathews, I.I., Begley, T.P., and Ealick, S.E. (2000). Crystal structure of 4-methyl-5-beta-hydroxyethylthiazole kinase from *Bacillus subtilis* at 1.5 Å resolution. *Biochemistry* **39**(27): 7868–7877.

Cheng, G., Bennett, E.M., Begley, T.P., and Ealick, S.E. (2002). Crystal structure of 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate kinase from *Salmonella typhimurium* at 2.3 Å resolution. *Structure* **10**(2): 225–235.

Cornish-Bowden, A. (1995). Fundamentals of Enzyme Kinetics. Portland Press, London.

Cuff, J.A., Clamp, M.E., Siddiqui, A.S., Finlay, M., and Barton, G.J. (1998). JPred: a consensus secondary structure prediction server. *Bioinformatics* **14**(10): 892–893.

Currie, M.A., Merino, F., Skarina, T., Wong, A.H.Y., Singer, A., Brown, G., Savchenko, A., Caniuguir, A., Guixé, V., Yakunin, A.F., and Jia, Z. (2009). ADP-dependent 6-phosphofructokinase from *Pyrococcus horikoshii* OT3: structure determination and biochemical characterization of PH1645. *J. Biol. Chem.* **284**(34): 22664–22671.

Datta, R., Das, I., Sen, B., Chakraborty, A., Adak, S., Mandal, C., and Datta, A.K. (2005). Mutational analysis of the active-site residues crucial for catalytic activity of adenosine kinase from *Leishmania donovani*. *Biochem. J.* **387**(Pt 3): 591–600.

Dekker, J.P., Fodor, A., Aldrich, R.W., and Yellen, G. (2004). A perturbation-based method for calculating explicit likelihood of evolutionary co-variance in multiple sequence alignments. *Bioinformatics* **20**(10): 1565–1572.

DeLano, W. (2002). The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA. .

Dörr, C., Zaparty, M., Tjaden, B., Brinkmann, H., and Siebers, B. (2003). The hexokinase of the hyperthermophile *Thermoproteus tenax*. ATP-dependent hexokinases and ADP-dependent glucokinases, two alternatives for glucose phosphorylation in Archaea. *J. Biol. Chem.* **278**(21): 18744–18753.

Duan, J. and Nilsson, L. (2005). Thermal unfolding simulations of a multimeric protein–transition state and unfolding pathways. *Proteins* **59**(2): 170–182.

Emsley, P. and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**(12): 2126–2132.

Evans, P.R., Farrants, G.W., and Hudson, P.J. (1981). Phosphofructokinase: structure and control. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **293**(1063): 53–62.

Eyring, H. (1935). The activated complex in chemical reactions. *J. Chem. Phys.* **3**: 107–115.

Foloppe, N. and MacKerell Jr, A. (2000). All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* **21**(2): 86–104.

Fong, J.H., Geer, L.Y., Panchenko, A.R., and Bryant, S.H. (2007). Modeling the evolution of protein domain architectures using maximum parsimony. *J. Mol. Biol.* **366**(1): 307–315.

Freisleben, H.J., Blöcher, D., and Ring, K. (1992). Calorimetry of tetraether lipids from *Thermoplasma acidophilum*: incorporation of alamethicin, melittin, valinomycin, and nonactin. *Arch. Biochem. Biophys.* **294**(2): 418–426.

Fuhrmann, M., Hausherr, A., Ferbitz, L., Schödl, T., Heitzer, M., and Hegemann, P. (2004). Monitoring dynamic expression of nuclear genes in *Chlamydomonas reinhardtii* by using a synthetic luciferase reporter gene. *Plant. Mol. Biol.* **55**(6): 869–881.

Fulton, K.F., Main, E.R., Daggett, V., and Jackson, S.E. (1999). Mapping the interactions present in the transition state for unfolding/folding of FKBP12. *J. Mol. Biol.* **291**(2): 445–461.

Gasteiger, J. and Marsili, M. (1980). Iterative partial equalization of orbital electronegativity–a rapid access to atomic charges. *Tetrahedron* **36**(22): 3219–3228.

Gogarten, J.P., Doolittle, W.F., and Lawrence, J.G. (2002). Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* **19**(12): 2226–2238.

Gordon, J.C., Myers, J.B., Folta, T., Shoja, V., Heath, L.S., and Onufriev, A. (2005). H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res* **33**: 368–371.

Guda, C., Scheeff, E.D., Bourne, P.E., and Shindyalov, I.N. (2001). A new algorithm for the alignment of multiple protein structures using Monte Carlo optimization. *Pac Symp Biocomput* pp. 275–286.

Guex, N. and Peitsch, M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**(15): 2714–2723.

Guixé, V., Rodríguez, P.H., and Babul, J. (1998). Ligand-induced conformational transitions in *Escherichia coli* phosphofructokinase 2: evidence for an allosteric site for $MgATP^{2-}$. *Biochemistry* **37**(38): 13269–13275.

Guixé, V. and Merino, F. (2009). The ADP-dependent sugar kinase family: kinetic and evolutionary aspects. *IUBMB Life* **61**(7): 753–761.

Hansen, T. and Schönheit, P. (2001). Sequence, expression, and characterization of the first archaeal ATP-dependent 6-phosphofructokinase, a non-allosteric enzyme related to the phosphofructokinase-B sugar kinase family, from the hyperthermophilic crenarchaeote *Aeropyrum pernix*. *Arch. Microbiol.* **177**(1): 62–69.

Hansen, T., Arnfors, L., Ladenstein, R., and Schönheit, P. (2007). The phosphofructokinase-B (MJ0406) from *Methanocaldococcus jannaschii* represents a nucleoside kinase with a broad substrate specificity. *Extremophiles* **11**(1): 105–114.

Hansen, T., Musfeldt, M., and Schönheit, P. (2002). ATP-dependent 6-phosphofructokinase from the hyperthermophilic bacterium *Thermotoga maritima*: characterization of an extremely thermophilic, allosterically regulated enzyme. *Arch. Microbiol.* **177**(5): 401–409.

Hansen, T. and Schönheit, P. (2004). ADP-dependent 6-phosphofructokinase, an extremely thermophilic, non-allosteric enzyme from the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus* strain 7324. *Extremophiles* **8**(1): 29–35.

Harms, M.J. and Thornton, J.W. (2010). Analyzing protein structure and function using ancestral gene reconstruction. *Curr. Opin. Struct. Biol.* **20**(3): 360–366.

Hayward, S. and Lee, R.A. (2002). Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *J. Mol. Graph. Model* **21**(3): 181–183.

Hellinga, H.W. and Evans, P.R. (1987). Mutations in the active site of *Escherichia coli* phosphofructokinase. *Nature* **327**(6121): 437–439.

Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008). Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**(3): 435–447.

Holm, L. and Park, J. (2000). DaliLite workbench for protein structure comparison. *Bioinformatics* **16**(6): 566–567.

Hub, J.S. and de Groot, B.L. (2009). Detection of functional modes in protein dynamics. *PLoS Comput. Biol.* **5**(8): e1000480.

Huelsenbeck, J.P. and Ronquist, F. (2001). MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**(8): 754–755.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* **14**(1): 33–38.

Huson, D.H., Richter, D.C., Rausch, C., Dezulian, T., Franz, M., and Rupp, R. (2007). Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics* **8**: 460.

Ito, S., Fushinobu, S., Yoshioka, I., Koga, S., Matsuzawa, H., and Wakagi, T. (2001). Structural basis for the ADP-specificity of a novel glucokinase from a hyperthermophilic archaeon. *Structure* **9**(3): 205–214.

Ito, S., Fushinobu, S., Jeong, J.J., Yoshioka, I., Koga, S., Shoun, H., and Wakagi, T. (2003). Crystal structure of an ADP-dependent glucokinase from *Pyrococcus furiosus*: implications for a sugar-induced conformational change in ADP-dependent kinase. *J. Mol. Biol.* **331**(4): 871–883.

Jain, R., Rivera, M., and Lake, J. (1999). Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Nat. Acad. Sci. USA* **96**(7): 3801–3806.

Jeong, J.J., Fushinobu, S., Ito, S., Shoun, H., and Wakagi, T. (2003). Archaeal ADP-dependent phosphofructokinase: expression, purification, crystallization and preliminary crystallographic analysis. *Acta Crystallogr. D Biol. Crystallogr.* **59**(Pt 7): 1327–1329.

Jones, W., Leigh, J., Mayer, F., Woese, C., and Wolfe, R. (1983). *Methanococcus jannaschii* sp. nov., an extremely thermophilic methanogen from a submarine hydrothermal vent. *Arch. Microbiol.* **136**(4): 254–261.

Kaneshiro, S.M. and Clark, D.S. (1995). Pressure effects on the composition and thermal behavior of lipids from the deep-sea thermophile *Methanococcus jannaschii*. *J. Bacteriol.* **177**(13): 3668–3672.

Kengen, S.W., de Bok, F.A., van Loo, N.D., Dijkema, C., Stams, A.J., and de Vos, W.M. (1994). Evidence for the operation of a novel Embden-Meyerhof pathway that involves ADP-dependent kinases during sugar fermentation by *Pyrococcus furiosus*. *J Biol Chem* **269**(26): 17537–17541.

Kengen, S.W., Tuininga, J.E., de Bok, F.A., Stams, A.J., and de Vos, W.M. (1995). Purification and characterization of a novel ADP-dependent glucokinase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Biol. Chem.* **270**(51): 30453–30457.

Koga, S., Yoshioka, I., Sakuraba, H., Takahashi, M., Sakasegawa, S., Shimizu, S., and Ohshima, T. (2000). Biochemical characterization, cloning, and sequencing of ADP-dependent (AMP-forming) glucokinase from two hyperthermophilic archaea, *Pyrococcus furiosus* and *Thermococcus litoralis*. *J. Biochem.* **128**(6): 1079–1085.

Koga, Y., Nishihara, M., Morii, H., and Akagawa-Matsushita, M. (1993). Ether polar lipids of methanogenic bacteria: structures, comparative aspects, and biosyntheses. *Microbiol. Rev.* **57**(1): 164–182.

Kotlarz, D. and Buc, H. (1981). Regulatory properties of phosphofructokinase 2 from *Escherichia coli*. *Eur. J. Biochem.* **117**(3): 569–574.

Krissinel, E. and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**(3): 774–597.

Kuzmic, P. (1996). Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal. Biochem.* **237**(2): 260–273.

de La Fortelle, E. and Bricogne, G. (1997). SHARP: A Maximum-Likelihood Heavy-Atom Parameter Refinement Program for the MIR and MAD Methods. *Methods Enzymol.* **276**: 472–494.

Lange, O.F. and Grubmüller, H. (2008). Full correlation analysis of conformational protein dynamics. *Proteins* **70**(4): 1294–1312.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Cryst* **26**: 283–291.

Lazaridis, T. and Karplus, M. (1997). "New view" of protein folding reconciled with the old through multiple unfolding simulations. *Science* **278**(5345): 1928–1931.

Lazaridis, T., Lee, I., and Karplus, M. (1997). Dynamics and unfolding pathways of a hyperthermophilic and a mesophilic rubredoxin. *Protein Sci.* **6**(12): 2589–2605.

Li, A. and Daggett, V. (1994). Characterization of the transition state of protein unfolding by use of molecular dynamics: chymotrypsin inhibitor 2. *Proc. Natl. Acad. Sci. USA* **91**(22): 10430–10434.

Li, M., Kwok, F., Chang, W., Lau, C., Zhang, J., Lo, S.C.L., Jiang, T., and Liang, D. (2002). Crystal structure of brain pyridoxal kinase, a novel member of the ribokinase superfamily. *J. Biol. Chem.* **277**(48): 46385–46390.

Luke, K.A., Higgins, C.L., and Wittung-Stafshede, P. (2007). Thermodynamic stability and folding of proteins from hyperthermophilic organisms. *FEBS J* **274**(16): 4023–4033.

Lüthy, R., Bowie, J.U., and Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature* **356**(6364): 83–85.

MacKerell Jr, A. and Banavali, N. (2000). All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comput. Chem.* **21**(2): 105–120.

MacKerell Jr, A., Bashford, D., Bellott, M., Dunbrack Jr, R., Evanseck, J., Field, M., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiherand, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998). All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **102**(18): 3586–3616.

Maj, M.C., Singh, B., and Gupta, R.S. (2000). Structure-activity studies on mammalian adenosine kinase. *Biochem. Biophys. Res. Commun.* **275**(2): 386–393.

Maj, M.C., Singh, B., and Gupta, R.S. (2002). Pentavalent ions dependency is a conserved property of adenosine kinase from diverse sources: identification of a novel motif implicated in phosphate and magnesium ion binding and substrate inhibition. *Biochemistry* **41**(12): 4059–4069.

Mamonova, T.B., Glyakina, A.V., Kurnikova, M.G., and Galzitskaya, O.V. (2010). Flexibility and mobility in mesophilic and thermophilic homologous proteins from molecular dynamics and FoldUnfold method. *J. Bioinform. Comput. Biol.* **8**(3): 377–394.

Margulis, L. (1996). Archaeal-eubacterial mergers in the origin of Eukarya: phylogenetic classification of life. *Proc. Natl. Acad. Sci. USA* **93**(3): 1071–1076.

Mathews, I.I., Erion, M.D., and Ealick, S.E. (1998). Structure of human adenosine kinase at 1.5 Å resolution. *Biochemistry* **37**(45): 15607–15620.

McCoy, A., Grosse-Kunstleve, R., Adams, P., Winn, M., Storoni, L., and Read, R. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* **40**(4): 658–674.

McInerney, J.O. (1998). GCUA: general codon usage analysis. *Bioinformatics* **14**(4): 372–373.

McRee, D. (1999). XtalView/Xfit–A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**(2-3): 156–165.

Merino, Rivas-Pardo, J.A., Caniuguir, A., García, I., and Guixé, V. (In editorial revision). Catalytic and regulatory roles of divalent metal cations on the phosphoryl-transfer mechanism of ADP-dependent sugar kinases from hyperthermophilic archaea. *Biochimie* .

Merino, F. and Guixé, V. (2008). Specificity evolution of the ADP-dependent sugar kinase family: in silico studies of the glucokinase/phosphofructokinase bifunctional enzyme from *Methanocaldococcus jannaschii*. *FEBS J.* **275**(16): 4033–4044.

Meyer, C., Schmid, R., Scriba, P.C., and Wehling, M. (1996). Purification and partial sequencing of high-affinity progesterone-binding site(s) from porcine liver membranes. *Eur. J. Biochem.* **239**(3): 726–731.

Mihalek, I., Res, I., and Lichtarge, O. (2004). A family of evolution-entropy hybrid methods for ranking protein residues by importance. *J. Mol. Biol.* **336**(5): 1265–1282.

Morris, G., Goodsell, D., Halliday, R., Huey, R., Hart, W., Belew, R., and Olson, A. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. comput. chem.* **19**(14): 1639–1662.

Mukund, S. and Adams, M.W. (1991). The novel tungsten-iron-sulfur protein of the hyperthermophilic archaebacterium, *Pyrococcus furiosus*, is an aldehyde ferredoxin oxidoreductase. Evidence for its participation in a unique glycolytic pathway. *J Biol Chem* **266**(22): 14208–14216.

Mukund, S. and Adams, M.W. (1995). Glyceraldehyde-3-phosphate ferredoxin oxidoreductase, a novel tungsten-containing enzyme with a potential glycolytic role in the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Biol. Chem.* **270**(15): 8389–8392.

Mulichak, A.M., Wilson, J.E., Padmanabhan, K., and Garavito, R.M. (1998). The structure of mammalian hexokinase-1. *Nat. Struct. Biol.* **5**(7): 555–560.

Nicholas, K., Nicholas, H., and Deerfield, D. (1997). GeneDoc: analysis and visualization of genetic variation. *Embnew News* **4**: 14.

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

van der Oost, J., Schut, G., Kengen, S.W., Hagen, W.R., Thomm, M., and de Vos, W.M. (1998). The ferredoxin-dependent conversion of glyceraldehyde-3-phosphate in the hyperthermophilic archaeon *Pyrococcus furiosus* represents a novel site of glycolytic regulation. *J. Biol. Chem.* **273**(43): 28149–28154.

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

Parducci, R.E., Cabrera, R., Baez, M., and Guixé, V. (2006). Evidence for a catalytic Mg^{2+} ion and effect of phosphate on the activity of *Escherichia coli* phosphofructokinase-2: regulatory properties of a ribokinase family member. *Biochemistry* **45**(30): 9291–9299.

Peng, Z.Y. and Mansour, T.E. (1992). Purification and properties of a pyrophosphate-dependent phosphofructokinase from *Toxoplasma gondii*. *Mol. Biochem. Parasitol.* **54**(2): 223–230.

Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kalé, L., and Schulten, K. (2005). Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**(16): 1781–1802.

Pigache, A., Cieplak, P., and Dupradeau, F.Y. (2004). Automatic and highly reproducible RESP and ESP charge derivation: application to the development of programs RED and X RED. *227th ACS National Meeting, Anaheim, CA, USA* .

Rivas-Pardo, J.A., Caniuguir, A., Wilson, C.A.M., Babul, J., and Guixé, V. (2011). Divalent metal cation requirements of phosphofructokinase-2 from *E. coli*. Evidence for a high affinity binding site for Mn^{2+}. *Arch. Biochem. Biophys.* **505**(1): 60–66.

Ronimus, R.S., Koning, J., and Morgan, H.W. (1999). Purification and characterization of an ADP-dependent phosphofructokinase from *Thermococcus zilligii*. *Extremophiles* **3**(2): 121–129.

Ronimus, R.S. and Morgan, H.W. (2004). Cloning and biochemical characterization of a novel mouse ADP-dependent glucokinase. *Biochem. Biophys. Res. Commun.* **315**(3): 652–658.

Ronquist, F. and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**(12): 1572–1574.

de Rosa, M., Gambacorta, A., and Bu'lock, J.D. (1975). Extremely thermophilic acidophilic bacteria convergent with *Sulfolobus acidocaldarius*. *J. Gen. Microbiol.* **86**(1): 156–164.

Safo, M.K., Musayev, F.N., Hunt, S., di Salvo, M.L., Scarsdale, N., and Schirch, V. (2004). Crystal structure of the PdxY Protein from *Escherichia coli*. *J. Bacteriol.* **186**(23): 8074–8082.

Sakuraba, H., Goda, S., and Ohshima, T. (2004). Unique sugar metabolism and novel enzymes of hyperthermophilic archaea. *Chem. Rec.* **3**(5): 281–287.

Sakuraba, H., Yoshioka, I., Koga, S., Takahashi, M., Kitahama, Y., Satomura, T., Kawakami, R., and Ohshima, T. (2002). ADP-dependent glucokinase/phosphofructokinase, a novel bifunctional enzyme from the hyperthermophilic archaeon *Methanococcus jannaschii*. *J. Biol. Chem.* **277**(15): 12495–12498.

Sali, A. and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**(3): 779–815.

Sapra, R., Bagramyan, K., and Adams, M.W.W. (2003). A simple energy-conserving system: proton reduction coupled to proton translocation. *Proc. Natl. Acad. Sci. USA* **100**(13): 7545–7550.

Schirmer, T. and Evans, P. (1990). Structural basis of the allosteric behaviour of phosphofructokinase. *Nature* **343**: 140–145.

Schumacher, M.A., Scott, D.M., Mathews, I.I., Ealick, S.E., Roos, D.S., Ullman, B., and Brennan, R.G. (2000). Crystal structures of *Toxoplasma gondii* adenosine kinase reveal a novel catalytic mechanism and prodrug binding. *J. Mol. Biol.* **298**(5): 875–893.

Schut, G.J., Brehm, S.D., Datta, S., and Adams, M.W.W. (2003). Whole-genome DNA microarray analysis of a hyperthermophile and an archaeon: *Pyrococcus furiosus* grown on carbohydrates or peptides. *J. Bacteriol.* **185**(13): 3935–3947.

Sigrell, J.A., Cameron, A.D., Jones, T.A., and Mowbray, S.L. (1998). Structure of *Escherichia coli* ribokinase in complex with ribose and dinucleotide determined to 1.8 Å resolution: insights into a new family of kinase structures. *Structure* **6**(2): 183–193.

Sippl, M.J. (1993). Recognition of errors in three-dimensional structures of proteins. *Proteins* **17**(4): 355–362.

van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., and Berendsen, H.J.C. (2005). GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26**(16): 1701–1718.

Sprott, G.D., Meloche, M., and Richards, J.C. (1991). Proportions of diether, macrocyclic diether, and tetraether lipids in *Methanococcus jannaschii* grown at different temperatures. *J. Bacteriol.* **173**(12): 3907–3910.

St Charles, R., Harrison, R.W., Bell, G.I., Pilkis, S.J., and Weber, I.T. (1994). Molecular model of human beta-cell glucokinase built by analogy to the crystal structure of yeast hexokinase B. *Diabetes* **43**(6): 784–791.

Tabtiang, R.K., Cezairliyan, B.O., Grant, R.A., Cochrane, J.C., and Sauer, R.T. (2005). Consolidating critical binding determinants by noncyclic rearrangement of protein secondary structure. *Proc. Natl. Acad. Sci. USA* **102**(7): 2305–2309.

Terwilliger, T.C. (2000). Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* **56**(8): 965–972.

Terwilliger, T. and Berendzen, J. (1999). Automated structure solution for MIR and MAD. *Acta Crystallogr. D Biol. Crystallogr.* **55**: 849–861.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997). The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**(24): 4876–4882.

Todd, M.J. and Gomez, J. (2001). Enzyme kinetics determined using calorimetry: a general assay for enzyme activity? *Anal. Biochem.* **296**(2): 179–187.

Torres, J.C., Guixé, V., and Babul, J. (1997). A mutant phosphofructokinase produces a futile cycle during gluconeogenesis in *Escherichia coli*. *Biochem. J.* **327**: 675–684.

Tsuge, H., Sakuraba, H., Kobe, T., Kujime, A., Katunuma, N., and Ohshima, T. (2002). Crystal structure of the ADP-dependent glucokinase from *Pyrococcus horikoshii* at 2.0-Å resolution: a large conformational change in ADP-dependent glucokinase. *Protein Sci.* **11**(10): 2456–2463.

Tuininga, J.E., Verhees, C.H., van der Oost, J., Kengen, S.W., Stams, A.J., and de Vos, W.M. (1999). Molecular and biochemical characterization of the ADP-dependent phosphofructokinase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Biol. Chem.* **274**(30): 21023–21028.

Uversky, V.N. (1993). Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry* **32**(48): 13288–13298.

Verhees, C.H., Kengen, S.W.M., Tuininga, J.E., Schut, G.J., Adams, M.W.W., De Vos, W.M., and Van Der Oost, J. (2003). The unique features of glycolytic pathways in Archaea. *Biochem. J.* **375**(2): 231–246.

Verhees, C.H., Tuininga, J.E., Kengen, S.W., Stams, A.J., van der Oost, J., and de Vos, W.M. (2001). ADP-dependent phosphofructokinases in mesophilic and thermophilic methanogenic archaea. *J. Bacteriol.* **183**(24): 7145–7153.

Verhees, C.H., Koot, D.G.M., Ettema, T.J.G., Dijkema, C., de Vos, W.M., and van der Oost, J. (2002). Biochemical adaptations of two sugar kinases from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Biochem. J.* **366**(Pt 1): 121–127.

Wallner, B. and Elofsson, A. (2003). Can correct protein models be identified? *Protein Sci* **12**(5): 1073–1086.

Woese, C.R. and Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. USA* **74**(11): 5088–5090.

Woese, C.R., Kandler, O., and Wheelis, M.L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA* **87**(12): 4576–4579.

Wolf-Watz, M., Thai, V., Henzler-Wildman, K., Hadjipavlou, G., Eisenmesser, E.Z., and Kern, D. (2004). Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nat Struct Mol Biol* **11**(10): 945–949.

Wu, L.F., Reizer, A., Reizer, J., Cai, B., Tomich, J.M., and Saier, M.H. (1991). Nucleotide sequence of the *Rhodobacter capsulatus* fruK gene, which encodes fructose-1-phosphate kinase: evidence for a kinase superfamily including both phosphofructokinases of *Escherichia coli*. *J. Bacteriol.* **173**(10): 3117–3127.

Yamauchi, K., Doi, K., Yoshida, Y., and Kinoshita, M. (1993). Archaebacterial lipids: highly proton-impermeable membranes from 1,2-diphytanyl-sn-glycero-3-phosphocholine. *Biochim. Biophys. Acta* **1146**(2): 178–182.

Zhang, R.G., Skarina, T., Katz, J.E., Beasley, S., Khachatryan, A., Vyas, S., Arrowsmith, C.H., Clarke, S., Edwards, A., Joachimiak, A., and Savchenko, A. (2001). Structure of *Thermotoga maritima* stationary phase survival protein SurE: a novel acid phosphatase. *Structure* **9**(11): 1095–1106.

Zhang, Y., Dougherty, M., Downs, D.M., and Ealick, S.E. (2004). Crystal structure of an aminoimidazole riboside kinase from *Salmonella enterica*: implications for the evolution of the ribokinase superfamily. *Structure* **12**(10): 1809–1821.

Zheng, R.L. and Kemp, R.G. (1992). The mechanism of ATP inhibition of wild type and mutant phosphofructo-1-kinase from *Escherichia coli*. *J. Biol. Chem.* **267**(33): 23640–23645.

Appendices

Appendix A

On the measurement of enzyme kinetics using isothermal titration calorimetry

Usually, the isothermal titration (micro)calorimeter (ITC) is used for the determination of the association constant of two molecular species through the heat change associated with the binding event. Given that the ITC can measure directly the equilibrium heat change related to the binding event, and also the association constant can be calculated from the dependence of the heat exchanged with the molar ratio between the species, it is possible also to obtain the entropy associated with the process from a single experiment. This makes ITC one of the most powerful techniques to study the thermodynamics of molecular binding. Unfortunately, the amount of protein (and ligand) needed to obtain a good measure is limited by the dissociation constant of the complex. As a rule of thumb, a concentration between 10 and 100 times K_d should give a very good estimate of the thermodynamic quantities governing the process. Considering protein concentrations (and total masses!) that are realistically attainable in the laboratory, the above limit restraints the use of this technique to complexes with dissociation constants below the 10^{-5} order of magnitude.

While the ITC is generally used for binding experiments, its use to measure the heat associated with an enzymatic reaction has been recognized in the past (for a detailed

study see Todd and Gomez (2001))

Unlike the binding case, it is possible to measure any reaction (it has not to be an enzyme catalyzed reaction) that has a detectable change of enthalpy associated, which make the ITC a very powerful equipment for kinetic studies as well.

A.1 Single injection method

For an enzymatic reaction, the change of enthalpy associated to the process can be measured by a microcalorimeter as the heat exchanged between the equipment and the reaction cell. Here, the reaction velocity is proportional to the change of power needed to maintain the isothermal condition.

Specifically, the power associated with the enzymatic reaction can be related to the reaction velocity through the equation

$$\frac{dQ}{dt} = \frac{d[P]}{dt} \cdot V \cdot \Delta H_{app} \tag{A.1}$$

where $[P]$ is the product concentration, V is the total cell volume and ΔH_{app} is the apparent enthalpy change associated with the reaction, which has to be determined empirically.

The apparent molar reaction enthalpy can be determined by doing a single injection of substrate (or enzyme) to initiate the reaction and then following the process until completion (A.1). There, it can be calculated using the relation

$$\Delta H_{app} = \frac{1}{[S]_T \cdot V} \int_{t=0}^{t=\infty} \frac{dQ}{dt} dt \tag{A.2}$$

where $[S]_T$ is the initial substrate concentration and V is the total volume of the reaction cell.

Additionally, given that the power and the enthalpy are now known, it is possible to calculate the substrate concentration for every time and the reaction velocity associated

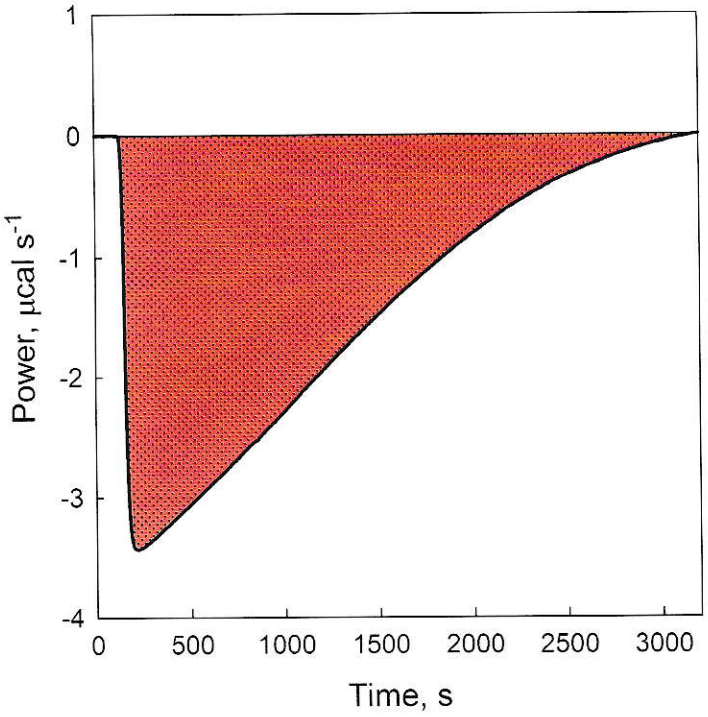


Figure A.1: Measurement of the apparent enthalpy change associated with an enzymatic reaction. The red area represents the total heat released by the process.

with this concentration, which makes it possible to obtain a saturation curve from this single experiment (Figure A.2C). Also, the amount of product produced at every time can be calculated as the fraction of the total heat exchanged by

$$[P](t) = [S]_T \frac{\int_0^t \frac{dQ}{dt} dt}{\int_0^\infty \frac{dQ}{dt} dt} \tag{A.3}$$

This makes it possible to obtain also the progress curve associated with the reaction (Figure A.2B).

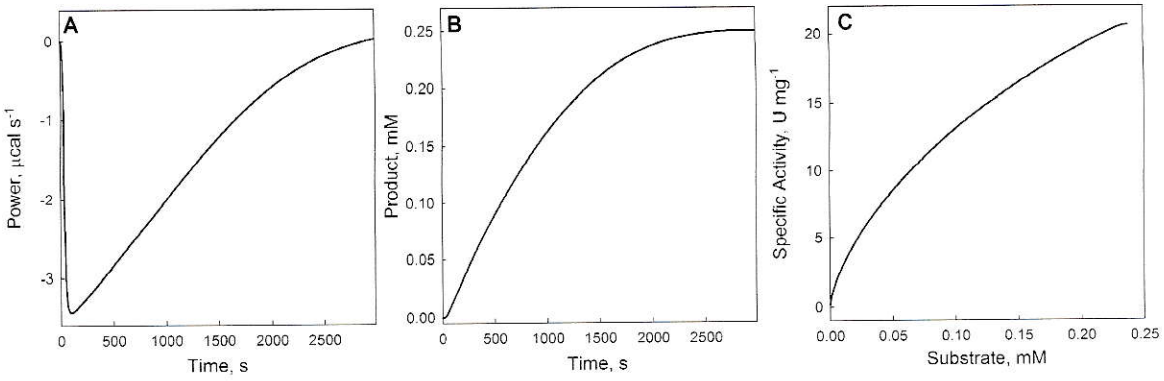


Figure A.2: Single injection method for enzyme kinetics. **A.** Typical isotherm obtained in a single injection ITC experiment. **B.** Progress curve obtained from (**A**) by using equation A.3. **C.** Saturation curve obtained from (**A**) by using equation A.1.

A.2 Multiple injections method

There is an alternative method to obtain a saturation curve by calorimetric means. Here, small amounts of substrate are injected in the reaction cell. After the initial power change due to the heat associated with the dilution the curve should reach a new baseline. This power change can be related directly to the reaction velocity associated with that substrate concentration through the equation A.1 (Figure A.3). The power change should be measured as quickly as possible in order to maintain the initial velocity condition. Repeated injections can be used to obtain the full saturation curve (Figure A.4). It is important to point out that for this method, the ΔH_{app} is necessary to process the data. So, in order to obtain the final result, a single injection experiment should also be performed.

While both methods are quite similar in the data treatment, they differ in how the product is accumulated through the reaction. In the first method, the highest product concentrations are present when the reaction velocities associated with the lowest substrate concentrations are being measured. On the hand, in the second method the highest product concentrations are only accumulated when the highest substrate

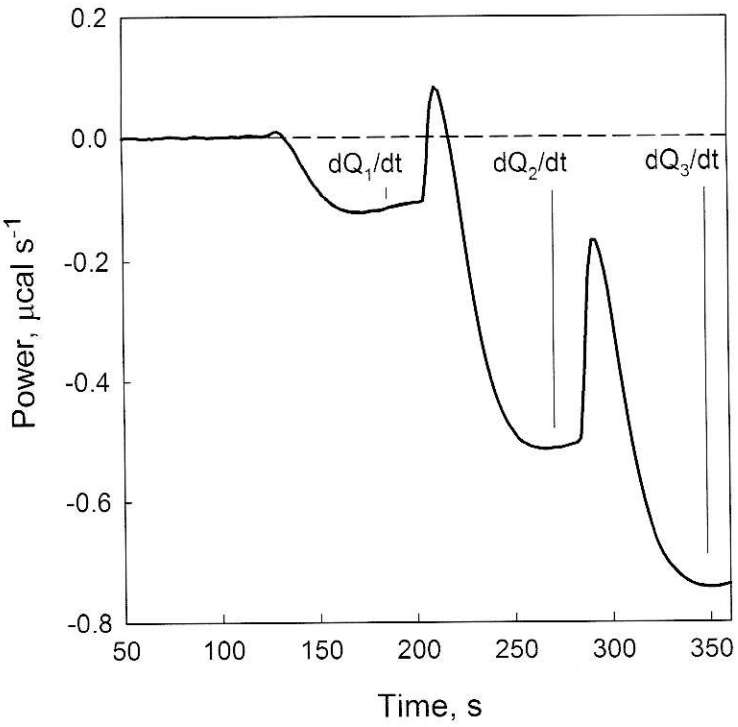


Figure A.3: Initial velocity measurement of enzyme kinetics. Three consecutive injections are shown as well as their associated power change.

concentrations are present. In this way, the first method is only recommended for enzymes with K_M values smaller that 10 μM.

As opposed to binding experiments where protein concentrations around ten times K_d are needed, kinetic assays need enzyme concentration that usually are around the nanomolar range.

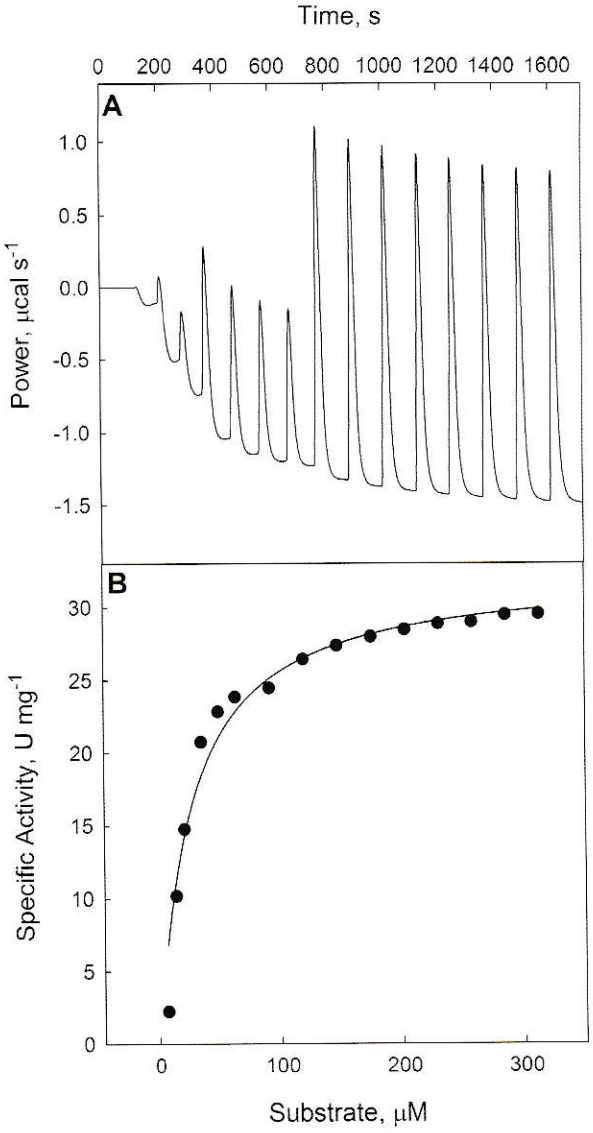


Figure A.4: Multiple injections method for enzyme kinetics. **A.** Typical isotherm obtained in a multiple injection experiment. Fifteen injections were performed, three 0.5 μL injections, three 1 μL injections, and nine 2 μL injections. **B.** Saturation curve obtained from (**A**) by using equation A.1.