

UCH-FC  
Doc-B.HCM  
C 355  
C.1

FACULTAD DE CIENCIAS

UNIVERSIDAD DE CHILE



DEPARTAMENTO DE BIOLOGÍA



**Resurrección de enzimas ancestrales: historia evolutiva de la  
especificidad por sustratos en la superfamilia Riboquinasa**

VÍCTOR HUGO CASTRO FERNÁNDEZ

Santiago, Chile

Enero 2015

**RESURRECCIÓN DE ENZIMAS ANCESTRALES: HISTORIA  
EVOLUTIVA DE LA ESPECIFICIDAD POR SUSTRATOS EN LA  
SUPERFAMILIA RIBOQUINASA**

Tesis entregada a la Universidad de Chile en cumplimiento parcial de los  
requisitos para optar al grado de

**Doctor en Ciencias con Mención en Biología Molecular, Celular y**

**Neurociencias**

Facultad de Ciencias

por

**VÍCTOR HUGO CASTRO FERNÁNDEZ**

Enero, 2015

Directora de Tesis

Dra. Victoria Guixé Leguía

**FACULTAD DE CIENCIAS**  
**UNIVERSIDAD DE CHILE**  
**INFORME DE APROBACIÓN**  
**TESIS DE DOCTORADO**

Se informa a la Escuela de Postgrado de la Facultad de Ciencias que la Tesis de Doctorado presentada por el candidato

**VÍCTOR HUGO CASTRO FERNÁNDEZ**

Ha sido aprobada por la comisión de Evaluación de la Tesis como requisito para optar al grado de Doctor en Ciencias con mención en Biología Molecular, Celular y Neurociencias, en el examen de Defensa Privada de Tesis rendido el día 27 de enero de 2015.

**Directora de Tesis:**

Dra. Victoria Guixé Leguía

.....  


**Comisión de Evaluación de la Tesis:**

Dr. Carlos Jerez (Presidente)

.....  


Dra. Marta Bunster

.....

Dr. Alexander Vargas

.....  


Dr. Marco Mendez

.....



Para  
Ángela y Fernanda



## BIOGRAFÍA



Nací en abril de 1986 en Concepción, aunque mis padres vivían en Curanilahue, un pueblo minero a dos horas al sur de la capital de la octava región. Cuando tenía 3 años mi padre obtuvo un trabajo de guardabosques en la cordillera de Nahuelbuta en un fundo denominado Pilin-Pilli, donde crecí rodeado de árboles nativos, araucarias milenarias y contrastantes plantaciones forestales de Pino y Eucaliptus. A los 6 años de edad ingresé a la escuela F-832 de Cañete, apodada por los cañetinos como “la escuela tres”, ahí estuve hasta octavo básico y creo que fue el nicho que forjó mi interés por las matemáticas, intereses impulsados por Jaime Pascal, profesor que estuvo guiándome de 1° a 8° básico. Posteriormente volví a Curanilahue para realizar la enseñanza media en el Liceo Polivalente Mariano Latorres B-51, lugar donde conocí a grandes personas, amigos y a Ángela, la mujer que sería mi esposa. Fue en este liceo público donde formaron mi espíritu científico, mi gran pasión por la Química y mi interés por la Biología y la Física, intereses avivados por profesores como Nelson González y Victor Iturrieta que de manera desinteresada compartían su conocimiento y su pasión por la ciencia fuera de las salas de clases. En el año 2004 tras rendir la primera versión de la PSU, ingresé a la carrera de Bioquímica de la Universidad de Concepción para cumplir un sueño, ser científico, en esa casa de estudios también encontré grandes amigos y grandes científicos guías como Elena A Uribe y otros que incitaron mi interés por las proteínas, en particular por las enzimas. Este interés me llevó a Santiago y en el año 2010 ingresé al doctorado BMCyN de la Universidad de Chile y me integré al grupo de la Dra. Victoria Guixé donde encontré la pasión por la evolución molecular. El 2012 nació Fernanda, por lo que ahora compatibilizo mi pasión por la ciencia con la pasión por mi familia.

## AGRADECIMIENTOS

Quisiera agradecer a la directora de esta tesis, la Dra. Victoria Guixé por su confianza y apoyo en los proyectos que hemos emprendido y por su guía no solo en el quehacer científico. También quiero agradecer al Dr. Ricardo Cabrera por las discusiones científicas y sobre la vida de un científico, al Dr. Jorge Babul por su apoyo para permitirme finalizar esta tesis y darme la posibilidad de seguir haciendo lo que me apasiona y a la Dra. Ana Preller por las conversaciones de la contingencia nacional durante los almuerzo en el laboratorio.

Agradezco a todos los miembros del Laboratorio de Bioquímica y Biología Molecular, en particular al grupo de la Dra. Guixé, los que actualmente están: Alejandra H, Diego Q, Ricardo Z, María José A, Cesar R, Evelin G, Felipe G, Macarena y los que estuvieron en un comienzo o durante este trabajo: Felipe M, Andrés R, Felipe B y por supuesto a los demás miembros del laboratorios que están o ya emprendieron otros rumbos: Pablo V, Rodrigo M, Pablo M, Matías F, Gabriel V, Exequiel, Alonso y Luis M personas que han apoyado de manera directa experimentalmente o con su amistad y compañía el desarrollo de este trabajo.

Parte de este trabajo no habría sido posible sin el apoyo del Dr Richard Garratt del Instituto de Física de São Carlos (Universidad de São Paulo, Brasil) quien junto a su grupo de cristalografía y al grupo de Biofísica Molecular permitieron que las estadias en Brasil fueran exitosas experimentalmente y muy agradables.

Por supuesto agradecer a mi familia quienes me han apoyado incondicionalmente y me han permitido dedicarme a esta pasión.

## PUBLICACIONES

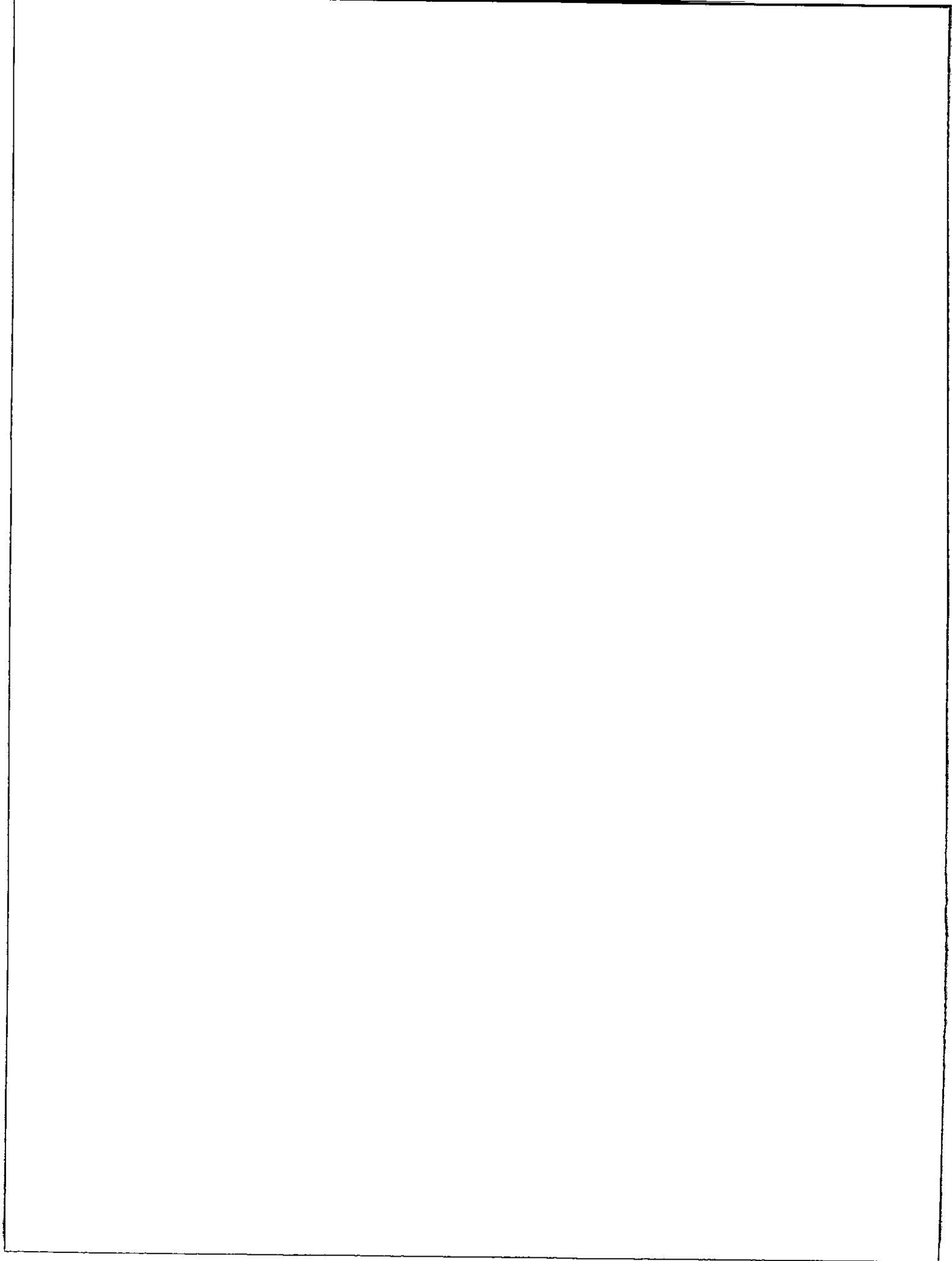
- Rivas-Pardo JA, Herrera-Morande A, **Castro-Fernandez V**, Fernandez FJ, Vega MC and Guixe V (2013). Crystal Structure, SAXS and Kinetic Mechanism of Hyperthermophilic ADP-Dependent Glucokinase from *Thermococcus litoralis* Reveal a Conserved Mechanism for Catalysis. PLoS ONE 8, e66687.
- **Castro-Fernandez V.**, Bravo-Moraga F., Herrera-Morande A. and Guixe V. (2014) Bifunctional ADP-dependent Phosphofructokinase/Glucokinase activity in the order *Methanococcales*: Biochemical characterization of the mesophilic enzyme from *Methanococcus maripaludis*. FEBS J. 281(8):2017-29.
- **Castro-Fernandez V.**, Bravo-Moraga F., Ramirez-Sarmiento C. A. and Guixe V. (2014) Emergence of pyridoxal phosphorylation through a promiscuous ancestor during the evolution of hydroxymethyl pyrimidine kinases, FEBS Letters. 588(17), 3068-3073.

## PREMIOS

- **Premio FERMELO SBBMCH** para la mejor charla de incorporación a la Sociedad Chilena de Bioquímica y Biología Molecular. Octubre 2012.
- **Medalla de Plata en conferencias Zing** (*Silver Poster Presentation Award*). Enzymes, Coenzymes and Methabolic Pathways Conferences. ZING Conference, Xcaret Mexico. Noviembre 2013.

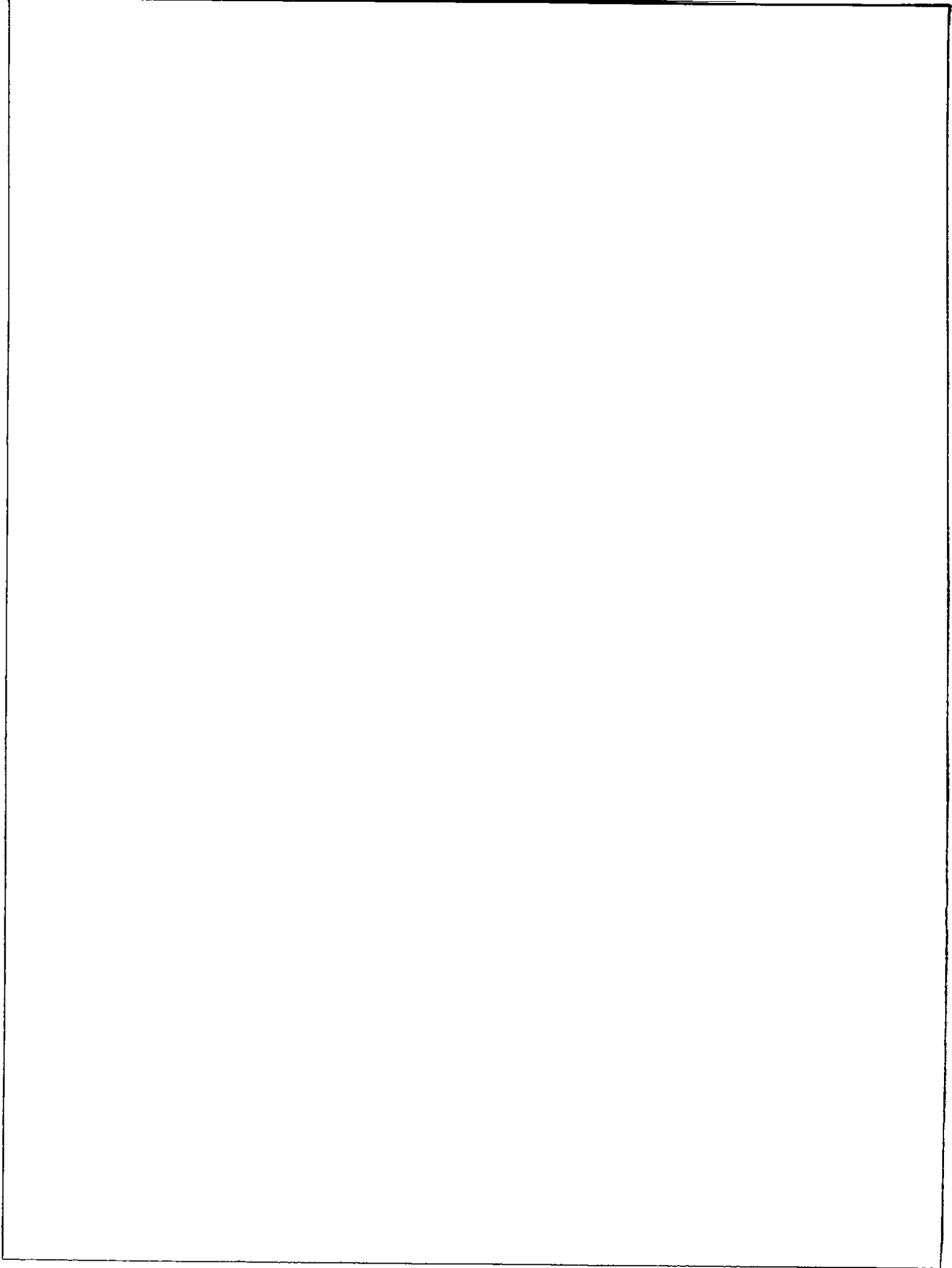
## **FINANCIAMIENTO**

- Proyecto FONDECYT 1110137.
- Beca para estudios de Doctorado en Chile, CONICYT 2010-2014.
- Proyecto de apoyo a la realización de tesis doctoral 24121448, CONICYT.
- Beca AMSUD-Pasteur, Programa sudamericano de intercambio regional para investigadores jóvenes, 2012.
- Beca de invitación de especialistas extranjeros, MESESUP UCH 0713, 2012.
- Beca ASMB/Prolabs, 2014 Promoting Research Opportunities for Latin American Biochemists.



## ÍNDICE DE CONTENIDOS

BIOGRAFÍA .....	v
AGRADECIMIENTOS .....	vi
PUBLICACIONES.....	vii
FINANCIAMIENTO.....	viii
ÍNDICE DE CONTENIDOS .....	ix
ÍNDICE DE TABLAS .....	xiv
ÍNDICE DE FIGURAS.....	xv
LISTA DE ABREVIATURAS .....	xvii
RESUMEN.....	xix
ABSTRACT .....	xxi
1 INTRODUCCIÓN GENERAL.....	1
1.1 Generalidades sobre evolución.....	1
1.2 Evolución de Proteínas.....	4
1.3 Evolución de enzimas .....	5
1.4 Reconstrucción de secuencias ancestrales (RSA).....	8
1.5 Miembros de la Superfamilia Riboquinasa como modelos para el estudio de la evolución de la especificidad por sustratos.....	11
2 MATERIALES Y MÉTODOS .....	16
2.1 Materiales.....	16
2.1.1 Vectores .....	16
2.1.2 Cepas.....	16
2.1.3 Medios de cultivo .....	17
2.1.4 Soluciones generales.....	17



2.1.5	Reactivos químicos generales .....	17
2.2	Métodos.....	19
2.2.1	Electroforesis en gel de agarosa.....	19
2.2.2	Electroforesis de proteínas en geles de poliacrilamida.....	19
2.2.3	Aislamiento de plásmidos.....	20
2.2.4	Purificación de fragmentos de ADN mediante geles de agarosa.....	20
2.2.5	Mutagénesis sitio dirigida.....	20
2.2.6	Cuantificación de proteínas.....	20
2.2.7	Reconstrucción de la filogenia e inferencia de las secuencias de enzimas ancestrales .....	21
2.2.8	Reconstrucción de estados ancestrales para rasgos discretos .....	22
2.2.9	Síntesis, expresión y purificación de proteínas ancestrales .....	23
2.2.10	Determinación de la actividad enzimática .....	24
2.2.11	Determinación de parámetros cinéticos: $k_{cat}$ , $K_m$ y $V_{max}$ .....	26
2.2.12	Dicroísmo circular y estabilidad térmica.....	27
2.2.13	Modelamiento molecular, ensayos de docking y dinámica molecular.....	27
2.2.14	Cristalografía de proteínas .....	29
3	CAPÍTULO UNO	
	Historia evolutiva del uso de sustratos en la Familia de quinasas de azúcar dependientes de ADP de arqueas .....	30
3.1	Introducción.....	31
3.2	Hipótesis.....	35
3.3	Objetivo.....	35
3.4	Resultados.....	36
3.4.1	Uso de sustratos en las enzimas actuales; <i>TIGK</i> , <i>PhPFK</i> , y <i>MjPFK/GK</i> ..	36



3.4.2	Reconstrucción filogenética por inferencia bayesiana y máxima verosimilitud .....	38
3.4.3	Reconstrucción de secuencias ancestrales de la familia de quinasas dependientes de ADP .....	42
3.4.4	Análisis de la robustez de las reconstrucciones de las secuencias ancestrales .....	44
3.4.5	Expresión y purificación de las enzimas ancestrales; ancGK/PFK, ancGK y ancPFK-MT .....	48
3.4.6	Caracterización cinética de las enzimas ancestrales; ancGK/PFK, ancGK y ancPFK-MT .....	48
3.4.7	Inferencia de las secuencias de los ancestros ancPFK-T y ancPFK-M .....	50
3.4.8	Resurrección experimental de ancPFK-T y ancPFK-M .....	50
3.4.9	Modelamiento y dinámica molecular de las interacciones entre proteína y ligando en los ancestros PFKs .....	52
3.4.10	Reconstrucción de estados ancestrales discretos para el uso de sustratos .....	56
3.4.11	Robustez de la reconstrucción de estados ancestrales discretos .....	60
3.4.12	Recreando la evolución, mutante E72A de ancPFK-MT .....	65
3.5	Discusión .....	67
3.6	Conclusiones .....	72
4	<b>CAPÍTULO DOS</b>	
	Caracterización experimental de la actividad bifuncional en fosfofructoquinasas dependientes de ADP provenientes de <i>Methanococcales</i> y <i>Methanosarcinales</i> .....	73
4.1	Introducción .....	74
4.2	Hipótesis .....	75
4.3	Objetivo .....	75

4.4	Actividad bifuncional fosfofructoquinasa/glucoquinasa en el orden <i>Methanococcales</i> : Caracterización bioquímica de la enzima mesófila de <i>Methanococcus maripaludis</i> .....	76
4.5	Análisis del uso de glucosa y fructosa-6P como sustratos en la enzima psicrófila de <i>Methanococcoides burtonii</i> .....	94
4.6	Discusión.....	96
4.7	Conclusiones.....	98
5	<b>CAPÍTULO TRES</b>	
	Estructura, estabilidad térmica y cambios conformacionales durante la evolución de la familia de quinasas de azúcar dependientes de ADP en arqueas.....	99
5.1	Introducción.....	100
5.2	Hipótesis.....	106
5.3	Objetivos .....	106
5.4	Resultados.....	107
5.4.1	Análisis del contenido de estructura secundaria a través de la evolución en la familia de quinasas dependientes de ADP mediante dicroísmo circular .....	107
5.4.2	Análisis de la estabilidad térmica de las proteínas a través de la evolución de la familia.....	111
5.4.3	Conservación de las interacciones involucradas en la transición abierta/cerrada en la glucoquinasa dependiente de ADP de <i>Thermococcus litoralis</i> en la familia de quinasas ADP dependiente .....	114
5.4.4	Cambios conformacionales evidenciados por determinación de la estructura de la fosfofructoquinasa dependiente de ADP de <i>Pyrococcus horikoshii</i> .....	130
5.5	Discusión.....	136

5.6 Conclusiones.....	141
6 CAPÍTULO CUATRO	
6.1 Aparición de la fosforilación de piridoxal durante la evolución de las quinasas de hidroximetil pirimidina a través de un ancestro promiscuo.....	143
7 DISCUSIÓN GENERAL.....	156
8 CONCLUSIONES GENERALES .....	163
REFERENCIAS .....	165

## ÍNDICE DE TABLAS

<b>Tabla 1.</b> Constantes cinéticas para las proteínas actuales. ....	37
<b>Tabla 2.</b> Probabilidades posteriores según el método de reconstrucción de ancGK/PFK .....	47
<b>Tabla 3.</b> Probabilidades posteriores de residuos del sitio activo de ancGK/PFK de acuerdo al método de reconstrucción .....	47
<b>Tabla 4.</b> Constantes cinéticas de las proteínas ancestrales; ancGK/PFK, ancGK y ancPFK-MT .....	49
<b>Tabla 5.</b> Constantes cinéticas para ancPFK-T y ancPFK-M a 30°C. ....	51
<b>Tabla 6.</b> Probabilidades para el uso de glucosa y fructosa-6P obtenidas por inferencia bayesiana .....	64
<b>Tabla 7.</b> Constantes cinéticas para ancPFK-MT E72A a 30°C. ....	66
<b>Tabla 8.</b> Constantes cinéticas para <i>Mb</i> PFK a 25°C. ....	95
<b>Tabla 9.</b> Porcentajes de estructura secundaria a partir de la estructura cristalográfica y espectros de DC .....	109
<b>Tabla 10.</b> Comparación de porcentajes de estructuras secundarias calculadas con CDSSTR.....	110
<b>Tabla 11.</b> Información cristalográfica y estadística del refinamiento para <i>Ph</i> PFK.....	134

## ÍNDICE DE FIGURAS

<b>Figura 1.</b> Superfamilia Riboquinasa .....	13
<b>Figura 2.</b> Comparación de estructuras de glucoquinasas y fosfofructoquinasas dependientes de ADP.....	32
<b>Figura 3.</b> Árbol filogenético consenso obtenido por inferencia bayesiana para secuencias de proteínas.....	39
<b>Figura 4.</b> Reconstrucción filogenética consenso obtenida por inferencia bayesiana para secuencias de ADN.....	40
<b>Figura 5.</b> Reconstrucción filogenética obtenida por máxima verosimilitud para secuencias de proteínas.....	41
<b>Figura 6.</b> Inferencia de las secuencias ancestrales ancGK/PFK, ancGK y ancPFK- MT.....	43
<b>Figura 7.</b> Reconstrucción de ancPFK/GK por distintos métodos.....	46
<b>Figura 8.</b> Análisis de la inferencia de ancPFK-T y ancPFK-M.....	51
<b>Figura 9.</b> Modelamiento e interacción proteína ligando de los ancestros PFKs.....	53
<b>Figura 10.</b> Simulaciones de dinámica molecular de la enzima ancestral ancPFK- MT.....	55
<b>Figura 11.</b> Análisis de la historia del uso de glucosa y fructosa-6P como rasgo discreto durante la historia evolutiva de la familia.....	57
<b>Figura 12.</b> Análisis de la historia del uso de glucosa y fructosa-6P como rasgo discreto durante la historia evolutiva de la familia reconsiderando el uso actual de sustratos.....	59

<b>Figura 13.</b> Análisis de la reconstrucción de estados ancestrales discretos obtenidos por inferencia bayesiana para ancGK/PFK.....	61
<b>Figura 14.</b> Distribución de probabilidades posteriores para la reconstrucción de estados ancestrales discretos para el uso de glucosa y fructosa-6P mediante inferencia bayesiana. ....	62
<b>Figura 15.</b> Determinación de los parámetros cinéticos de la mutante E72A de ancPFK-MT.....	66
<b>Figura 16.</b> Espectros de DC para diferentes tipos de estructuras secundarias encontradas en proteínas. ....	104
<b>Figura 17.</b> Comparación entre espectro experimental de <i>TIGK</i> y <i>PhPFK</i> y los espectros calculados por los programas SELCON3, CDSSTR y CONTINLL.....	108
<b>Figura 18.</b> Espectros de dicroísmo circular para las proteínas sin estructura cristalográfica. ....	110
<b>Figura 19.</b> Curvas de desplegamiento térmico.....	113
<b>Figura 20.</b> Superposición de estructuras de <i>PhPFK</i> depositadas en el banco PDB. ....	132
<b>Figura 21.</b> Cristales de <i>PhPFK</i> . ....	133
<b>Figura 22.</b> Determinación de la estructura de <i>PhPFK</i> . ....	133
<b>Figura 23.</b> Superposición entre la estructura de <i>PhPFK</i> determinada y las estructuras depositadas en el PDB.....	135

## LISTA DE ABREVIATURAS

aa	Aminoácido
ancC	Ancestro en común de enzimas HMPKs específicas y bifuncionales
ancB	Ancestro de enzimas HMPKs bifuncionales
ancS	Ancestro de enzimas HMPKs específicas
ancGK	Ancestro de glucoquinasas de <i>Thermococcales</i>
ancPFK/GK	Ancestro en común de actividad GK y PFK
ancPFK-MT	Ancestro de enzimas PFKs de <i>Methanococcales</i> y <i>Thermococcales</i>
ancPFK-M	Ancestro de PFKs de <i>Methanococcales</i>
ancPFK-T	Ancestro de PFKs de <i>Thermococcales</i>
CD	Dicroísmo Circular
DDC	Modelos de evolución por duplicación, degeneración y complementación.
EAC	Modelo de evolución por escape de conflicto adaptativo
EDTA	Ácido etilendiaminotetraacético
F6P	Fructosa-6-fosfato
GFP	Proteína fluorescente verde
GK	Glucoquinasa
HEPES	Acido 4-(2-hydroxyethyl)-1-piperazineetanosulfónico
HMP	4-amino-5-hidroximetil-2-metil pirimidina
HMPK	Quinasa de 4-amino-5-hidroximetil-2-metil pirimidina
IAD	Modelo de evolución por innovación, amplificación y divergencia
IPTG	Isopropil $\beta$ -D-1-tiogalactopiranosido
kDa	kilo Dalton
$K_i$	Constante de Inhibición
$k_{cat}$	Constante catalítica
$K_m$	Constante de <i>Michaelis-Menten</i>
$k_{cat}/K_m$	Constante de eficiencia catalítica o de especificidad
LB	Luria-Bertani
MCMC	Método de Montecarlo y Cadenas de Markov

<i>Mb</i> PFK/GK	Fosfofructoquinasa/glucoquinasa de <i>Methanococcoides burtonii</i>
<i>Mj</i> PFK/GK	Fosfofructoquinasa/glucoquinasa de <i>Methanocaldococcus jannaschii</i>
ML	Máxima Verosimilitud
<i>Mm</i> PFK/GK	Fosfofructoquinasa/glucoquinasa de <i>Methanococcus maripaludis</i>
NAD <sup>+</sup>	Forma oxidada de Nicotinamida dinucleotido
NADH	Forma reducida de Nicotinamida dinucleotido
PAGE	Gel de electroforesis de poliacrilamida
pb	Pares de bases
PCR	Reacción en cadena de la polimerasa
PEG	Polietilenglicol
PFK	Fosfofructoquinasa
PIPES	Ácido piperazina-N,N'-bis (2-etanosulfónico)
<i>Ph</i> PFK	Fosfofructoquinasa de <i>Pyrococcus horikoshii</i>
PL	Piridoxal
PLK	Quinasa de piridoxal
PLP	Piridoxal 5' fosfato
SAXS	Dispersión de rayos X a bajo ángulo
SDS	Dodecil sulfato de sodio
TAE	Tampón Tris-Ácido acético-EDTA
TEV	Proteasa del virus del tabaco
THG	Transferencia horizontal de genes
THZK	Hidroxietiltiazol quinasa
<i>Tl</i> GK	Glucoquinasa de <i>Thermococcus litoralis</i>
TRIS	Tris(hidroximetil)aminometano



## RESUMEN

Las enzimas juegan un rol fundamental en el metabolismo y la señalización celular. Su importancia radica en la capacidad de catalizar la transformación de un sustrato de manera específica, en el contexto de su papel en la vía metabólica en la que participa. Esta especificidad se ha alcanzado durante el curso de la evolución mediante cambios en la secuencia de aminoácidos y eventos de duplicación génica, lo que ha llevado a la aparición de superfamilias de enzimas con diversas especificidades, pero con una conservación estructural que evidencia un origen evolutivo en común.

En este trabajo se utilizó la metodología de resurrección de proteínas ancestrales para estudiar la evolución de la especificidad por sustratos en enzimas homólogas de la familia de quinasas de azúcar dependientes de ADP y en la familia de quinasas de vitaminas, ambas familias pertenecientes a la superfamilia Riboquinasa. Se identificaron determinantes moleculares responsables de especializaciones y aparición de actividades bifuncionales por sustratos, lo que nos permitió entender la actual distribución de las actividades enzimáticas en las proteínas actuales de las familias.

Los resultados obtenidos en este trabajo confirman la importancia del papel de las actividades promiscuas en la generación de nuevas funciones en proteínas y apoyan la idea que la adaptación de funciones a nivel molecular surge con un número reducido de mutaciones de gran efecto, con posterior surgimiento de un gran número de mutaciones de pequeño efecto. También resaltan la transferencia horizontal de genes como un mecanismo generador de redundancia génica en procariontes y por ende como otro motor de la generación de nuevas funciones, además de la duplicación génica. Así, las

historias evolutivas de las familias estudiadas concuerdan con un mecanismo de generación de nuevas funciones en proteínas de tipo innovación, amplificación y divergencia (IAD).

Los resultados de conservación de elementos de estructura secundaria como los análisis de estructura terciaria, en conjunto con estudios de cinética enzimática, mostraron que los cambios conformacionales serían relevantes para la unión de ligandos y por ende para la catálisis en la familia de quinasas dependientes de ADP y que estos cambios conformacionales estarían conservados en la familia, aunque con posibles diferencias en los mecanismos de unión de los sustratos. Por otra parte, los análisis de estructura terciaria concuerdan con un mecanismo de evolución tipo interrumpido o puntuado en la evolución de las estructuras de la superfamilia Riboquinasa.

## ABSTRACT

Enzymes play a key role in metabolism and cell signaling. Its importance lies in the ability to catalyze the specific transformation of a substrate, in the context of its role in the metabolic pathway in which it participates. This specificity has been achieved during the course of evolution by sequence mutations and gene duplication events, and this has led to the emergence of superfamily of enzymes with different specificities, but with a structural conservation that evidences a common evolutionary origin.

In this work the resurrection of ancestral protein methodology was used to study the evolution of substrate specificity in homologous enzymes of the ADP-dependent sugar kinases family and vitamins kinases family, both families belonging to Ribokinase superfamily. We identified key molecular determinants in activity specialization and in the emergence of bifunctional activity, allowing us to understand the current distribution of enzyme activities in the current protein of these families.

The results obtained in this study confirm the importance of promiscuous activities in the generation of new protein functions and support the idea that adaptation of functions at the molecular level involves few mutations with large effects that are later adjusted by numerous mutations with small effect. They also emphasize the horizontal transfer of genes as a mechanism for gene redundancy in prokaryotes and thereby as another motor for generation of new functions, besides gene duplication. Thus, the evolutionary histories of the families studied here are consistent with the mechanism called innovation, amplification and divergence (IAD) for generation of new functions in proteins.

The conservation of secondary structure elements as well as the tertiary structure analysis, in conjunction with enzyme kinetic studies, showed that conformational changes would be relevant for ligand binding and thereby for catalysis in ADP-dependent kinases family and these conformational changes would be conserved in the family, but with possible differences in the substrates binding mechanisms. Moreover, analysis of the evolution of the tertiary structure of proteins belonging to the Ribokinase superfamily is consistent with a punctuated equilibrium mechanism for protein structure evolution.

# 1 INTRODUCCIÓN GENERAL

## 1.1 Generalidades sobre evolución

La característica que todos los organismos vivos de la tierra compartan mecanismos moleculares como la utilización del ADN como material genético, rutas metabólicas, la conservación de las estructuras de proteínas, así como la utilización de estereoisómeros específicos como el enantiómero D de la desoxirribosa en el ADN y las proteínas formadas solo por aminoácidos L, por nombrar algunos, apoya la teoría que todos los organismos vivos descendemos de un ancestro en común y que por los tanto los mecanismos moleculares involucrados son extrapolables entre organismos, llegando a expresarse en aforismos tales como “*lo que es válido para la bacteria es válido para el elefante*”[1]. Esto es sorprendente desde un punto de vista taxonómico, desde el cual se observa la gran variedad de tipos y formas de vida, pero desde una mirada química es claro que las moléculas y mecanismos que sustentan las formas de vida sean compartidos por todos los organismos presentes en la tierra y estén restringidos por las condiciones físico-químicas en que se han desarrollado.

Una reflexión rápida sobre la diversidad de la vida basta para evidenciar la gran inmensidad estructural y funcional de organismos vivos, por lo que la pregunta; ¿Cómo se originó esta diversidad? tiene una larga tradición a lo largo de la historia de la ciencia. Unos de los grandes avances en cuanto a explicar la diversidad biológica la propuso Charles Darwin en 1859 con su obra *El origen de las especies*, en donde propone la selección natural como mecanismo por medio del cual las diferencias en la

competitividad, directa o indirecta, de los individuos resulta en la capacidad de una población de modificarse evolutivamente a través del tiempo [2]. Entre 1930 y 1940 surgió la síntesis moderna o Neodarwinismo la cual integró a la teoría de la evolución de especies, la genética y la mutación génica aleatoria como fuente de variación [3]. La síntesis moderna propuso que la velocidad y la dirección del proceso evolutivo están determinadas por la selección natural, que las mutaciones son en su mayoría perjudiciales y que las ventajosas ocurren en menor medida, por lo que las mutaciones neutrales o sin efecto no existirían o serían muy escasas. Así también la evolución es gradual, lenta y se presenta como un patrón lineal en el tiempo. Actualmente esta teoría también es conocida como teoría seleccionista o sintética [1]. Bajo estas mismas bases pero criticando el patrón lineal de la evolución, la temporalidad en la cual ocurren los cambios y el modo en que se despliegan, se ha propuesto la teoría de equilibrio interrumpido (o puntuado), en el cual los cambios ocurren en períodos cortos y las especies permanecen con cambios menores la mayor parte del tiempo, negando así el ritmo uniforme de la teoría sintética [4].

Posterior a 1950 con la aparición de la biología molecular y la capacidad de obtener secuencias de ADN y de proteínas fue posible estudiar la evolución a nivel molecular. De esta manera en 1965 Zuckerkandl y Pauling notaron que la velocidad de sustitución de aminoácidos en hemoglobinas o en citocromos *c* era similar en varios linajes de mamíferos, por lo que propusieron que las proteínas acumulan mutaciones a una velocidad constante y característica de cada tipo de proteína, lo que permitiría fechar divergencias entre especies, surgiendo así la propuesta del reloj molecular de la

evolución [5]. Estas observaciones, junto a otras como el aparente azar de las sustituciones, la alta velocidad de sustituciones de nucleótidos en genomas de mamíferos y la alta heterogeneidad y polimorfismos, hicieron surgir la teoría neutralista de la evolución propuesta por Kimura, King y Jukes en 1968. Esta teoría postula que la gran mayoría de los cambios evolutivos se producen por fijación al azar de mutaciones selectivamente neutrales o casi neutras, y aunque no niega el rol de la selección natural en la determinación de las adaptaciones durante la evolución, sugiere que solo una porción muy pequeña de las mutaciones tienen efectos sobre la adaptación, así tampoco niega el rol de las mutaciones desfavorables y su desaparición por selección purificadora [1,6]. La teoría neutralista se enfrentó a un intenso debate con partidarios de la teoría seleccionista la cual descarta el rol de las mutaciones neutras. Recientemente se ha propuesto que ambas teorías se pueden reconciliar considerando que las mutaciones neutrales y la robustez a mutaciones son importantes para las innovaciones evolutivas en proteínas; a mayor robustez de una molécula mayor es la probabilidad de que las mutaciones sean neutras y mayor es la posibilidad de adquirir una nueva función [7]. Esta adquisición de nuevas funciones está ligada a la adaptación del organismo a nuevas condiciones ambientales. Estudios experimentales de adaptación han mostrado que en el comienzo del proceso adaptativo las nuevas funciones que permiten adaptación involucran un pequeño número de cambios genéticos con mutaciones discretas que tienen un gran efecto sobre el fenotipo o *fitness* y que luego siguen mutaciones de pequeño efecto [8].

La evolución de proteínas es responsable en gran parte de la diversidad de formas y funciones encontradas en los organismos actuales, es por esto que entender la naturaleza de la diversificación funcional de las proteínas es primordial para la biología evolutiva.

## ***1.2 Evolución de Proteínas***

Los mecanismos genéticos y bioquímicos por los cuales las proteínas evolucionan y aparecen nuevas funciones son un objetivo central en la evolución molecular [9]. En 1979 el modelo de Ohno propone una explicación para la aparición de nuevas funciones. Este modelo propuso que la duplicación génica crea una copia redundante que no estaría bajo la presión de la selección natural y por lo tanto podría acumular mutaciones que estarían “prohibidas” en la función ancestral. Esto permitiría eventualmente llevar a una nueva función en una de las copias duplicadas [10]. Actualmente sabemos que no todas las duplicaciones de genes involucran aparición de nuevas funciones, de hecho muchos de los genes duplicados son mantenidos en los genomas probablemente por mecanismos pasivos como la sub-funcionalización, en la cual las copias hijas comparten partes de la función de la proteína ancestral [11]. Desde la propuesta de Ohno, se han propuesto mecanismos para explicar cómo aparecen las nuevas funciones en proteínas, entre los cuales podemos mencionar los que involucran neo-funcionalización, como el modelo de mutación durante la no-funcionalidad (MDN en inglés *mutation during non-functionality*) o aquellos que involucran sub-funcionalización a través de la deriva génica, como el modelo de duplicación, degeneración y complementación (DDC en inglés *duplication, degeneration, complementation*), o el modelo de selección positiva denominado como escape de conflicto adaptativo (EAC en inglés *escape from adaptive*



*conflict*). Todos estos modelos consideran la duplicación génica como motor de la aparición de nuevas funciones [9], aunque evidencia reciente ha postulado que la transferencia horizontal de genes (THG) podría jugar un rol muy importante en la evolución de funciones de proteínas en los procariontes, ya que este mecanismo también permite crear copias redundantes de genes [12,13]. Recientemente, se ha propuesto un modelo más general denominado innovación, amplificación y divergencia (IAD) que contempla las actividades promiscuas como el material de partida para la generación de actividades duales o bifuncionales, además de considerar el aumento de la expresión génica mediante la redundancia de genes (por duplicación o THG) como una posibilidad para que ocurra selección positiva para la fijación de estos genes en una población [14].

### ***1.3 Evolución de enzimas***

Una de las principales funciones de las proteínas es la catálisis de reacciones químicas involucradas en el metabolismo. Estas proteínas, denominadas enzimas, se pueden agrupar ya sea en familias, en base a similitud de secuencia/funciones o en superfamilias, en base a su estructura y mecanismo catalítico. Aunque funcionalmente diversos, los miembros de una superfamilia, comparten un mecanismo químico de catálisis, en donde sus sitios activos poseen residuos conservados respecto a sus propiedades físico-químicas, los que permiten activar ciertos grupos químicos en los sustratos y estabilizar los estados de transición de las reacciones, disminuyendo así la energía de activación y por consiguiente aumentando la velocidad de la reacción en ordenes de magnitud [15].

Una superfamilia de enzimas puede tener miembros con distintas funciones en diferentes vías metabólicas y dichas funciones radican en determinantes estructurales que dictan distintas especificidades por sustratos. El término superfamilia supone homología entre todos los miembros, por lo que se considera que todos los miembros comparten un ancestro en común y que a través de la evolución se habrían generado las distintas familias y miembros presentes en la actualidad con distintas especificidades por sustratos y por ende con distintas funciones [16]. Por tanto, la catálisis de nuevas reacciones durante la evolución no serían creadas *de novo*, sino que utilizan residuos y mecanismos catalíticos ya presentes en enzimas [17]. Esta característica de la evolución a nivel molecular sería análoga a la evolución de estructuras anatómicas complejas en organismos, donde una estructura adaptada para un propósito en particular es utilizada en una nueva función la cual le resulta selectivamente beneficiosa (exaptación). Esta situación es denominado como el *principio del pulgar del panda*, ya que este animal no tiene un pulgar como tal, pero tiene una extensión del hueso sesamoideo que cumple la función de pulgar [18]. Otro buen ejemplo de esto es el descubrimiento de las plumas de dinosaurios no aviares, las cuales habrían evolucionado para la regulación de la temperatura, pero habrían sido adaptadas para volar en las aves actuales [19].

Esta exaptación de estructuras moleculares en la creación de nuevas funciones en enzimas es apoyada por estudios de evolución dirigida con familias y superfamilias de enzimas. Estos estudios muestran que la aparición de nuevas funciones proviene de la evolución de enzimas promiscuas o inespecíficas, las cuales mantienen su función original mientras desarrollan una nueva función, generalmente pasando por un estado

bifuncional o multifuncional. En este camino evolutivo la promiscuidad catalítica cumple un rol fundamental en la aparición de nuevas actividades enzimáticas [20,21]. En este contexto es importante hacer la diferencia entre una enzima promiscua y una enzima que cumple dos funciones; la promiscuidad catalítica se define como otra actividad catalítica para la cual la enzima no ha evolucionado y que por tanto no es relevante para la fisiología del organismo, mientras que una enzima que ha evolucionado para catalizar la transformación de más de un sustrato en diferentes vías metabólicas o catalizar la transformación de una amplia gama de sustratos, es una enzima multifuncional o de amplia especificidad [20].

El efecto de la selección natural en las enzimas ha sido estudiado desde los años 70's; los primeros trabajos fueron publicados por Fersht [22], Crowley [23] y Cornish-Bowden [24]. En estos trabajos se postula que las enzimas tienden a incrementar sus eficiencias catalíticas dadas principalmente por aumentos en sus  $k_{cat}$  en desmedro de un aumento de las  $K_m$  para los sustratos. En ese entonces, estos autores hacen la diferencia entre enzimas del metabolismo central, las cuales en su mayoría no se encuentran saturadas *in vivo*, de otras enzimas, como las proteasas del sistema digestivo, las cuales se encuentran en abundancia de sustrato. Para las primeras proponen que la evolución tiende a aumentar sus  $k_{cat}$  y a mantener diferencias entre concentraciones de sustratos y  $K_m$ , de un orden de magnitud como máximo; mientras que en las segundas solo  $k_{cat}$  es la que estaría sometida a la restricción selectiva. Actualmente, se sostiene que la evolución de la eficiencia catalítica es dependiente de la función de la enzima y del contexto metabólico en el que trabaja, encontrándose que las enzimas de vías metabólicas

principales están fuertemente sometidas a selección, mientras que enzimas de vías metabólicas secundarias estarían bajo un régimen menos restrictivo, lo que se evidencia por  $k_{cat}$  menores y una especificidad por sustrato más amplia [25]. Pero además de las restricciones evolutivas, los parámetros cinéticos tienen límites teóricos de acuerdo a ciertas restricciones fisicoquímicas como la formación y ruptura de ciertos enlaces y el tiempo requerido para ciertos cambios conformacionales, lo que limita teóricamente la  $k_{cat}$  a un máximo de  $10^6 - 10^7 \text{ s}^{-1}$  [26]. La difusión de las moléculas en el medio impone un límite superior para la constante de segundo orden  $k_{cat}/K_m$   $10^8 - 10^9 \text{ M}^{-1}\text{s}^{-1}$ . Así también las energías de activación de las reacciones son otro gran obstáculo, ya que si una reacción no catalizada es extremadamente lenta, una enzima puede mejorar la velocidad en varios órdenes de magnitud pero seguirá siendo lenta en términos prácticos. Por otra parte, hay restricciones termodinámicas como la impuesta por la relación de Haldane ( $K_{eq} = [k_{cat} * K_m]_{directa} / [k_{cat} * K_m]_{reversa}$ ), por lo que si la reacción directa tiene una  $k_{cat}/K_m$  limitada por la difusión, la velocidad de su reacción reversa estará muy lejos de este máximo. También se han encontrado correlaciones entre la afinidad de los sustratos con los pesos moleculares y la hidrofobicidad de éstos, limitando las mayores afinidades a los sustratos pequeños y más hidrofóbicos.

#### ***1.4 Reconstrucción de secuencias ancestrales (RSA)***

En 1963 Linus Pauling y Emile Zuckerkandl [20] publicaron un artículo visionario titulado '*Chemical paleogenetics molecular restoration studies of extinct forms of life*', en el que proponen la idea de reconstruir la secuencia aminoácidica de proteínas ancestrales en virtud de una comparación entre secuencias relacionadas encontradas en

organismos actuales. Aunque la tecnología del momento no permitió que plasmaran sus ideas, ellos señalaron que: *“una vez que se conozca la secuencia de la cadena polipeptídica (ancestral), en el futuro será posible sintetizar este presunto componente de un organismo extinto. Así uno será capaz de estudiar las propiedades físico-químicas de estas moléculas y hacer inferencias sobre sus funciones. Por ejemplo, se podrá estudiar la afinidad del oxígeno y su dependencia con el pH en hemoglobinas ancestrales, así como la afinidad de enzimas ancestrales por varios sustratos y la probable naturaleza de estos sustratos en el pasado de la historia evolutiva”*. Desde aquella publicación hasta nuestros días, los avances en los métodos relacionados con la inferencia filogenética, los modelos de evolución, algoritmos, tecnología computacional e ingeniería genética han hecho posible llevar a la práctica la idea de Pauling y Zuckerkandl. Se han propuesto varios métodos para reconstruir secuencias de ancestros en un árbol filogenético. Unos de los primeros métodos fue el de parsimonia, el cual busca una combinación de estados ancestrales (caracteres) en un nodo interior de la filogenia, tal que minimice el número de cambios necesarios para explicar todo el árbol evaluado. Con este método a menudo resultan varias secuencias con la misma parsimonia y en estos casos se dice que la reconstrucción es ambigua. Posteriormente, se desarrollaron métodos para la reconstrucción de secuencias de aminoácidos y ADN usando modelos estocásticos, con modelos de sustitución empíricos y paramétricos. El método de máxima verosimilitud utiliza estos modelos de sustitución y busca el estado del carácter ancestral en un nodo interno del árbol que maximiza la probabilidad de observar los caracteres actuales, por lo que no calcula directamente la probabilidad del carácter ancestral y por lo tanto no permite evaluar la confiabilidad de la reconstrucción.

En 1995 Yang publicó un método para actualizar la verosimilitud de los caracteres ancestrales calculados por máxima verosimilitud a una probabilidad posterior utilizando el teorema de Bayes [27]. Esta aproximación utiliza máxima verosimilitud para estimar los parámetros del modelo de sustitución (transiciones/transversiones), topología y largo de ramas del árbol y luego usa estos parámetros para calcular la probabilidad posterior de los caracteres ancestrales. De esta manera se obtiene directamente la probabilidad del carácter ancestral dado los datos actuales. Esta aproximación es conocida como el método de Bayes empírico o el método de máxima verosimilitud para reconstrucción de secuencias ancestrales. Un método más reciente llamado Bayes jerárquico, considera la incertidumbre en la estimación de los parámetros del modelo de evolución, en la topología y en el largo de rama y por ende esta aproximación infiere el estado del carácter ancestral sin condicionarse a un simple árbol ni a un grupo de valores de los parámetros de éste, utilizando poblaciones de árboles y de parámetros que expliquen la filogenia, mediante la metodología de MCMC (del inglés *Markov Chain Monte Carlo*) [28]. Actualmente, hay varios ejemplos exitosos de resurrección de proteínas ancestrales, entre los que se encuentran ancestros de opsinas [29], proteínas tipo GFP [30,31], receptores de esteroides [32], alcohol deshidrogenasas [33], por nombrar solo algunos. Por ejemplo, las proteínas tipo GFP de arrecifes de coral, son proteínas que fluorescen a una longitud de onda que está determinada por su secuencia de aminoácidos. Mediante la resurrección de proteínas ancestrales, Fiel y Matz (2010), encontraron que la emisión de fluorescencia del ancestro de estas proteínas correspondía al color verde y que luego se diversificó en una variedad de otros colores. En ese trabajo además identificaron las mutaciones necesarias para intercambiar el color de la emisión

de fluorescencia entre las proteínas actuales, en donde algunas de estas mutaciones presentaban fuertes interacciones epistáticas, que según los autores no hubiesen sido encontradas sin la ayuda de la reconstrucción ancestral [30]. Como mencionan Harms y Thornton (2010) : *“un beneficio de la reconstrucción de secuencias ancestrales es que provee un puente entre la bioquímica mecanística y la biología evolutiva, campos que en gran parte han estado separados, a pesar de la prolongada existencia de fascinantes preguntas en su interface”* [34].

### ***1.5 Miembros de la Superfamilia Riboquinasa como modelos para el estudio de la evolución de la especificidad por sustratos***

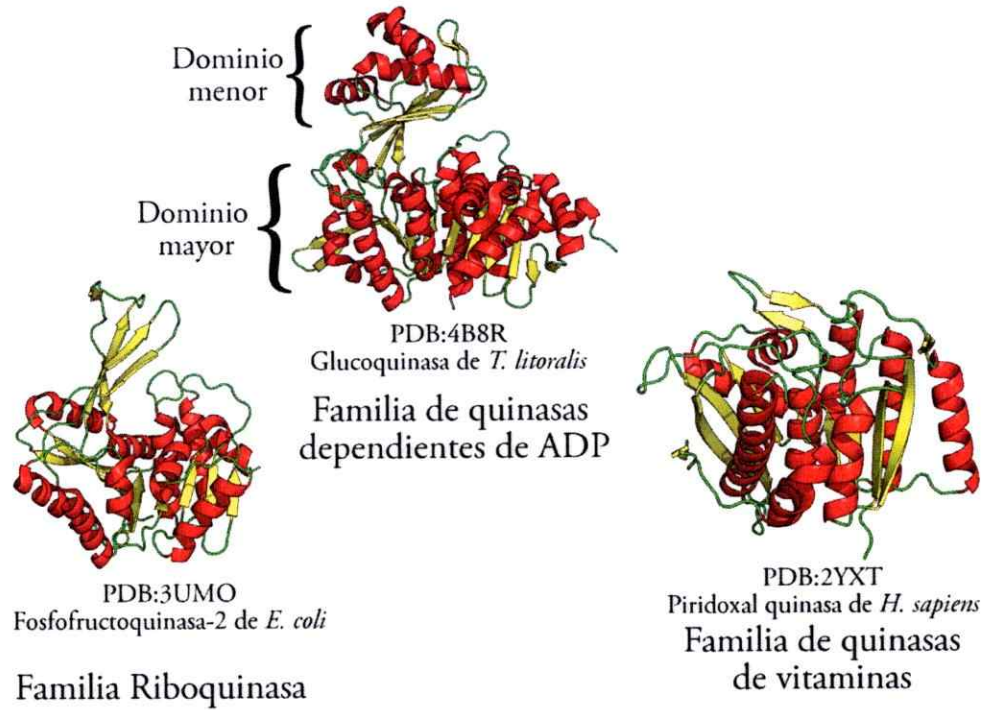
Los miembros de la superfamilia Riboquinasa muestran una notable conservación estructural, a pesar de la baja identidad de secuencia entre ellos. La arquitectura común entre estas enzimas consiste en un plegamiento tipo Rossman, con ocho hojas beta en una disposición predominantemente paralela, rodeadas de hélices alfa, lo que constituye el dominio base (dominio mayor). Además, en algunas proteínas de esta superfamilia, es usual encontrar inserciones de elementos de estructura secundaria los cuales forman un segundo dominio más pequeño (dominio menor).

En una primera instancia esta superfamilia fue propuesta solamente en base a información de secuencia, e incluía únicamente a enzimas dependientes de ATP de la actual familia Riboquinasa [35]. Posteriormente, basados en la similitud estructural, enzimas con baja identidad de secuencia también fueron identificadas como miembros de esta superfamilia; entre ellas están las quinasas de vitaminas dependientes de ATP y las quinasas de azúcares dependientes de ADP, estas últimas identificadas en un

comienzo en arqueas, aunque actualmente se han encontrado representantes en eucariontes superiores [36,37]. Actualmente, la superfamilia Riboquinasa se divide en tres grupos o familias: quinasas de vitaminas dependientes de ATP, las cuales no presentan dominio menor; quinasas de azúcares dependientes de ATP (familia Riboquinasa) y quinasas de azúcares dependientes de ADP, estas dos últimas con la presencia de un dominio menor en su estructura (**Figura 1**) [38].



## Superfamilia Riboquinasa



**Figura 1. Superfamilia Riboquinasa.**

Esta superfamilia está compuesta por tres familias; familia riboquinasa, familia de quinasas dependientes de ADP y familia de quinasas de vitaminas, esta última familia solo presenta el dominio mayor en su estructura.



De acuerdo a alineamientos de secuencia basados en comparación de estructuras tridimensionales se han encontrado motivos estructurales y de secuencia involucrados en un mecanismo catalítico similar para todos los miembros de esta superfamilia. Entre éstos se encuentra un ácido aspártico estrictamente conservado dentro de un motivo llamado GXGD, el cual es propuesto como la base catalítica que activa a un hidroxilo del sustrato para el ataque nucleofílico. En algunos miembros de la familia de las quinasas de vitaminas se encuentra una cisteína en lugar del aspártico cumpliendo la misma función [39–42]. Otro motivo conservado es el NXXE, el que ha sido asociado a la unión de metales. Para varios miembros de la superfamilia se ha demostrado que no es posible transferir el grupo fosforilo en ausencia de un metal catiónico divalente y que estos además tienen un rol regulatorio.

El presente estudio se centró en un inicio en la familia de quinasas dependientes de ADP presentes en arqueas. En esta familia se han descrito solo dos actividades catalíticas; enzimas que fosforilan glucosa (glucoquinasas) y enzimas que utilizan como sustrato fructosa-6-fosfato (F6P) denominadas fosfofructoquinasas (PFKs). Se ha descrito que estas enzimas poseen una alta especificidad por sus respectivos sustratos y solo se ha reportado una enzima relacionada a las PFKs que es capaz de fosforilar glucosa y F6P, la enzima bifuncional de *Methanococcus jannaschii* [43]. Sin embargo, se desconoce si el rasgo bifuncional de esta enzima es un rasgo ancestral en la familia, es una novedad evolutiva o si la enzima de este organismo es la única enzima de la familia con esta propiedad. Además en esta familia se realizó un estudio de conservación estructural y de la importancia de cambios conformacionales en la catálisis realizada por estas enzimas.

Posteriormente, extendimos nuestro estudio a la familia de quinasas de Vitaminas, la cual presenta actividades piridoxal quinasa (PLK), hidroximetil pimimidina quinasa (HMPK) e hidroxietil tiazol quinasa (THZK). En esta familia también se ha descrito un grupo de enzimas con actividad bifuncional HMPK/PLK, las cuales estarían filogenéticamente relacionadas a las HMPK específicas [44]. En esta familia nuestro estudio se centró en determinar el origen de los rasgos actuales (bifuncional y específico) en el grupo de las HMPKs y la importancia de la promiscuidad catalítica en la aparición de nuevas funciones.

## 2 MATERIALES Y MÉTODOS

### 2.1 *Materiales.*

#### 2.1.1 Vectores

Como vector de expresión se utilizó el plásmido pET-TEV que es una versión modificada del plásmido pET-28a (NOVAGEN), en el cual la secuencia que codifica el sitio de corte por trombina fue remplazada por una secuencia de corte por la proteasa del virus del tabaco (TEV). El vector tiene una longitud de 5369 pb, posee el promotor T7/lac, otorga resistencia a kanamicina, presenta la posibilidad de incorporar un *tag* de histidina (HIS-tag) en el N-terminal o C-terminal y además el HIS-tag del N-terminal puede eliminarse mediante la proteasa TEV. Las proteínas expresadas en este estudio fueron sub-clonadas en este vector en los sitios de restricción NdeI y BamHI.

Los genes fueron sintetizados por la compañía comercial (GenScript USA Inc., Piscataway, NJ, Estados Unidos) y recibidos en el vector de clonamiento pUC57. Este plasmido tiene una longitud de 2710 pb, posee un sitio de clonamiento múltiple y otorga resistencia a ampicilina. Los genes sintetizados fueron sub-clonados entre los sitios de restricción NdeI y BamHI.

#### 2.1.2 Cepas

Para las metodologías de biología molecular como sub-clonamientos y mutagénesis sitio dirigida, se utilizó la cepa de *E. coli* DH5 $\alpha$ , que tiene un genotipo; F-80dlacZ M15 (lacZYA-argF) U169 recA1 endA1hsdR17(rk-, mk+) phoA<sup>sup</sup>E44 -thi-1 gyrA96 relA1.

Para la sobreexpresión de proteínas se utilizó la cepa de *E. coli* BL21(D3), cuyo genotipo es *fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHI ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5* y la cepa BL21(DE3)pLyss con el mismo genotipo anterior, pero que incorpora el plasmido pLyss que confiere resistencia a Cloranfenicol y codifica para la lisozima del fago T7, la cual es un inhibidor de la T7 polimerasa, lo que reduce la expresión basal del promotor T7.

### 2.1.3 Medios de Cultivo

El medio Luria Bertani (*LB*) se preparó utilizando 10 gr/lt de Bacto-triptona, 5 gr/lt de extracto de levadura y 5 gr/L de NaCl. Luego, se ajustó a pH 7.0 con NaOH 5N y se autoclavó.

Las placas LB-agar se prepararon utilizando medio LB y agregando 15 gr/lt de agar, luego se autoclavó, se enfrió a 50 °C y se le adicionaron los antibióticos esterilizados por filtración.

### 2.1.4 Soluciones generales

*TAE 10X*: un litro de solución contiene 48.4 gr de Tris, 11.4 ml de ácido acético glacial y 20 ml de EDTA 0.5 M pH 8.0.

*Buffer TG-SDS10X*: un litro de solución contiene 30.2g de Tris, 144.0g de glicina y 10.0 g de dodecilsulfato de sodio (SDS), se ajusta a pH 8.4 a temperatura ambiente.

### 2.1.5 Reactivos químicos generales

Los reactivos químicos generales, amortiguadores, sales y sustratos glucosa, fructosa-6-fosfato, glucosa-6-fosfato, fructosa 1,6-bifosfato, ADP, ATP y PL utilizados para los

ensayos enzimáticos fueron adquiridos a la compañía Sigma-Aldrich (Sigma-Aldrich Corp, St. Louis, MO, Estados Unidos).

El sustrato 4-amino-5-hidroximetil-2-metil pirimidina (HMP; CAS:73-67-6) fue adquirido de la empresa Finetech (Fine Tech Ind Ltd, <http://www.finetechnology-ind.com>).

## 2.2 *Métodos*

### 2.2.1 Electroforesis en gel de agarosa

La agarosa se preparó a una concentración de 1% peso/volumen en TAE 1X y Gel Red 1X (Biotium Inc, Hayward, CA, Estados Unidos). A las muestras analizadas se les agregó solución de carga 1X (Fermentas Inc, Hanover, MD, Estados Unidos) y la electroforesis se realizó a un voltaje constante de 90 V. Como patrones para estimar el tamaño de los fragmentos de ADN, se usaron marcadores de masa molecular MassRuler (Bio-Rad Laboratories, Hercules, CA, Estados Unidos). El ADN en los geles de agarosa fue detectado mediante un transiluminador UV.

### 2.2.2 Electroforesis de proteínas en geles de poliacrilamida

Las muestras se re-suspendieron en una solución que contenía Tris-HCl pH 7.6 60 mM, SDS al 10%, glicerol al 10%,  $\beta$ -mercaptoetanol al 5% y azul de bromofenol al 0.02%, y fueron calentadas a 100 °C por 5 minutos. La electroforesis se realizó en geles de poliacrilamida al 12% en presencia de solución TG-SDS 1X, a un amperaje constante de 30 mA. Una vez finalizada la electroforesis, los geles fueron teñidos con azul de Coomassie al 0.1% y destañidos con una mezcla de metanol 40%, ácido acético 10% y agua.

### 2.2.3 Aislamiento de plásmidos

Los plásmidos fueron purificados utilizando kit comerciales (E.Z.N.A. Plasmid Miniprep Kit II; Omega Bio-Tek, Norcross, GA, Estados Unidos), siguiendo las recomendaciones del fabricante.

### 2.2.4 Purificación de fragmentos de ADN mediante geles de agarosa

Los fragmentos de ADN obtenidos mediante la digestión de plásmidos con las enzimas de restricción NdeI y BamHI (Fermentas), fueron purificados desde geles de agarosa utilizando el kit UltraClean 15 DNA Purification Kit (MO BIO Laboratories Inc, Carlsbad, CA, Estados Unidos).

### 2.2.5 Mutagénesis sitio dirigida

La mutante E72A del ancestro ancMT se realizó con las instrucciones de *QuikChange Site-Directed Mutagenesis kit* (Agilent Thechnologies, Santa Clara, USA), utilizando los siguientes oligos:

Directo: 5-CGGGTAAACCGGCAGCAGTCCCGCTGGACAACGAAG-3

Reverso: 5-CTTCGTTGTCCAGCGGGACTGCTGCCGGTTTACCCG-3

### 2.2.6 Cuantificación de proteínas

Para los análisis cinéticos las concentraciones de proteínas fueron determinadas mediante el método de Bradford, empleando albúmina de suero de bovino como estándar. Para los análisis espectroscópicos la concentración fue determinada a 280 nm



utilizando los coeficientes de extinción teóricos de las proteínas, obtenidos desde la herramienta ProtParam de ExPasy Proteomic Server [45].

#### 2.2.7 Reconstrucción de la filogenia e inferencia de las secuencias de enzimas ancestrales

La construcción de los árboles filogenéticos por inferencia bayesiana y las reconstrucciones de secuencias ancestrales por el método de bayes jerárquico, se llevaron a cabo con el programa MrBayes3.1 [28][46]. El análisis filogenético de secuencias de ADN se basó en un alineamiento de las proteínas, estrategia que ha mostrado ser adecuada [47]. Se extrajeron todas las secuencias disponibles de cada familia en la base de datos de proteínas no redundantes mediante un PSI-BLAST utilizando como plantados secuencias de proteínas caracterizadas experimentalmente (ejemplo, GK de *T. litoralis*). Luego, se realizó un alineamiento estructural con las proteínas a las que se les conoce su estructura tridimensional y sobre este perfil se alinearon las demás secuencias conocidas. Además, se realizó un alineamiento basado en estructura en el servidor PROMALS-3D[48], comparándose los resultados y editando manualmente las posiciones mal alineadas. Este alineamiento de secuencias de proteínas fue traducido hacia ADN de acuerdo a los genes de cada proteína utilizando el servidor RevTrans [49]. El modelo de evolución se seleccionó de acuerdo a JModeltest [50]. Los análisis se realizaron con 5.000.000 de generaciones, con una frecuencia de muestreo de 1000 y con un burn-in de 1500 muestras. La reconstrucción de secuencias ancestrales por Bayes jerárquico utilizando ADN o aminoácidos se realizó según lo descrito por Hall 2006 [51,52]. Para los análisis bayesianos basados en el alineamiento de

aminoácidos se utilizaron modelos de evolución empíricos de acuerdo a lo seleccionado por ProtTest [53]. La reconstrucción filogenética por máxima verosimilitud para secuencias de aminoácidos se realizó con el programa MEGA 6.06, utilizando el modelo de evolución de acuerdo a la selección de modelos por el factor de información de akaike (AIC) en el programa ProtTest [53]. La inferencia de secuencias ancestrales por Bayes empírico se realizó con el programa PAML 4.4 [54] utilizando la herramienta Lazarus, de acuerdo a lo descrito por Hanson-Smith et al. [55].

#### 2.2.8 Reconstrucción de estados ancestrales para rasgos discretos

La reconstrucción de los estados ancestrales de los rasgos se realizó por máxima verosimilitud utilizando el programa Mesquite 2.75 [56]. Para discriminar entre el modelo de un parámetro MK1 y el modelo markoviano de estados k con dos parámetros asimétricos se usó la prueba de razones de verosimilitud. Para todos los rasgos evaluados el estadístico (LR) no fue mayor al parámetro p de la distribución chi cuadrado con 1 grado de libertad, por lo que se trabajó con el modelo más sencillo, el MK1.

Las inferencias de estados ancestrales discretos por inferencia bayesiana se realizó en el programa BayesTraits [57] mediante la metodología MCMC. Los rasgos evaluados se mapearon en 500 árboles obtenidos en MrBayes3.1. Para ello se utilizó el modelo de solo una velocidad de cambio ( $q_{01}=q_{10}$ ) en contraste con el modelo con dos velocidades ( $q_{01}\neq q_{10}$ ) y se utilizó el factor de bayes para discriminar entre ambos modelos. Los análisis se hicieron con 10 millones de iteraciones, con un burn-in de 100 mil y con una frecuencia de muestreo de 1000. El parámetro de ratedev se ajustó en cada corrida para

que la “*acceptance*” se mantuviera entre 20-40%, en tanto que la secuencia de *Homo sapiens* se excluyó de todos los análisis ya que fue usada para enraizar los árboles. Los archivos de parámetros y el cálculo del factor de bayes fueron evaluados mediante el programa Tracer [58].

#### 2.2.9 Síntesis, expresión y purificación de proteínas ancestrales

Las secuencias de genes inferidas fueron enviadas para su síntesis a una compañía comercial (GenScript USA Inc., Piscataway, NJ, Estados Unidos). Posteriormente, los genes fueron subclonados en el vector pET-TEV (pET-28a modificado) en los sitios de restricción *NdeI* y *BamHI* y las proteínas fueron expresadas en la cepa de *Escherichia coli* BL21 (DE3). El vector pET-TEV introduce un tag de 6 histidinas (His) en el extremo N terminal, lo que permitió purificar las proteínas por columnas de afinidad. Además, este vector insertó un sitio de corte para la proteasa del virus del tabaco (TEV) para remover el *tag* luego de la purificación de la proteína. Las bacterias transformadas con los respectivos plásmidos fueron crecidas en medio LB con 35 ug/mL de kanamicina a 37°C hasta alcanzar OD<sub>600</sub> de 0.4 aproximadamente. La expresión de las proteínas fue inducida por 1 mM de Isopropil-β-D-tiogalactopiranosido (IPTG) durante toda la noche a 30°C. Las bacterias fueron centrifugadas por 15 minutos a 3.210 g y resuspendidas en Buffer de unión (Tris 50mM pH 7,8 NaCl 0,5mM, MgCl 5mM, imidazol 20mM). Posteriormente, las bacterias resuspendidas se sonicaron en hielo, con 10 pulsos de 20 segundos a 40% de amplitud con 1 minuto de descanso (Digital Sonifier 450, Branson). Luego, el lisado se centrifugó a 18.514 g por 30 minutos y la fracción soluble se filtró y cargó en una columna HisTrap (GE Healthcare) equilibrada con buffer

de unión. Para las proteínas termófilas el lisado se calentó a 70°C por 10 minutos, luego se centrifugó a 16.514 g por 10 min y el sobrenadante se cargó a la columna. Las proteínas fueron eluidas con un gradiente de 20 a 500 mM de imidazol y las fracciones con actividad enzimática se juntaron y dializaron contra PIPES 25 mM pH 6,5 y MgCl<sub>2</sub> 5 mM. Luego de la diálisis se agregó 50% de glicerol y 5 mM de ADP o ATP y la muestra se almacenó a -20°C.

#### 2.2.10 Determinación de la actividad enzimática

La actividad glucoquinasa dependiente de ADP se midió espectrofotométricamente de manera continua y en condiciones aeróbicas. El ensayo enzimático contenía amortiguador HEPES 25 mM pH 7,8, NAD<sup>+</sup> 0,5 mM, MgCl<sub>2</sub> libre 5 mM y concentraciones variables de glucosa y Mg-ADP. La producción de glucosa-6-P se acopló a la producción de NADH usando la enzima glucosa-6-P deshidrogenasa de *Leuconostoc mesenteroides* y se siguió espectrofotométricamente por el aumento de la absorbancia a 340 nm. Para el cálculo de las unidades de enzima se utilizó un coeficiente de extinción del NADH de 6,22 mM<sup>-1</sup> cm<sup>-1</sup>.

La actividad fosfofructoquinasa dependiente de ADP se midió espectrofotométricamente de manera continua y en condiciones aeróbicas. El ensayo enzimático contenía amortiguador PIPES 25 mM pH 6,5, NADH 0,16 mM, MgCl<sub>2</sub> libre 5 mM y concentraciones variables de fructosa-6P y Mg-ADP. La producción de fructosa-1,6-bisP fue acoplada al consumo de NADH utilizando como enzimas auxiliares aldolasa, triosa-P isomerasa y α-glicerofosfato deshidrogenasa (todas de músculo de conejo) y seguida por la disminución de la absorbancia a 340 nm [59]. Tanto para el ensayo de la

actividad glucoquinasa como fosfofructoquinasa las enzimas auxiliares no fueron limitantes y la unidad enzimática (U) corresponde a  $\mu\text{mol}/\text{min}$  de producto.

Para el ensayo de la actividad quinasa dependiente de ADP con distintos azúcares se empleó un ensayo discontinuo, en el cual se cuantificó la desaparición de ADP. La mezcla de reacción contenía de Tris-HCl 25 mM pH 7.0,  $\text{MgCl}_2$  20 mM, ADP 25 mM y cada azúcar a 10 mM. La mezcla fue incubada por 15 min a  $30^\circ\text{C}$  y la reacción detenida con una concentración de ácido perclórico al 3% final y puesta en hielo por 5 minutos, para luego ser neutralizada con  $\text{NaHCO}_3$ , 0.5 M final. Las muestras fueron centrifugadas a 18.620 g por 10 minutos y una alícuota del sobrenadante fue titulada en 500  $\mu\text{L}$  para determinar la concentración de ADP a  $37^\circ\text{C}$ . La mezcla de titulación de ADP contenía piruvato quinasa 0,8 U, lactato deshidrogenasa 2,4 U, fosfoenolpiruvato 0,3 mM, KCl 125 mM, Tris-HCl 25 mM pH 7,0,  $\text{MgCl}_2$  5 mM y NADH 0,2 mM en un volumen final de 0,5 mL. El consumo de ADP fue calculado usando como referencia el control sin azúcar.

Para medir la actividad PLK, se preparó un stock de piridoxal 100 mM, el cual fue cuantificado a una longitud de onda de 316 nm utilizando un  $\epsilon$  de  $8,61\text{mM}^{-1}\text{cm}^{-1}$  [60] a pH 7,2. La mezcla de reacción contenía ATP 10 mM,  $\text{MgCl}_2$  15 mM, PIPES 25 mM pH 6,5 y concentraciones variables de piridoxal. La actividad PLK fue medida espectrofotométricamente siguiendo la aparición de piridoxal fosfato (PLP), a una longitud de onda de 388 nm durante 10 minutos a  $37^\circ\text{C}$ .

La actividad HMPK fue evaluada siguiendo la aparición de ADP con un ensayo acoplado que emplea piruvato quinasa y lactato deshidrogenasa (PK/LDH), enzimas que utilizan ADP y NADH como sustratos, respectivamente. El sustrato HMP fue preparado a 100 mM y el ensayo de reacción contenía ATP 10 mM, MgCl<sub>2</sub> 15 mM NADH 0,15 mM, Tris-HCl 25 mM pH 7,8, KCl 125 mM, PEP 0,3 mM y 1,6 U/mL PK 4,8 U/mL LDH (piruvato quinasa - lactato deshidrogenasa) y concentraciones variables de HMP. El ensayo se realizó por 5 minutos a 37 °C. La titulación del HMP se realizó a 268 nm utilizando un  $\epsilon$  4.266 mM<sup>-1</sup>cm<sup>-1</sup> a pH 7,0 [55].

#### 2.2.11 Determinación de parámetros cinéticos: $k_{cat}$ , $K_m$ y $V_{max}$

La determinación de los parámetros cinéticos se realizó mediante estudios de velocidad inicial variando la concentración de un sustrato a concentraciones saturantes del co-sustrato. Para la obtención de  $V_{max}$  y  $K_m$  aparentes, la información fue analizada mediante el ajuste no lineal con GraphPad Prism versión 5.0 (GraphPad Software Inc., San Diego, CA, USA), utilizando las ecuaciones de Michaelis-Menten ( Ecuación 1) o de inhibición por sustrato (Ecuación 2). La constante  $k_{cat}$  (s<sup>-1</sup>) se determinó a partir de  $V_{max}$  (U/mg) considerando los peso moleculares teóricos de cada proteína estimados en la herramienta ProtParam del servidor ExPASy[45] (Ecuación 3).

Ecuación 1: 
$$v = \frac{V_{max} \cdot [S]}{K_m + [S]}$$

Ecuación 2: 
$$v = \frac{V_{max} \cdot [S]}{K_m + [S] \cdot (1 + \frac{[S]}{K_i})}$$

Ecuación 3: 
$$k_{cat} (s^{-1}) = \frac{V_{max} \left( \frac{U}{mg} \right) \cdot PM \left( \frac{g}{mol} \right)}{1000 \cdot 60 (s)}$$

### 2.2.12 Dicroísmo Circular y Estabilidad Térmica

Los experimentos de dicroísmo circular fueron ensayados en un espectropolarímetro Jasco J-815 acoplado a un sistema Peltier de control de temperatura. Todos los experimentos fueron realizados en buffer fosfato 25 mM pH 7,8 y 5 mM de MgCl<sub>2</sub>. Las medidas de los espectros fueron realizadas a 30 °C, a una velocidad de 50 nm/min, cada 0,2 nm, con 1 seg de respuesta, 1 nm de ancho de banda y una acumulación de 16 espectros. Para disminuir el ruido de los espectros se utilizó el filtro de la transformada de Fourier (FFT) implementado en el programa Origin 8.0. El cálculo de porcentajes de estructura secundaria se realizó con el software SDSSTR. Para los experimentos de estabilidad térmica, todas las transiciones fueron monitoreadas por el cambio de elipticidad a 222 nm con un ancho de banda de 1 nm y se utilizó un tiempo de integración de 32 segundos.

### 2.2.13 Modelamiento molecular, ensayos de docking y dinámica molecular

Los modelos de las proteínas dependientes de ADP fueron realizados en la conformación cerrada. Como templados utilizamos la estructura de ADP-GK de *P. furiosus* (1UA4) en conformación cerrada y la estructura de ADP-PFK de *P. horikoshii* (3DRW) la cual se encuentra en la conformación abierta. Para la construcción del alineamiento de las dos proteínas cortamos la estructura 3DRW en sus dominios mayor y menor y alineamos estructuralmente cada dominio a la estructura 1UA4 con STAMP implementado en Multiseq [61] de VMD [62]. El alineamiento de la secuencia resultante se inspeccionó visualmente y se editaron las posiciones mal alineadas. Este alineamiento y las tres estructuras resultantes se utilizaron para realizar el modelamiento por

homología con Modeller 8 [63]. Se escogieron los 10 modelos con mejor potencial DOPE y luego se eligió el mejor modelo de acuerdo a la evaluación con PROSA2003 [64], Procheck [65] y VERIF3D [66].

Los ensayos de docking con glucosa y fructosa-6P fueron realizados con Autodock Vina 1.0 [67]. El estado de protonación de los residuos ionizables fue calculado con el servidor web H++ [68]. La derivación de cargas parciales se realizó con el método Gasteiger implementado en AutodockTool [69]. Se seleccionaron los resultados con energías de interacción más bajas y con el grupo aceptor del fosforilo orientado hacia el motivo GXGD. El aspártico del motivo GXGD es considerado la base catalítica en la transferencia del grupo fosforilo en toda la superfamilia riboquinasa [70]. Luego, el complejo Mg-ADP fue puesto en el sitio activo por superposición, de acuerdo a lo descrito por Merino y cols [71]. Posteriormente, las estructuras fueron puestas en una caja de agua de 1.2 nm desde cada lado de la proteína y el sistema fue neutralizado con NaCl a una concentración final de 0,1M. La minimización de energía se realizó con 10.000 iteraciones sin restricciones con NAMD 2.7 [72] y utilizando el campo de fuerza CHARMM27 [73]. Para las simulaciones de dinámica molecular se realizó un equilibrado para ajustar el sistema a una temperatura y presión constante. El equilibrado del sistema se realizó en un ensamble isotérmico isobárico (NPT) con una temperatura de 300 K y presión de 1 atmósfera (1,01325 bar), el paso de tiempo empleado fue de 2 fs. Luego del equilibrado, se simularon los sistemas por 8 ns y se analizó la energía no enlazante compuesta por las interacciones electrostáticas y de Van



Der Waals a través de las trayectorias, utilizando NAMDenergy implementado en VMD [62].

#### 2.2.14 Cristalografía de proteínas

El *screening* de cristalización se realizó en el equipo “Honeybee 963 robot (Genomic Solutions)”, con los kit de cristalización Classic I, Classic II, PEG Suite I, PEG suite II de Quiagen y SaltRX de Hampton Research. Los ensayos de cristalización se realizaron por el método difusión de vapor por gota sentada en una mezcla 1:1 de las proteínas y las diversas soluciones de cristalización en placas Greiner CrystalQuick. Los cristales crecieron durante 72 horas a 18 °C. Las rondas de optimización para las condiciones en las que se obtuvieron cristales se realizaron por el método de gota colgante a la misma temperatura. Los cristales que fueron difractados se pasaron por solución de crioprotección (solución madre con 20% de glicerol) y se montaron en “cryo-loops” y se congelaron directamente en nitrógeno líquido. Los cristales se difractaron en el sincrotrón Diamon (Inglaterra), las reflexiones observadas se indexaron, integraron y escalaron mediante el programa Xia2 v0.3.4.0 [74]. Las fases se resolvieron mediante reemplazo molecular utilizando Phaser-2.5.5 [75] con la estructura de la PFK de *P. horikoshii* (PDB:3DRW) en la Plataforma CCP4 [76], en tanto que el refinamiento de *PhPFK* se realizó de manera iterativa mediante Coot [77] y Phenix 1.8.4 [78], la evaluación del modelo se realizó con MolProbity [79].

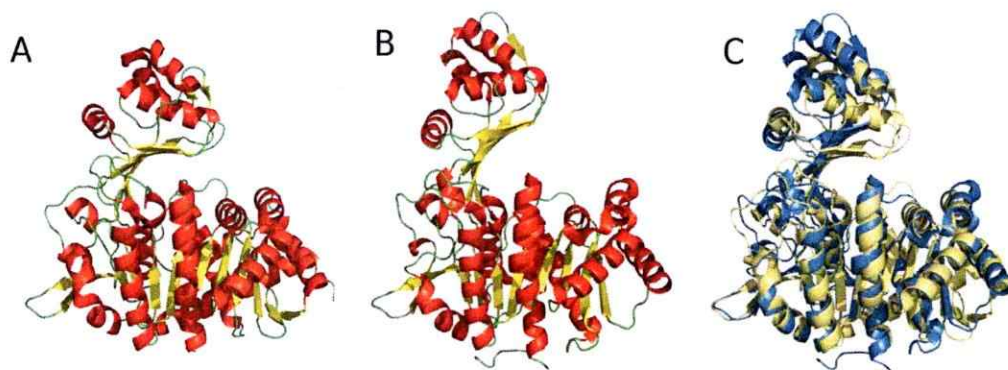
### **3 CAPÍTULO UNO**

**Historia evolutiva del uso de sustratos en la  
Familia de quinasas de azúcar dependientes  
de ADP de arqueas**

### 3.1 Introducción

Varias arqueas del reino *Euryarqueota* presentan una vía modificada de Embden-Meyerhof (EM) la cual solo involucra cuatro de las enzimas clásicas de la vía canónica. Una de las características más interesantes de esta vía modificada es que la fosforilación de glucosa y fructosa-6P (F6P) es realizada con ADP como dador del fosforilo, en vez de ATP, como en la vía clásica. Estas quinasas dependientes de ADP fosforilan glucosa (ADP-GKs) y fructosa-6P (ADP-PFKs) y son homólogas entre ellas, pero no muestran similitud de secuencia con ninguna quinasas dependiente de ATP conocida. Por esta razón, estas proteínas fueron inicialmente propuestas como una nueva familia de enzimas [80], pero posteriormente fueron clasificadas dentro de la superfamilia Riboquinasa, dada la semejanza estructural con los miembros de esta superfamilia [81].

En el dominio arquea las quinasas de azúcares dependientes de ADP. han sido encontradas en los géneros *Pyrococcus*, *Thermococcus*, *Methanosarcina*, *Methanosaeta*, *Methanococcoides*, *Methanococcus*, *Metanocaldococcus* y *Archaeoglobus* [59,80,82–84]. También ha sido posible identificar homólogos distantes de estas enzimas en genomas de eucariontes superiores, como los descritos en ratón y humano, los que se ha demostrado que son glucoquinasas dependientes de ADP [36,37].



**Figura 2. Comparación de estructuras de glucoquinasas y fosfofructoquinasas dependientes de ADP.**

A, ADP-GK de *T. litoralis*. B, ADP-PFK de *P. horikoshii*. C, Superposición de ADP-GK (amarillo) y ADP-PFK (azul).

En el banco de estructuras de proteínas (PDB) se encuentran las estructuras tridimensionales de la glucoquinasa dependiente de ADP (ADP-GK) de *Thermococcus litoralis* [81], *Pyrococcus horikoshii* (Tsuge et al., 2002), *Pyrococcus furiosus* [40] y de la fosfofructoquinasa dependiente de ADP (ADP-PFK) de *Pyrococcus horikoshii* [85], las que muestran una alta conservación de la estructura a pesar que entre las ADP-GKs y ADP-PFKs la identidad de secuencia no supera el 30%. En la **Figura 2** se muestra la comparación entre la estructura de la ADP-GK de *T. litoralis* con la ADP-PFK de *P. horikoshii*.

En literatura, las enzimas del grupo termófilo de los *Thermococales* son las mejor caracterizadas experimentalmente. Las glucoquinasas de *Thermococcus litoralis* (TIGK) y de *Pyrococcus furiosus* mostraron ser altamente específicas por glucosa. En cuanto a las fosfofructoquinasas, las provenientes de *Pyrococcus horikoshii* (PhPFK), *Thermococcus zilligii*, *Pyrococcus furiosus* y *Archaeoglobus fulgidus*, también mostraron ser altamente específicas, pero en este caso por fructosa-6P [70]. Todas las demás proteínas clasificadas como glucoquinasas o fosfofructoquinasas dependientes de ADP, han sido asignadas solo por su similitud de secuencia.

Una de las características más interesantes en esta familia de enzimas dependientes de ADP, es la presencia de una enzima con función dual o inespecífica denominada fosfofructoquinasa/glucoquinasa, la cual se encuentra presente en la arquea termófila *Methanocaldococcus jannaschii* (MjPFK/GK) [43]. Inicialmente, se propuso a esta enzima como un estado ancestral para toda esta familia, desde donde habrían surgido, después de un evento de duplicación génica, las dos especificidades conocidas

actualmente [43]. Esta hipótesis fue puesta a prueba por Merino y Guixé utilizando un método bayesiano de inferencia filogenética, el cual mostró que la enzima *Mj*PFK/GK no representa una forma ancestral de toda la familia, sino más bien el análisis filogenético muestra que es una enzima que se encuentra en el grupo de las PFK, pero que además de utilizar fructosa-6P, cataliza también la fosforilación de glucosa [86].

Si bien se descarta que esta enzima bifuncional sea un estado ancestral de la familia, aún se desconoce cuándo apareció este rasgo bifuncional y si esta es la única enzima de la familia que presenta esta característica.

### ***3.2 Hipótesis***

La última proteína ancestral en común para las glucoquinasas y fosfofructoquinasas dependientes de ADP de arqueas, es una enzima bifuncional capaz de usar tanto glucosa como fructosa-6P como aceptor del fosforilo, en tanto que las proteínas ancestrales de cada sub-familia son específicas para el respectivo sustrato aunque con una eficiencia catalítica menor que la de las enzimas actuales.

### ***3.3 Objetivo***

Reconstrucción y caracterización de enzimas ancestrales de la familia dependiente de ADP de la superfamilia Riboquinasa para establecer los determinantes moleculares que dictaron la especialización por el sustrato azúcar durante la evolución de esta familia.

### 3.4 Resultados

#### 3.4.1 Uso de sustratos en las enzimas actuales; TIGK, PhPFK, y MjPFK/GK

Como modelos de proteínas actuales hemos caracterizado la glucoquinasa de *Thermococcus litoralis* (TIGK), la fosfofructoquinasa de *Pyrococcus horikoshii* (PhPFK) y la enzima bifuncional PFK/GK de *M. jannaschii* (MjPFK/GK) (Tabla 1). Las enzimas TIGK y PhPFK son altamente específicas por glucosa y fructosa-6P respectivamente, no detectándose actividad con el otro sustrato aún a una concentración de 10 mM. Sin embargo, la enzima MjPFK/GK puede utilizar ambos sustratos aunque con una eficiencia catalítica ( $k_{cat}/K_m$ ) 50 veces mayor para fructosa-6P comparada con la obtenida para glucosa. Cabe destacar que esta es la única enzima de la familia descrita con esta propiedad. Además, evaluamos la capacidad de fructosa-6P de inhibir la actividad de la enzima TIGK, no encontrándose alteración de la actividad aun a una concentración de 50 mM, lo que sugiere que esta enzima no es capaz de unir fructosa-6P a las concentraciones evaluadas. Tampoco se observó inhibición significativa de la actividad PFK de PhPFK con 100 mM de glucosa, aunque ensayos realizados sobre este rango muestran una pequeña inhibición con un IC<sub>50</sub> de aproximadamente 200 mM, la que pudiese deberse a la unión de glucosa al sitio activo o a efectos de osmolaridad considerando las concentraciones utilizadas. En cualquier caso, los ensayos muestran que TIGK y PhPFK son altamente específicas por sus respectivos sustratos y que no unen el sustrato contrario en rangos de concentraciones que se consideran fisiológicos para estos azúcares (rango  $\mu$ M para fructosa-6P y unidades de mM para glucosa).



**Tabla 1. Constantes cinéticas para las proteínas actuales**

	Fructosa-6P		Glucosa		Fructosa-6P	Glucosa
	$K_m(\mu\text{M})$	$k_{\text{cat}}(\text{s}^{-1})$	$K_m(\text{mM})$	$k_{\text{cat}}(\text{s}^{-1})$	$k_{\text{cat}} / K_m$ ( $\text{M}^{-1}\text{s}^{-1}$ )	$k_{\text{cat}} / K_m$ ( $\text{M}^{-1}\text{s}^{-1}$ )
<i>TlGK</i> <sup>†</sup>	---	---	200±10	43±1	---	2.1 x10 <sup>5</sup>
<i>PhPFK</i> <sup>*</sup>	15±2	41±3	---	---	2.7 x10 <sup>6</sup>	---
<i>MjPFK/GK</i> <sup>*</sup>	5±3	3±2	0.78±0.09	9.6±0.3	6.0 x10 <sup>5</sup>	1.2 x10 <sup>4</sup>

<sup>†</sup>valores determinados a 40°C, \*valores determinados a 50 °C.

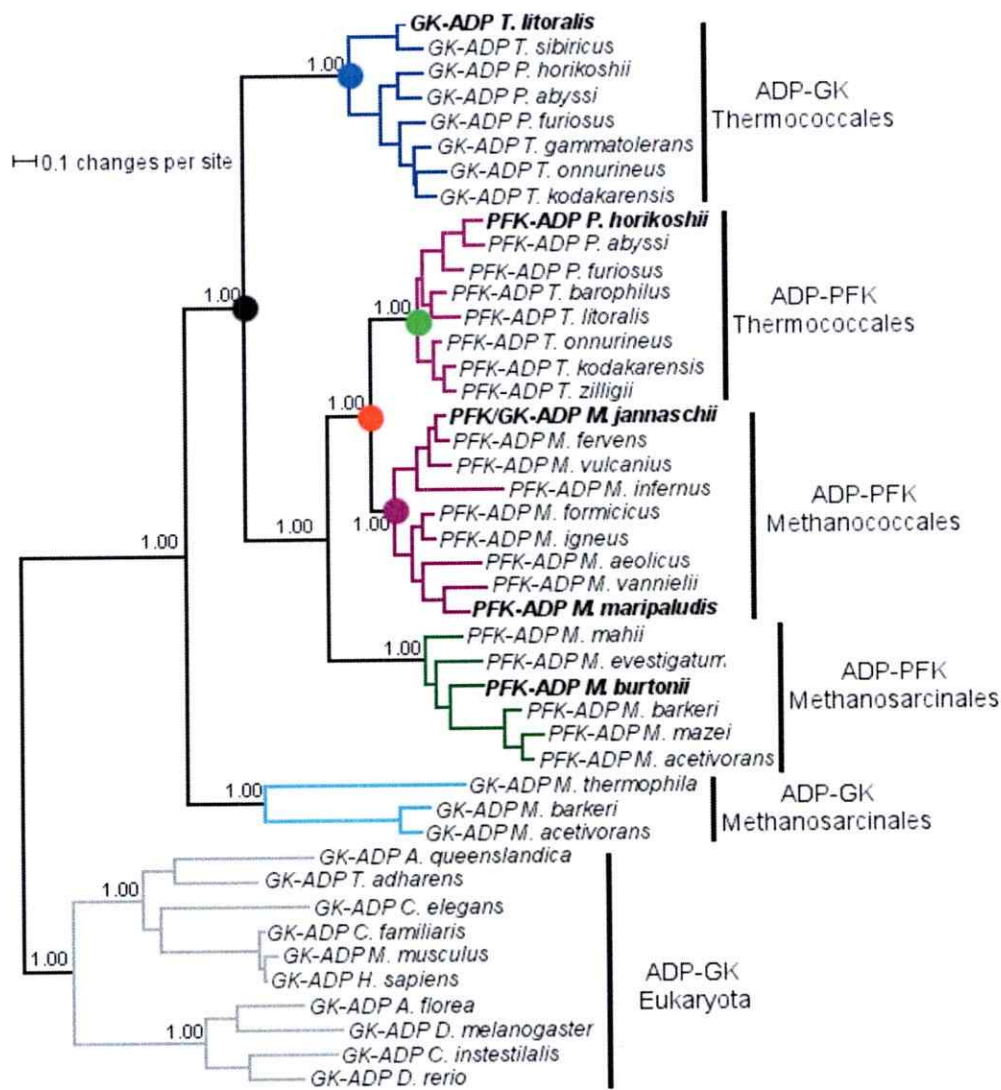
--- Actividad no detectada.

### 3.4.2 Reconstrucción filogenética por inferencia bayesiana y máxima verosimilitud

Hemos reconstruido la filogenia de esta familia mediante inferencia bayesiana tanto para secuencias de proteínas (**Figura 3**) como para secuencias de genes (**Figura 4**). Dada la incerteza de los métodos filogenéticos, también se reconstruyó la filogenia de esta familia por el método de máxima verosimilitud (ML) (**Figura 5**). El método de máxima verosimilitud encuentra el árbol que maximiza la probabilidad de observar los caracteres en el alineamiento. Por otra parte, la inferencia bayesiana calcula directamente la probabilidad posterior desde poblaciones de árboles, dado los caracteres observados en el alineamiento, y considera además la incerteza de la reconstrucción al calcularla desde poblaciones de árboles.

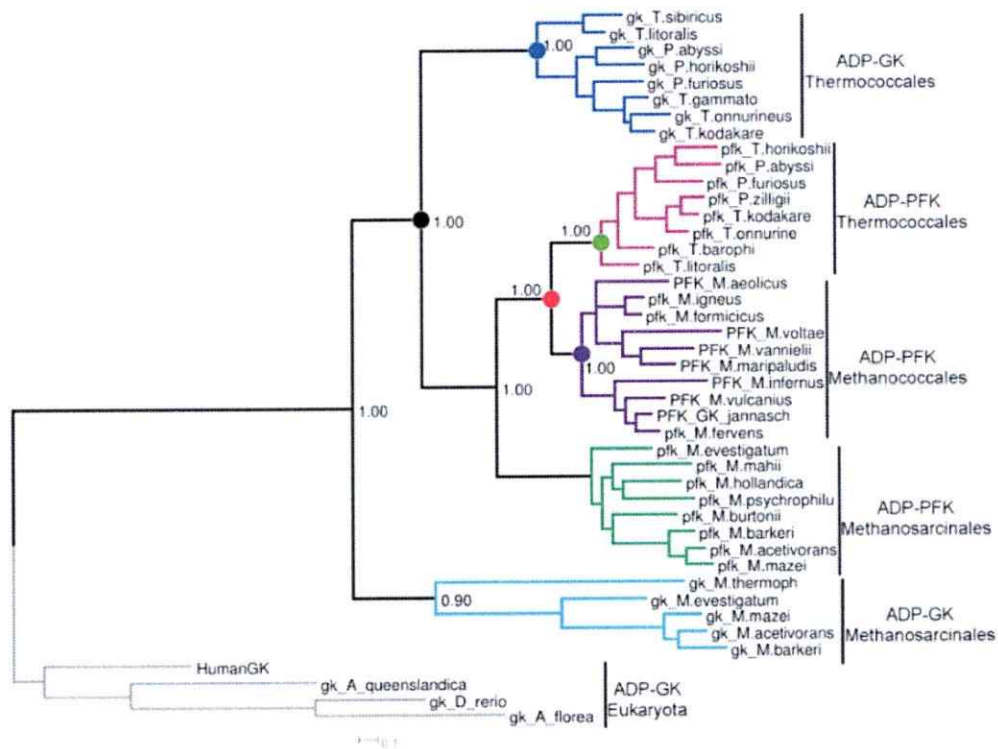
La reconstrucción por máxima verosimilitud (**Figura 5**) muestra las mismas agrupaciones que las reconstrucciones bayesianas, lo cual muestra el buen soporte que tiene la topología del árbol y los nodos utilizados para realizar la inferencia de las secuencias ancestrales.

La búsqueda de secuencias fue realizada el segundo semestre de 2010 encontrándose secuencias de GKs dependientes de ADP solo en arqueas y eucariontes. En cuanto a las secuencias de PFKs solo fueron encontradas en arqueas. En el dominio arqueas se encontraron GKs en los órdenes *Thermococcales* y *Methanosarcinales* y las secuencias de PFKs fueron encontradas en los órdenes de *Thermococcales*, *Methanosarcinales* y *Methanococcales*.



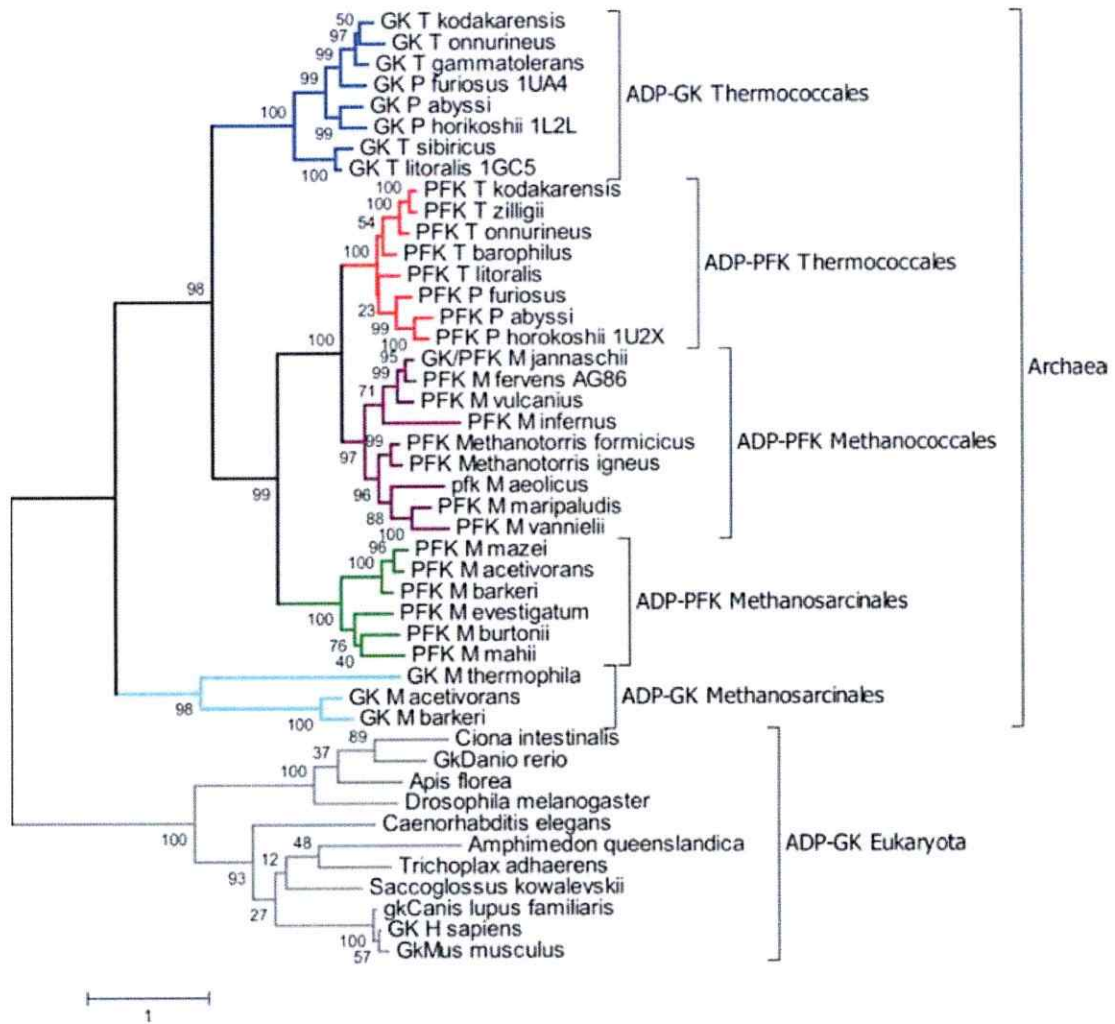
**Figura 3. Árbol filogenético consenso obtenido por inferencia bayesiana para secuencias de proteínas.**

En color se muestran las enzimas de los siguientes ordenes de arqueas; GKs de *Thermococcales* (azul), PFKs de *Thermococcales* (rosado), PFKs de *Methanococcales* (morado), PFKs de *Methanosarcinales* (verde) y ADP-GKs de *Methanosarcinales* (cian). Se muestran en color los nodos correspondientes a; ancGK/PFK (negro), ancGK (azul), ancPFK-MT (rojo), ancPFK-T (verde) y ancPFK-M (morado). La probabilidad posterior para los grupos de interés se muestra en los respectivos nodos. Las secuencias de eucariontes (gris) fueron utilizadas como grupo externo. En negrita se muestran las enzimas actuales estudiadas.



**Figura 4. Reconstrucción filogenética consenso obtenida por inferencia bayesiana para secuencias de ADN.**

La probabilidad posterior para los grupos de interés se muestra en los respectivos nodos. Se muestran en color los nodos correspondientes a; ancGK/PFK (negro), ancGK (azul), ancPFK-MT (rojo), ancPFK-T (verde) y ancPFK-M (morado).



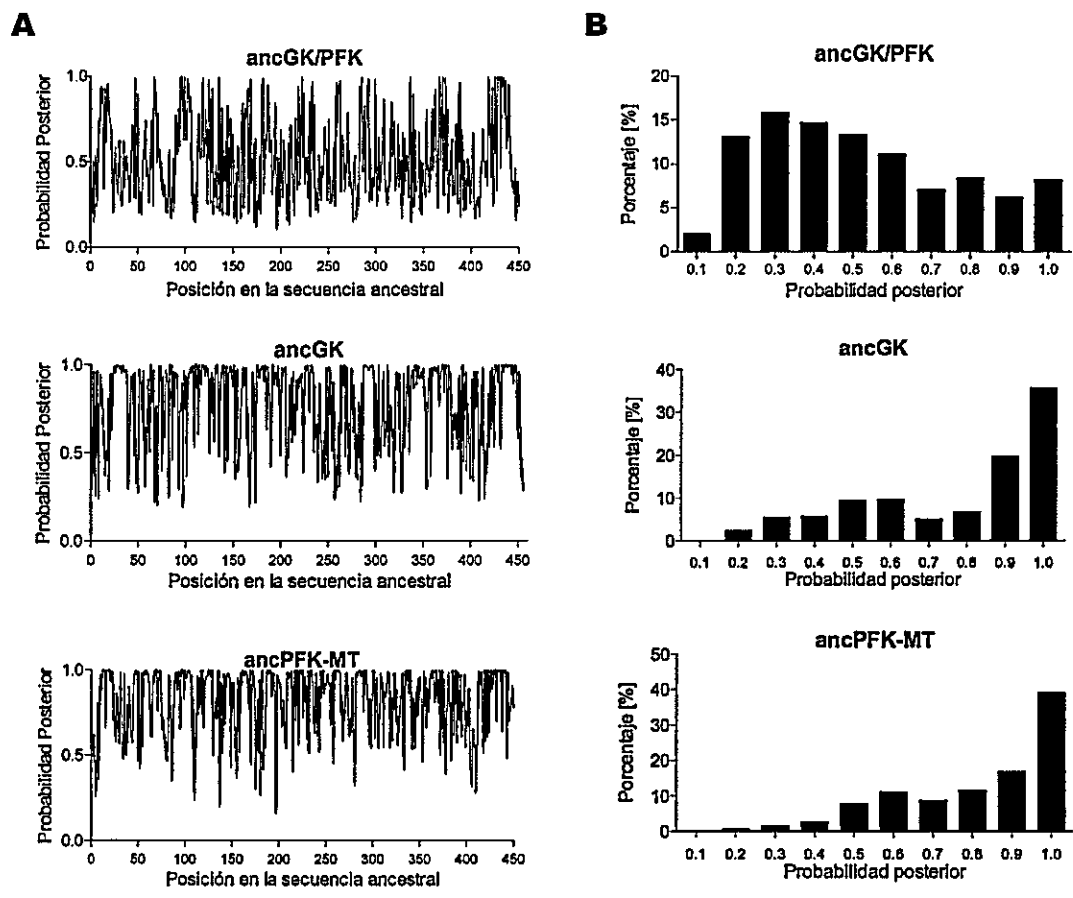
**Figura 5. Reconstrucción filogenética obtenida por máxima verosimilitud para secuencias de proteínas.**

Se utilizó el modelo Cprev+I+G de acuerdo a la selección de modelos por el factor de información de akaike (AIC). Los números de las ramas representan los valores de bootstrap calculados utilizando MEGA (versión 6.06) con 1000 replicados. Código de colores como en figura 4.

### 3.4.3 Reconstrucción de secuencias ancestrales de la familia de quinasas dependientes de ADP

Utilizando el método bayesiano jerárquico y el alineamiento de genes inferimos las siguientes secuencias ancestrales; último ancestro en común entre GKs y PFKs (ancGK/PFK) nodo coloreado en negro; ancestro de las ADP-GKs de *Thermococcales* (ancGK) nodo en color azul, y último ancestro en común para las ADP-PFKs de *Methanococcales* y *Thermococcales* (ancPFK-MT) nodo rojo (**Figura 3**).

En la **Figura 6** se muestra un análisis de la inferencia para las tres secuencias. La distribución de probabilidades mostró que la secuencia de ancGK presenta un 86% de residuos con una probabilidad posterior superior a 0.5, en tanto que para el caso de ancPFK-MT se encontró un 93% de los residuos con esta probabilidad. Para el ancestro ancGK/PFK solo el 54% de los residuos presentan una probabilidad sobre 0.5. El promedio de probabilidad (y desviación estándar) para las secuencias son  $0.77\pm 0.24$ ,  $0.80\pm 0.21$  y  $0.52\pm 0.25$  respectivamente.



**Figura 6. Inferencia de las secuencias ancestrales ancGK/PFK, ancGK y ancPFK-MT.**  
**A**, Probabilidad posterior del residuo respecto a la posición en la secuencia ancestral. **B**, Histogramas de distribución de frecuencias de las probabilidades posteriores de cada secuencia. La probabilidad posterior representa la probabilidad que tiene un residuo (codon) de haber estado en esa posición en la secuencia ancestral, dados los codones (nucleotidos) que están en esa posición en las secuencias actuales de acuerdo al modelo evolutivo utilizado y la población de arboles que explica la filogenia.

#### 3.4.4 Análisis de la robustez de las reconstrucciones de las secuencias ancestrales.

Una manera de analizar la robustez de las reconstrucciones ancestrales es utilizar distintos métodos como el método de ML (Bayes empírico) o método de Bayes jerárquico con modelos de evolución para ADN y para aminoácidos y comparar las reconstrucciones.

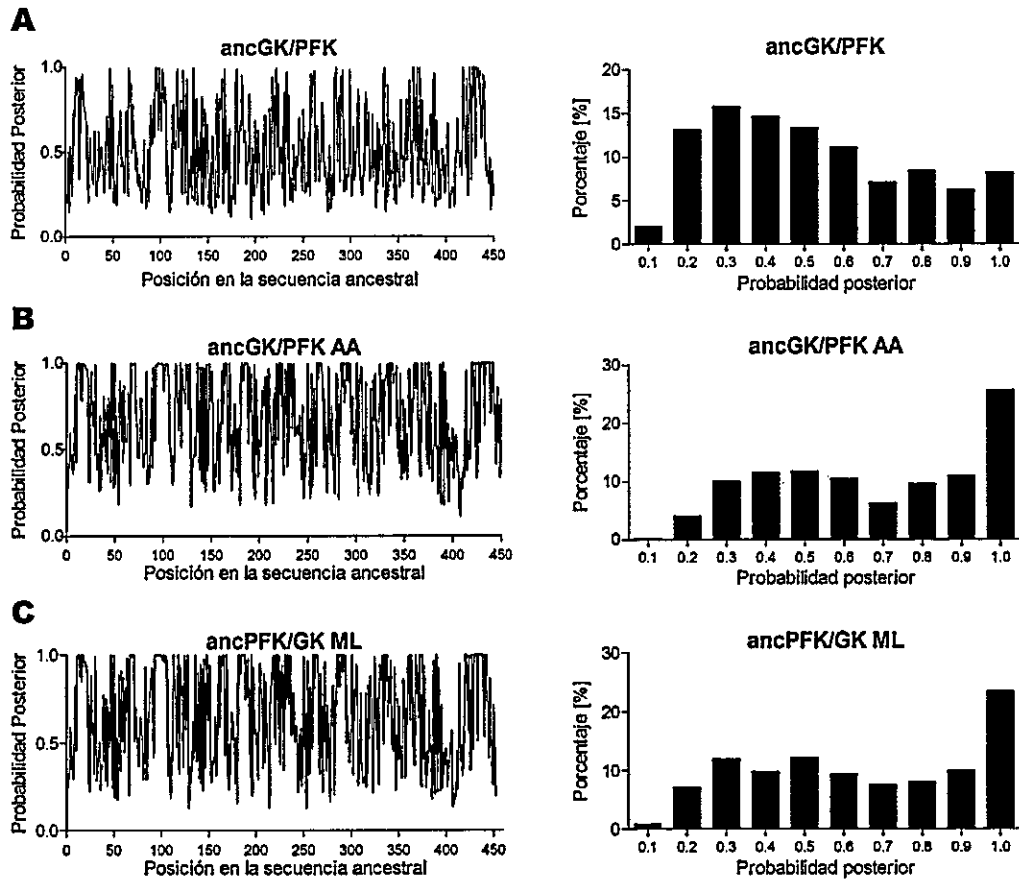
Dentro de las secuencias reconstruidas por el método de Bayes jerárquico utilizando el modelo de evolución GTR+I+G para secuencias de ADN, el último ancestro en común entre glucoquinasas (GK) y fosfofructoquinasas (PFK) (ancestro ancGK/PFK) mostró el menor promedio de probabilidad posterior por residuo ( $0.52 \pm 0.25$ ) (**Figura 6A**) (**Tabla 2**), quizás debido a la profundidad del nodo. Para analizar la confiabilidad en la reconstrucción de los residuos del sitio activo y por consiguiente en la especificidad por el sustrato azúcar, reconstruimos el ancGK/PFK por Bayes jerárquico utilizando el modelo de Markov de tiempo continuo para aminoácidos (Cprev+I+G), en el cual las velocidades de sustitución son fijas a un valor específico (sec AA, **Figura 7B**). Se escogió este modelo porque mostró una probabilidad posterior de 1.00 respecto a los otros modelos (Poisson, Jones, Dayhoff, Mtrev, Mtmam, Rtrev, Vt y Blosum) en un análisis de MCMC (Markov Chain Monte Carlo) en MrBayes 3.1.2. También se reconstruyó esta secuencia mediante el método de Bayes empírico (o también denominado máxima verosimilitud) utilizando el modelo Cprev+I+G en el programa PAML 4.4 (ML, **Figura 7C**).

En la **Tabla 2** se muestran el promedio y mediana de las distribuciones de frecuencias de la **Figura 7**. La secuencia por Bayes jerárquico/sec AA muestra mejores probabilidades



en promedio que las otras dos secuencias, pero este mayor valor puede estar sesgado por el modelo de evolución con velocidades de sustitución fijas.

Ya que el principal objetivo de las reconstrucciones es analizar la especificidad por sustrato, hemos comparado los residuos del sitio activo para el sustrato azúcar en las tres secuencias. La comparación se muestra en la **Tabla 3** donde se aprecia que todos los residuos del sitio activo son idénticos en todos los métodos. Esto muestra que independiente de las bajas probabilidades obtenidas por el método jerárquico/sec genes los residuos del sitio activo se sustentan en todas las reconstrucciones y por ende la especificidad por el sustrato azúcar no estaría sesgada por del método de reconstrucción utilizado.



**Figura 7. Reconstrucción de ancPFK/GK por distintos métodos.**

A, inferencia utilizando bayes jerárquico y secuencias de genes. B, inferencia utilizando bayes jerárquico y secuencias aminoácidos. C, inferencia utilizando bayes empírico y secuencias de aminoácidos.

**Tabla 2. Probabilidades posteriores según el método de reconstrucción de ancGK/PFK**

Metodología	Promedio de PP ( $\pm$ SD)	Mediana de PP
Bayes jerárquico/ sec. genes	0.52 $\pm$ 0.25	0.47
Bayes jerárquico / sec AA	0.67 $\pm$ 0.26	0.69
Bayes empírico-ML / sec AA	0.64 $\pm$ 0.27	0.63

**Tabla 3. Probabilidades posteriores de residuos del sitio activo de ancGK/PFK de acuerdo al método de reconstrucción**

Jerárquico Genes	PP	Jerárquico AA	PP	Empírico-ML AA	PP
N30	0,92	N13	1,00	N13	1,00
N32	0,93	N15	1,00	N15	1,00
D34	0,93	D17	1,00	D17	1,00
E88	0,78	E71	1,00	E71	1,00
G112	1,00	G96	1,00	G96	1,00
Q113	0,84	Q97	0,99	Q97	0,99
I116	0,79	I100	1,00	I100	1,00
H176	0,84	H161	0,96	H161	0,99
I178	0,82	I163	1,00	I163	0,99
R201	0,62	R186	1,00	R186	1,00
I203	0,82	I188	1,00	I188	1,00
N210	0,84	N195	0,99	N195	1,00
Q239	0,90	Q224	0,99	Q224	0,99
V447	1,00	V431	1,00	V431	1,00
G448	1,00	G432	1,00	G432	1,00
D451	1,00	D435	1,00	D435	1,00
<b>Promedio</b>	<b>0,88</b>		<b>1,00</b>		<b>1,00</b>

#### 3.4.5 Expresión y purificación de las enzimas ancestrales; ancGK/PFK, ancGK y ancPFK-MT

Los genes de las proteínas ancestrales fueron enviados a sintetizar y optimizados para la expresión en *Escherichia coli*, utilizando el servicio de GeneScript USA. Las tres proteínas ancestrales se purificaron, de acuerdo a lo descrito en materiales y métodos; la proteína ancestral ancGK/PFK presentó un bajo rendimiento en la purificación de solo 1.8 mg de proteína no homogénea a partir de 1 litro de cultivo. Para la proteína ancGK se obtuvieron 5 mg de proteína purificada por litro de cultivo y para ancPFK-MT se obtuvieron alrededor de 70 mg de proteína por litro de cultivo. La pureza se evaluó mediante electroforesis denaturante SDS-PAGE.

#### 3.4.6 Caracterización cinética de las enzimas ancestrales; ancGK/PFK, ancGK y ancPFK-MT

La caracterización cinética de los ancestros con respecto al uso de glucosa y fructosa-6P como sustratos aceptores de fosforilo se muestra en la **Tabla 4**. Los ensayos de la actividad glucoquinasa y fosfofructoquinasa y las posteriores curvas de saturación de sustrato, mostraron que ancGK/PFK es capaz de utilizar ambos sustratos con constantes catalíticas muy inferiores a las enzimas actuales, pero con eficiencias catalíticas similares para ambos sustratos. Como se esperaba, el ancestro ancGK solo utilizó glucosa como sustrato e interesantemente, el ancestro de las PFKs de *Thermococales* y *Methanococales* (ancPFK-MT) fue capaz de utilizar ambos sustratos, sin preferencia significativa por el uso de glucosa o fructosa-6P.

**Tabla 4. Constantes cinéticas de las proteínas ancestrales; ancGK/PFK, ancGK y ancPFK-MT**

	F6P		Glucosa		F6P	Glucosa	Rasgo
	$K_m(\mu\text{M})$	$k_{\text{cat}}(\text{s}^{-1})$	$K_m(\text{mM})$	$k_{\text{cat}}(\text{s}^{-1})$	$k_{\text{cat}} / K_m$ ( $\text{M}^{-1}\text{s}^{-1}$ )	$k_{\text{cat}} / K_m$ ( $\text{M}^{-1}\text{s}^{-1}$ )	
ancGK/PFK	29±3	0,05 <sup>†</sup>	0,67±0,1	0,48 <sup>†</sup>	2,8 x 10 <sup>1</sup>	1,2 x 10 <sup>1</sup>	B
ancGK	---	---	0,8±0,1	8,3±0,3	---	1,0 x 10 <sup>4</sup>	E
ancPFK-MT	1646±284	0,25±0,01	19±2	3,5±0,1	1,5 x 10 <sup>2</sup>	1,8 x 10 <sup>2</sup>	B

†, valores en  $\text{min}^{-1}$ . Constantes determinadas a 30 °C. --- Actividad no detectada.

B: Bifuncional

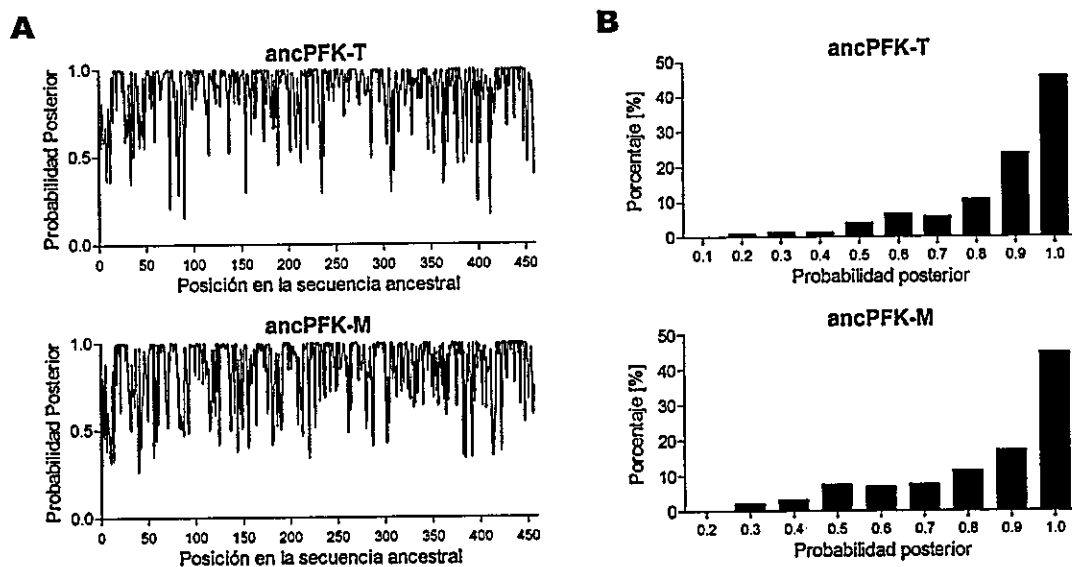
E: Especifico

### 3.4.7 Inferencia de las secuencias de los ancestros ancPFK-T y ancPFK-M

La capacidad del ancPFK-MT de utilizar ambos sustratos fue totalmente inesperada e indica que la característica dual del ancestro ancGK/PFK se conservó hasta el ancestro PFK-MT. Para buscar donde se originó la alta especificidad por fructosa-6P mostrada por las PFKs actuales y rastrear hasta donde se ha conservado la característica dual decidimos inferir la secuencia del último ancestro en común de las PFK de *Thermococcales* (ancPFK-T) y el último ancestro de las PFKs de *Methanococcales* (ancPFK-M). Se obtuvo un promedio de probabilidad de  $0.86 \pm 0.17$  para ancPFK-T y  $0.84 \pm 0.19$  para ancPFK-M. Ambas reconstrucciones muestran un soporte robusto respecto al residuo más probable (**Figura 8**).

### 3.4.8 Resurrección experimental de ancPFK-T y ancPFK-M

Se sintetizaron los genes correspondientes a ancPFK-T y ancPFK-M, expresamos y, purificamos las proteínas ancestrales y caracterizamos su uso de sustratos (**Tabla 5**). La caracterización cinética de ancPFK-T mostró que a concentraciones consideradas fisiológicas de los sustratos evaluados, sólo es capaz de utilizar fructosa-6P como sustrato aceptor de fosforilo, el  $K_m$  estimado para glucosa está sobre 1M, lo que hace imposible que esta enzima pueda unir adecuadamente glucosa a concentraciones fisiológicas de este azúcar. Por otra parte, el ancestro ancPFK-M mostró actividad con ambos sustratos y su caracterización cinética (**Tabla 5**) dejó en evidencia la capacidad de utilizar ambos, por lo que el rasgo bifuncional se mantuvo hasta este nodo y sustenta la idea que las enzimas que descienden de este ancestro (*Methanococcales*) serían también bifuncionales.



**Figura 8. Análisis de la inferencia de ancPFK-T y ancPFK-M.**

A, Probabilidad posterior del residuo respecto a la posición en la secuencia ancestral. B, Histogramas de distribución de frecuencias de las probabilidades posteriores de cada secuencia.

**Tabla 5. Constantes cinéticas para ancPFK-T y ancPFK-M a 30°C**

	Fructosa-6P		Glucosa		Fructosa-6P	Glucosa	Rasgo
	$K_m(\mu\text{M})$	$k_{cat}(\text{s}^{-1})$	$K_m(\text{mM})$	$k_{cat}(\text{s}^{-1})$	$k_{cat} / K_m$ ( $\text{M}^{-1}\text{s}^{-1}$ )	$k_{cat} / K_m$ ( $\text{M}^{-1}\text{s}^{-1}$ )	
ancPFK-T	18±1	23,8±0,2	(1244*)	(14*)	1,3×10 <sup>6</sup>	9,7†	E
ancPFK-M	364±20	4,20±0,05	2,2±0,2	6,38±0,24	1,1×10 <sup>4</sup>	2,9×10 <sup>3</sup>	B

\*Valores estimados desde gráficos de doble recíproco.

†Valor determinado desde la pendiente inicial de la curva de saturación.

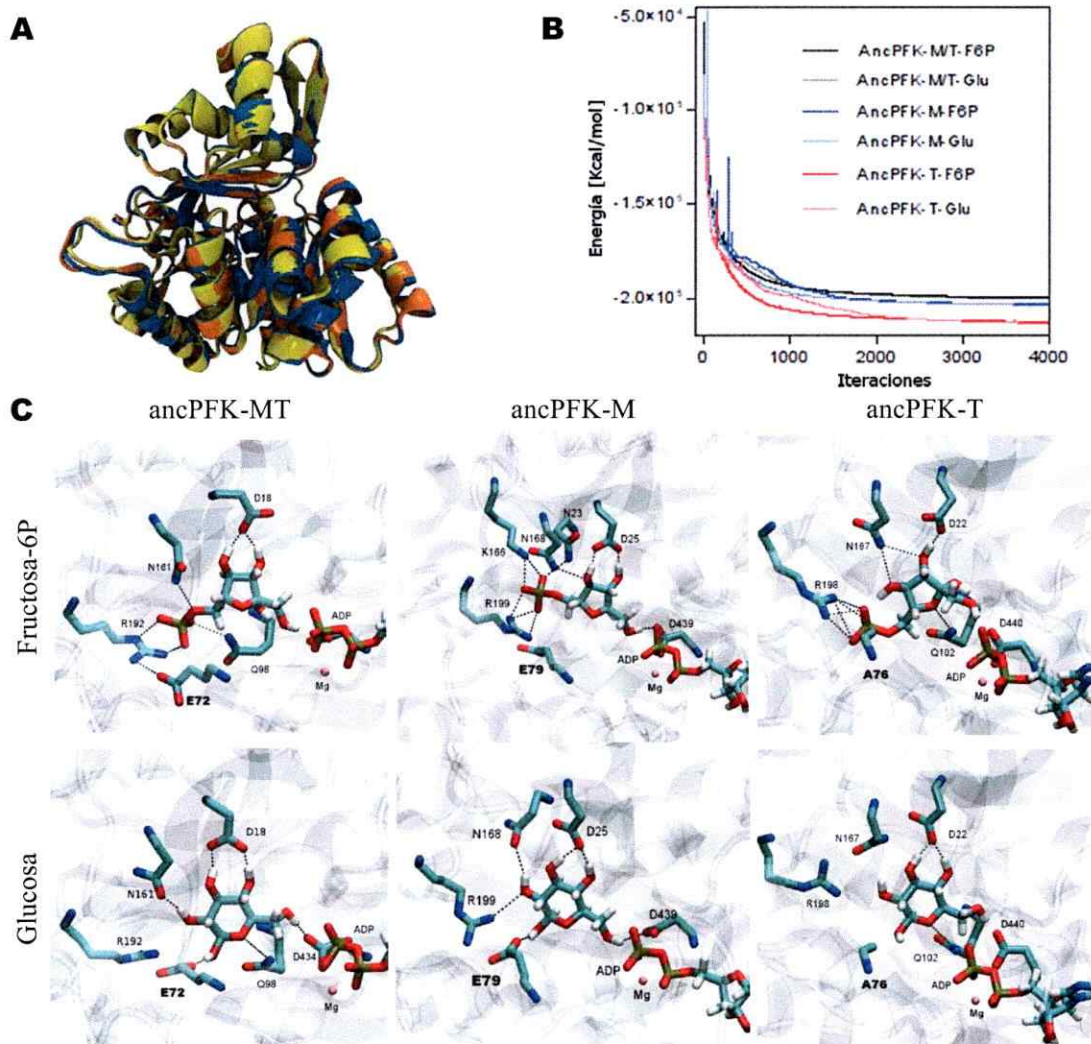
E: Especifico

B: Bifuncional

#### 3.4.9 Modelamiento y dinámica molecular de las interacciones entre proteína y ligando en los ancestros PFKs

Dado el interesante resultado experimental arrojado por el ancestro ancPFK-MT, el cual utiliza glucosa y fructosa como aceptor de fosforilo, realizamos modelos por homología de los tres ancestros PFKs correspondientes a: ancPFK-MT, ancPFK-T y ancPFK-M. Además realizamos *docking* con ambos sustratos y una minimización de energía utilizando el campo de fuerza CHARMM27, para permitir el acomodo adecuado de ambos sustratos (**Figura 9**). Los ensayos de *docking*, después de la minimización de energía, mostraron que en los ancestros ancPFK-MT y ancPFK-M, ambos sustratos se unen de manera prácticamente idéntica. En tanto, en el ancestro ancPFK-T sólo fructosa-6P presenta interacciones con todos sus grupos hidroxilos, mientras glucosa sólo presenta 3 puentes de hidrógeno, lo que indicaría que la unión de este sustrato al sitio activo podría ser débil. Esta falta de interacciones con glucosa podría deberse a la presencia de A76 en el sitio activo de ancPFK-T, ya que en los otros dos ancestros se encuentra un ácido glutámico estructuralmente equivalente (en ancPFK-MT E72 y en ancPFK-M E79), el que muestra una clara interacción con el hidroxilo del C1 de glucosa.

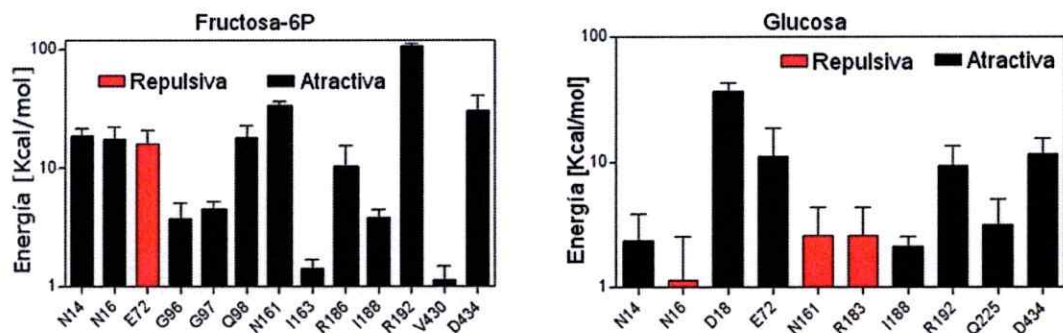




**Figura 9. Modelamiento e interacción proteína ligando de los ancestros PFKs.**

A, Superposición de los modelos por homología de las enzimas ancestrales; ancPFK-MT (azul), ancPFK-M (naranja), ancPFK-T (amarillo). B, minimización de energía de los modelos de las proteínas ancestrales luego del docking con  $\beta$ -D-glucosa y  $\beta$ -D-Fructosa-6P. Los ensayos de docking fueron realizados con autodock vina y las minimizaciones se realizaron con NAMD 2.8 utilizando el campo de fuerza CHARMM27. C, interacciones para  $\beta$ -D-Fructosa-6P y  $\beta$ -D-glucosa con las enzimas ancestrales después del procedimiento de minimización de energía.

Para estudiar la dinámica del sitio activo del ancestro ancPFK-MT y analizar las interacciones atractivas y repulsivas que presentan los residuos con glucosa y fructosa-6P, simulamos por 8 ns la proteína en complejo con ambos azúcares y Mg-ADP. Con esta simulación se realizó un análisis de la energía de interacción no enlazante entre los residuos del sitio activo y los azúcares (Figura 10). El análisis mostró que el residuo E72 presenta una gran repulsión en la interacción con fructosa-6P, mientras que en la interacción con glucosa este residuo presenta una de las mayores energías atractivas. Para glucosa, los residuos N161 y R183 también presentaron energías repulsivas significativas.

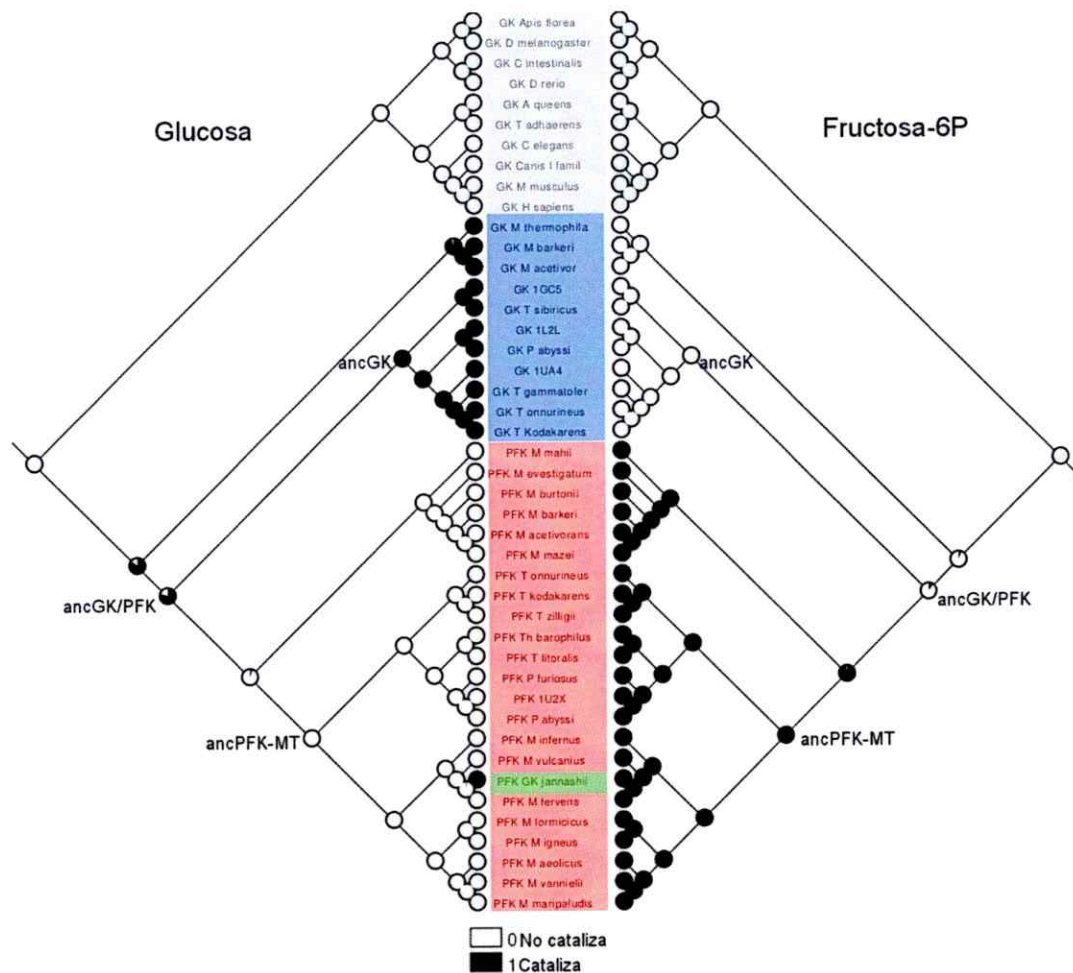


**Figura 10. Simulaciones de dinámica molecular de la enzima ancestral ancPFK-MT.**

Se simuló por 8 ns el complejo Mg-ADP con  $\beta$ -D-fructosa-6P y  $\beta$ -D-glucosa. Se calculó el promedio de la energía no enlazante para todos los residuos con al menos un átomo dentro de un radio de 5 Å del sustrato. Las interacciones que promediaron valores positivos de energía se muestran en rojo y en negro se muestran los promedios de valores negativos. Ambas simulaciones fueron realizadas con el campo de fuerza CHARMM27 y el software NAMD 2.8.

#### 3.4.10 Reconstrucción de estados ancestrales discretos para el uso de sustratos

Para analizar de manera general la historia evolutiva que ha tenido el uso de los dos sustratos (glucosa y fructosa-6P) e intentar explicar por qué el ancestro ancPFK-MT es bifuncional, inferimos la utilización de ambos sustratos como rasgos discretos en los estados ancestrales de la familia en arqueas, basándonos en la información actual del uso de fructosa-6P y glucosa por las enzimas actuales, la cual considera que todas las enzimas PFKs son específicas por F6P y que solo la enzima de *M. jannaschii*, es capaz de fosforilar ambos sustratos. El análisis se realizó con el software Mesquite mediante la aproximación de verosimilitud con el modelo MK1 para caracteres discretos binarios, se utilizó el árbol consenso obtenido de la reconstrucción filogenética bayesiana de genes (Figura 11). De acuerdo al análisis, el ancestro ancGK presenta un 99,9% de probabilidad para el uso de glucosa y solo un 0,001% para el uso de fructosa-6P, lo cual corresponde al uso experimental de sustrato por parte del ancGK. Por otra parte, el ancestro ancGK/PFK tiene una probabilidad de 79,7% de haber utilizado glucosa y solo un 10,0% de haber utilizado fructosa-6P. De acuerdo a esto ancPFK-MT debería ser específico por fructosa-6P ya que presenta un 99,9% de probabilidad para el uso de fructosa-6P y tan solo un 0,001% para el uso de glucosa. Sin embargo, este ancestro presentó actividad con ambos sustratos. Claramente el análisis no es capaz de reconstruir la verdadera historia evolutiva del uso de sustratos en la familia con la información actual recogida de la literatura. La incongruencia entre el análisis teórico y los resultados experimentales de las enzimas ancestrales se puede deber a que el uso de sustratos en las enzimas actuales no es correcto por la falta de caracterización de enzimas o que el modelo evolutivo para rasgos discretos es muy simple.

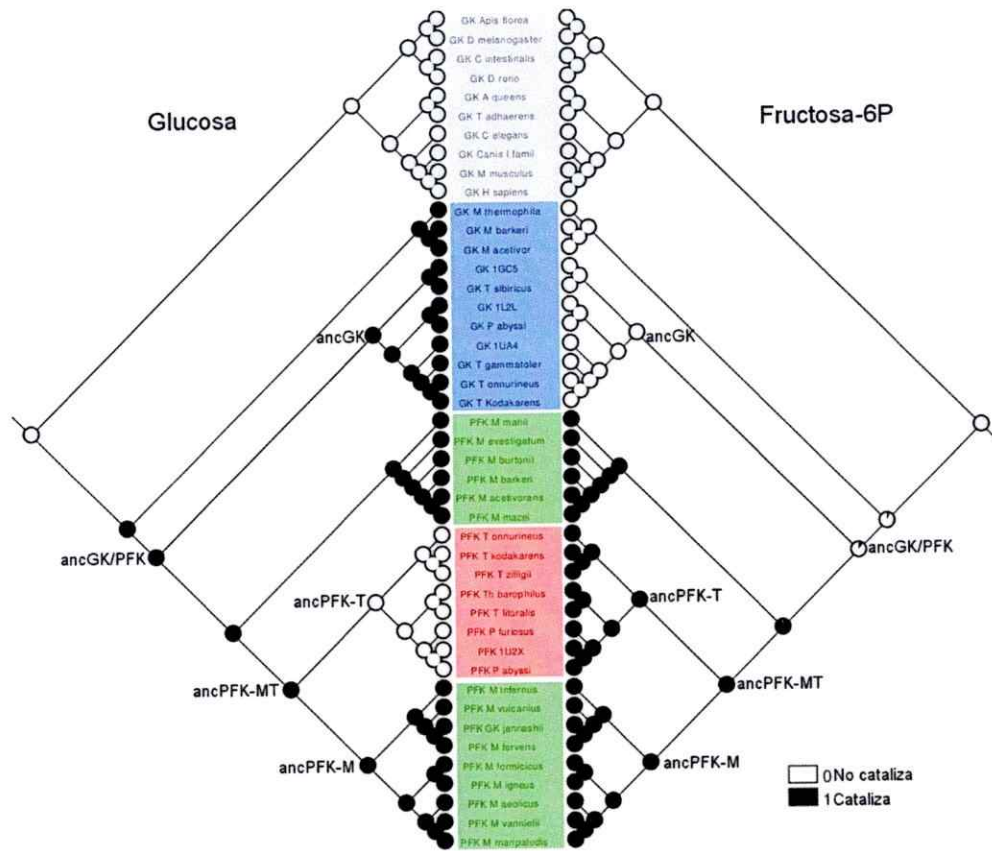


**Figura 11. Análisis de la historia del uso de glucosa y fructosa-6P como rasgo discreto durante la historia evolutiva de la familia.**

Izquierda, uso de glucosa como aceptor del fosforilo, derecha, uso de fructosa-6P. La proteína de *M. Jannaschii* (verde) es la única descrita hasta la fecha con la capacidad de utilizar ambos azúcares. Glucoquinasas (azul), fosfofructoquinasas (rojo) y en gris las GKs de eucariontes utilizadas como grupo externo y no consideradas en el análisis. Los nodos se representan como círculos color negro y blanco con áreas proporcionales a las probabilidades de catálisis (negro) o no catálisis (blanco).

Considerando los resultados de la interacción proteína ligando en los ancestros PFKs, analizamos la conservación del residuo glutámico, el que sería el responsable de la unión de glucosa en el ancestro ancPFK-MT (E72) y al analizar la secuencia de la única proteína actual caracterizada como bifuncional (*M. jannaschii*), encontramos que esta proteína presenta un ácido glutámico en aquella posición (E82). Además, encontramos que en el alineamiento múltiple de secuencias de toda la familia, este residuo se conserva prácticamente en todas las enzimas actualmente definidas como PFKs (*Methanococcales* y *Methanosarcinales*), excepto en las PFKs de *Thermococcales*, las cuales presentan una alanina en esa posición, la cual se habría conservado desde el ancestro ancPFK-T. Las enzimas PFKs cuya caracterización experimental ha sido reportada en la literatura mostraron ser altamente específicas por fructosa-6P [70], (PFK de *Pyrococcus horikoshii*, PFK de *Thermococcus zilligii* y PFK de *Pyrococcus furiosus*), al igual que lo corroborado por nuestro grupo con la proteína de *P. horikoshii* (Tabla 1). La excepción hasta la fecha es la enzima de *M. jannaschii*, la cual es bifuncional ya que utiliza glucosa y fructosa-6P. En su gran mayoría, todas las enzimas PFKs caracterizadas son de organismos del orden *Thermococcales*. De acuerdo a estos antecedentes propusimos que todas las enzimas de *Methanococcales* y de *Methanosarcinales*, actualmente anotadas como PFKs, pueden ser enzimas bifuncionales capaz de utilizar ambos sustratos. Con esta hipótesis hemos reevaluado la historia evolutiva basada en rasgos discretos, considerando en el análisis que todas las enzimas PFKs de *Methanococcales* y de *Methanosarcinales* son capaces de utilizar glucosa y fructosa-6P, mientras que las GKs son específicas por su sustrato al igual que las PFKs de *Thermococcales* (Figura 12).





**Figura 12. Análisis de la historia del uso de glucosa y fructosa-6P como rasgo discreto durante la historia evolutiva de la familia reconsiderando el uso actual de sustratos.**

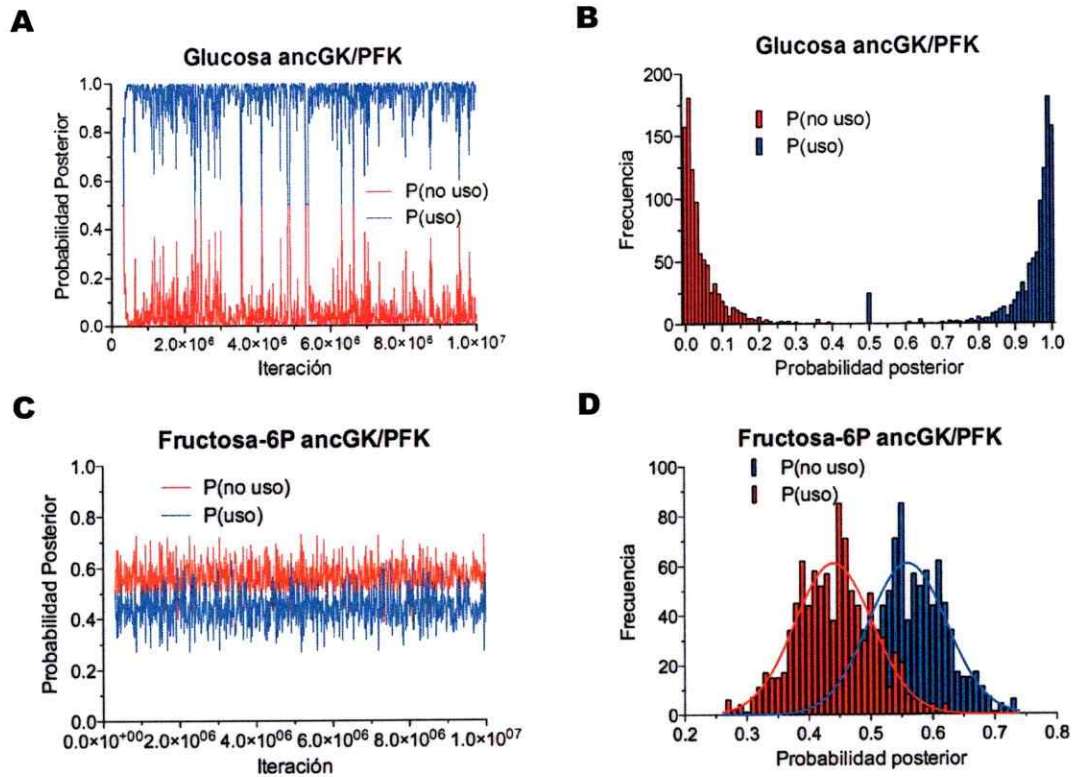
Izquierda, uso de glucosa como aceptor del fosforilo, derecha, uso de fructosa-6P. Uso sólo de glucosa (azul), uso sólo de fructosa-6P (rojo), uso de ambos azúcares (verde) y en gris las GKs de eucariontes utilizadas como grupo externo y no consideradas en el análisis. Los nodos se representan como círculos color negro y blanco con áreas proporcionales a las probabilidades de catálisis (negro) o no catálisis (blanco).

El análisis mostró que el ancestro ancGK tiene una probabilidad casi absoluta de utilizar solo glucosa y el ancPFK-MT presenta una probabilidad de 98,9% de utilizar glucosa y 99,9% de utilizar fructosa-6P, lo cual está en total acuerdo con los resultados experimentales del uso de sustratos. Por otra parte, el ancestro ancGK/PFK presentó una probabilidad de 99,9 % de utilizar glucosa y tan solo 10 % de utilizar fructosa-6P. Aunque el cálculo no concuerda totalmente con el resultado experimental para este ancestro, hay que considerar que el modelo predice que es en este ancestro donde es más probable que comience el uso por fructosa-6P. Estos resultados teóricos para rasgos ancestrales discretos dan una visión general de la evolución del uso de sustratos en esta familia, la que concuerda con los resultados experimentales obtenidos para las enzimas resucitadas y apoya la hipótesis de la bifuncionalidad de las enzimas de *Methanococcales* y *Methanosarcinales*.

#### 3.4.11 Robustez de la reconstrucción de estados ancestrales discretos

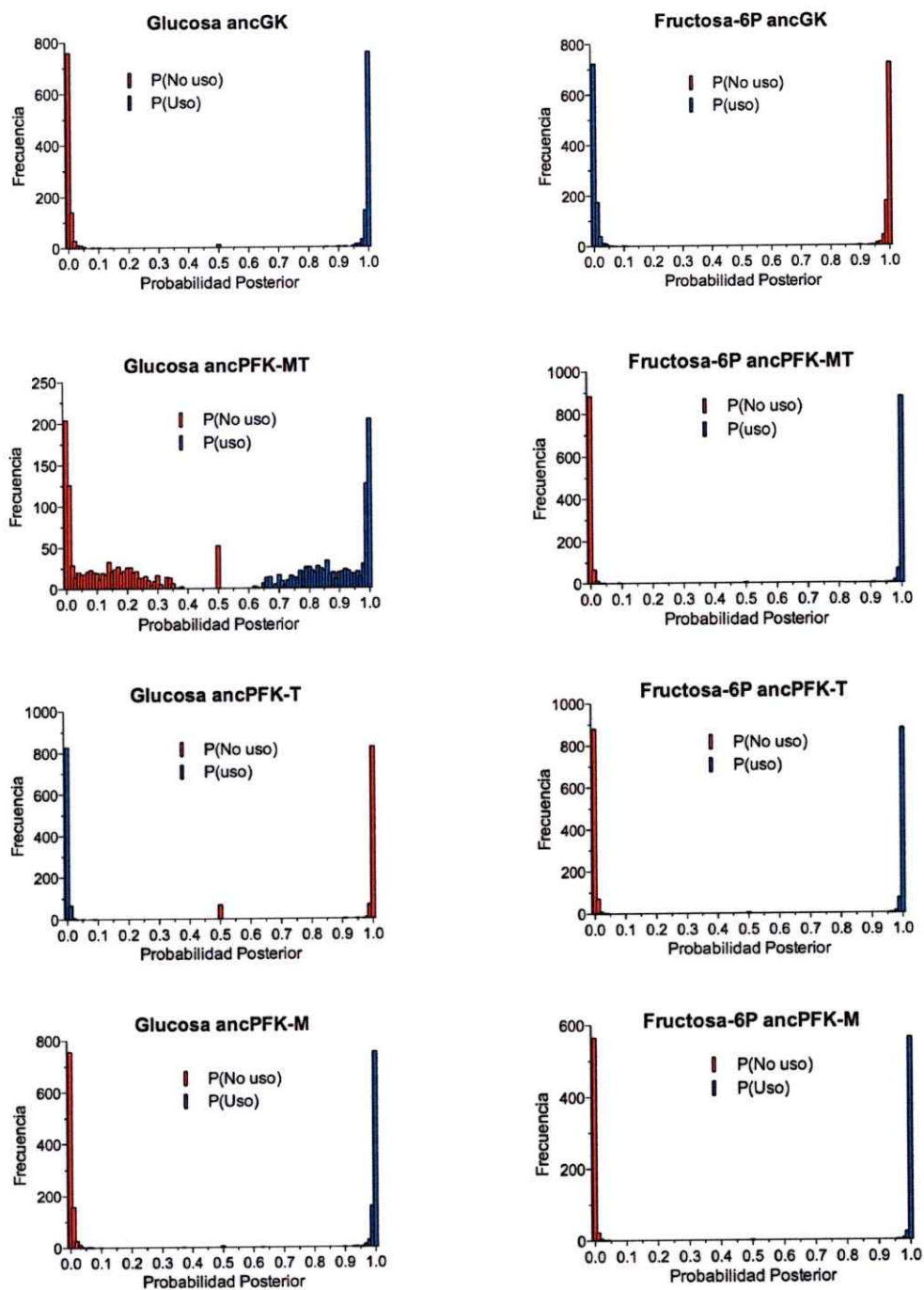
Se analizó la robustez de las reconstrucciones de estados discretos mediante máxima verosimilitud, dado que estos análisis se basaron en un solo árbol y con solo un set de parámetros. A continuación (**Figura 13**) se muestran las reconstrucciones realizadas mediante inferencia bayesiana con el programa BayesTrait V2.0 [87] sobre una población de 500 árboles obtenidos en Mrbayes 3.1.2 desde secuencias de aminoácidos, donde se utilizó solo una velocidad de cambio ( $q_{01}=q_{10}$ ) en contraste con el modelo con dos velocidades ( $q_{01}\neq q_{10}$ ) de acuerdo al factor de bayes entre ambos modelos. Se realizaron  $10^7$  iteraciones de las cuales se descartaron las primeras  $3 \times 10^5$  y las muestras analizadas fueron recogidas cada  $10^4$  iteraciones.





**Figura 13. Análisis de la reconstrucción de estados ancestrales discretos obtenidos por inferencia bayesiana para ancGK/PFK.**

A y C, probabilidad posterior obtenida para las muestras recogidas durante el análisis de MCMC. B y D, distribución de probabilidades posteriores para el uso y no uso de los azúcares como sustratos aceptores de fosforilo.



**Figura 14. Distribución de probabilidades posteriores para la reconstrucción de estados ancestrales discretos para el uso de glucosa y fructosa-6P mediante inferencia bayesiana.** Se muestran los histogramas para para el uso (azul) y no uso (Rojo) de los azúcares para los ancestros ancGK, ancMT, ancT y ancM.

Para el ancestro ancGK/PFK los análisis de MCMC y sus correspondientes histogramas se muestran en la **Figura 13**, en tanto que para los demás ancestros los histogramas se muestran en la **Figura 14**. De los histogramas se calculó el promedio y la mediana de las probabilidades posteriores para el uso de glucosa y fructosa-6P para cada ancestro (**Tabla 6**).

Los análisis realizados mediante inferencia bayesiana muestran esencialmente los mismos resultados que los obtenidos por máxima verosimilitud. La mayor diferencia en las probabilidades las mostró el ancestro ancGK/PFK para el uso de fructosa-6P, ya que el análisis bayesiano mostró en promedio una  $PP=0.44$  para el uso de F6P, mientras el análisis de máxima verosimilitud arrojó una probabilidad de 0.10. Esta incerteza muestra la incapacidad para inferir el uso de fructosa-6P en este ancestro, quizás dada la profundidad del nodo. Todos los demás ancestros concuerdan en el uso de azúcares inferido por Bayes o por ML, incluso el ancPFK-MT muestra un promedio de  $PP= 0.87$  para el uso de glucosa, lo que concuerda con el hecho de que las PFKs del orden *Methanococcales* son bifuncionales y que las únicas PFKs específicas son las del orden *Thermococcales*.



**Tabla 6. Probabilidades para el uso de glucosa y fructosa-6P obtenidas por inferencia bayesiana**

Nodo	PP uso de Glucosa		PP uso de Fructosa-6P	
	Promedio	Mediana	Promedio	Mediana
ancGK/PFK	0.938	0.972	0.441	0.444
ancGK	0.989	0.998	<b>0.004</b>	<b>0.001</b>
ancMT	0.873	0.910	0.996	0.999
ancT	<b>0.035</b>	<b>0.000</b>	0.996	0.999
ancM	0.992	0.998	0.996	0.999

PP: Probabilidad Posterior. En negrita se resaltan las probabilidades cercanas a cero, es decir mayor probabilidad de no uso.

#### 3.4.12 Recreando la evolución, mutante E72A de ancPFK-MT

Considerando los resultados de los modelos por homología y las energías de interacción de las dinámicas moleculares, el residuo E72 en ancPFK-MT sería el responsable de la unión de glucosa en esta enzima, en tanto que la presencia de este residuo en las enzimas PFKs sería el responsable de la bifuncionalidad. Esto sugiere que el cambio de este residuo por una alanina en el grupo de las PFKs de *Thermococcales* habría sido el cambio más relevante para llegar a la especialización por fructosa-6P en este grupo. Para probar esta hipótesis mutamos el gen de ancPFK-MT generando la mutante E72A; la caracterización cinética (**Figura 15**) de esta mutante mostró que la unión de fructosa-6P no se alteró con la mutación respecto a la enzima silvestre, pero el cambio del residuo E72 alteró dramáticamente la unión de glucosa, desde un valor de  $K_m$  de 19 mM en la enzima silvestre a uno por sobre 1M en la mutante (**Tabla 7**). Esta mutación recrearía el camino evolutivo que llevó a que el grupo de las PFKs de *Thermococcales* fuera específico por fructosa-6P. Además esta mutación explica la actividad bifuncional de la mutante A71E reportada en la PFK de *Pyrococcus horikoshii* [85].

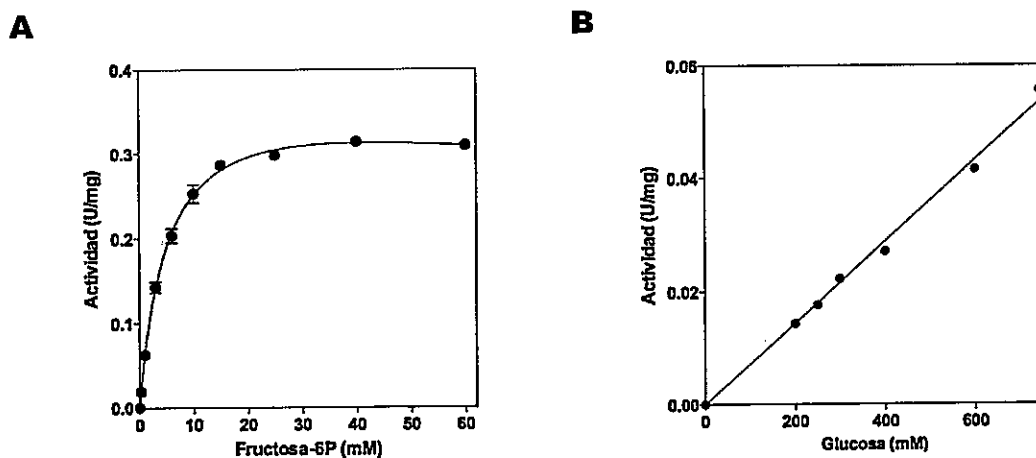


Figura 15. Determinación de los parámetros cinéticos de la mutante E72A de ancPFK-MT. A, curva de saturación para fructosa-6P ajustado al modelo de Michaelis-Menten. B, curva de saturación para glucosa.

Tabla 7. Constantes cinéticas para ancPFK-MT E72A a 30°C

Fructosa-6P		Glucosa		Fructosa-6P	Glucosa	Rasgo
$K_m(\mu\text{M})$	$k_{\text{cat}}(\text{s}^{-1})$	$K_m(\text{mM})$	$k_{\text{cat}}(\text{s}^{-1})$	$k_{\text{cat}} / K_m$ ( $\text{M}^{-1}\text{s}^{-1}$ )	$k_{\text{cat}} / K_m$ ( $\text{M}^{-1}\text{s}^{-1}$ )	
5180±450	0,336±0,013	~1x10 <sup>6</sup> *	~0,8*	6,5 x10 <sup>1</sup>	7,2 x10 <sup>-2</sup> #	Específica

\*valores estimados de gráficos de doble recíproco.

# valor determinado desde la pendiente inicial de la curva de saturación

### 3.5 *Discusión*

Las enzimas de la familia de quinasas dependientes de ADP son descritas en literatura como enzimas con una alta especificidad por glucosa o F6P, siendo la enzima de *M. jannaschii* la única enzima de la familia reportada como capaz de utilizar ambos sustratos [43,70]. Nuestros resultados de inhibición por glucosa en la *Ph*PFK y por F6P en la *TIGK* confirman esta gran especificidad por sustrato de las GKs y PFKs provenientes de organismos *Thermococcales*, las que son las enzimas dependientes de ADP mejor caracterizadas experimentalmente. Así también confirmamos la bifuncionalidad de la enzima de *M. jannaschii* respecto al uso de glucosa y fructos-6P como sustratos.

La caracterización de las enzimas ancestrales reconstruidas en este trabajo mostró que las enzimas específicas por glucosa o F6P provienen de enzimas bifuncionales. De acuerdo a la reconstrucción de rasgos discretos, la novedad evolutiva en esta familia sería la aparición de la especificidad por F6P en el grupo de los *Thermococcales*, en tanto que la mayoría de las enzimas anotadas como PFKs serían enzimas bifuncionales capaz de utilizar glucosa y F6P (PFKs de *Methanosarcinales* y *Methanococcales*). Así también el modelamiento molecular de los ancestros PFKs mostró que el cambio más relevante en ancPFK-MT para la especialización por F6P, es el cambio del residuo E72 por alanina, lo cual fue comprobado por mutagénesis sitio dirigida. Los resultados mostraron que el residuo E72 (ancPFK-MT) es fundamental para la unión de glucosa en las enzimas descritas como PFKs.

Un análisis de las especies que poseen en sus genomas estas enzimas ADP-dependientes muestra que los *Thermococcales* y *Methanosarcinales* tienen dos genes (PFK y GK). En el caso de los *Thermococcales* se ha demostrado que el producto de ambos genes es específico por sus sustratos; en cambio no hay evidencia experimental sobre las especificidades de estas enzimas en los *Methanosarcinales*. De acuerdo a la conservación del residuo glutámico en el sitio activo de las enzimas PFKs y a los análisis de rasgos discretos, las enzimas anotadas como PFKs de los *Methanosarcinales* serían bifuncionales y las GKs de este orden serían específicas por glucosa, lo que llevaría a una redundancia en la actividad glucoquinasa en estos organismos. Esta redundancia fue propuesta por Merino y Guixé [86] para explicar por qué la arquea *Methanococcoides burtonni* no presenta problemas en su metabolismo, aun cuando la enzima GK de este organismo no sería funcional ya que presenta una gran remoción en el extremo carboxilo terminal, donde se encuentra el motivo GXGD, cuyo aspártico es la base catalítica de estas enzimas.

Por otra parte, los organismos del orden *Methanococcales* solo presentan un gen que corresponde a una PFK. Los resultados de este trabajo indicarían que todos los organismos de este orden, y no solo la arquea *M. jannaschii* tienen PFKs bifuncionales. Esto implica que en estos organismos se utiliza una sola enzima para fosforilar glucosa y F6P dado la ausencia de genes homólogos a otras quinasas de azúcar que son dependientes de ATP o pirofosfato.

Para explicar la topología de la filogenia de esta familia Merino y Guixé proponen que hubo un evento de duplicación génica después de la divergencia entre



*Methanosarcinales* y *Thermococcales* (ancGK/PFK) (Figura 3). Además, proponen un evento de transferencia horizontal de genes (THG) para explicar la presencia de las PFKs en *Methanosarcinales*. Considerando las evidencias de este trabajo que muestran que las enzimas PFKs en su mayoría son bifuncionales y que las enzimas específicas por F6P de *Thermococcales* provienen de un ancestro bifuncional (ancPFK-MT), se puede reconsiderar la explicación de la topología del árbol de esta familia. Para explicar la presencia de GKs y PFKs específicas en los *Thermococcales*, se puede proponer la THG de una PFK bifuncional entre un organismo ancestro de los *Methanococcales* hacia un ancestro de los *Thermococcales*. La redundancia del gen transferido habría dado la libertad para la especialización del gen bifuncional por F6P con la mutación del residuo E72 por alanina, así como también la especialización por glucosa del gen original. La presencia de las glucoquinasas específicas en los *Methanosarcinales* se podrían deber a otro evento de THG desde los *Thermococcales*. No hay evidencia para la THG desde los *Methanococcales* hacia los *Thermococcales*, ya sea por organización de genes vecinos o por uso de codones, pero si hay evidencia proveniente del uso de codones para THG de glucoquinasas de esta familia desde *Thermococcales* hacia *Methanosarcinales*. Merino y Guixé describen que la GK de *Methanotherix thermophila* proviene de los *Thermococcales* dada su gran diferencia en el uso de codones con el genoma de *M. thermophila* y a la gran semejanza con el uso de codones de la glucoquinasas de *Thermococcales* [86]. En este mismo sentido, considerando la amplia presencia de estas enzimas solo en arqueas *Euriarqueotas* y la carencia de ortólogos en eucariontes inferiores como levaduras y *Entamoebas*, Ronimus y cols [36] propone un evento de HTG para la presencia de estas enzimas en eucariontes. Por otra parte, una búsqueda

actualizada (2do semestre 2014) de secuencias homólogas de esta familia en nuevos genomas secuenciados, arrojó la presencia de estas enzimas en algunas proteobacterias, las cuales parecen ser GKs específicas y una posibilidad es que habrían sido adquiridas por THG desde arqueas mesófilas del grupo de los *Methanosarcinales* o podrían ser un vestigio evolutivo de que estas enzimas se encontraban en el ancestro de los tres dominios de la vida. Todos estos eventos de THG muestran lo común que es este tipo de mecanismo en la evolución de procariotas y apoyan recientes investigaciones que muestran la importancia y la alta velocidad con que ocurren estos eventos durante la evolución de estos microorganismos [13].

Respecto a la transición desde enzimas bifuncionales a enzimas específicas por F6P, los resultados de este trabajo muestran que la sustitución del residuo E72 por alanina en el ancestro ancPFK-MT habría eliminado la actividad glucoquinasa de esta enzima a concentraciones fisiológicas de este sustrato. Sin embargo, la mutante (E72A), al igual que el ancestro ancPFK-T, todavía presentarían promiscuidad por glucosa ya que es posible detectar actividad a concentraciones sobre 100 mM, aunque en las enzimas actuales de *Thermococcales*, como es el caso de PhPFK, no es posible detectar esta actividad. Estos resultados concuerdan con el papel de la promiscuidad de sustrato como punto de partida para nuevas actividades enzimáticas o como vestigio evolutivo de actividades ancestrales [20]. También evidencian, que si bien hay mutaciones discretas que provocan grandes cambios en las actividades enzimáticas las que podrían estar involucradas en el *fitness* de un organismo y por lo tanto otorgar ciertas ventajas metabólicas en determinados ambientes ecológicos, hay otras mutaciones que se podrían

denominar neutrales o casi-neutrales que si bien no otorgan grandes cambios en una actividad enzimática pueden mejorar discretamente la actividad catalítica o afinidad por un sustrato en desmedro de actividades promiscuas que no tienen relevancia para el metabolismo [1]. Estas mutaciones casi-neutrales serían las responsables del aumento de la actividad catalítica y afinidad por F6P en las enzimas actuales de los *Thermococcales*.

Los parámetros cinéticos de las enzimas resucitadas y de las enzimas actuales muestran que la eficiencia catalítica ha ido en aumento a través de la evolución, lo que concuerda con las teorías de evolución de la catálisis enzimática [24]. La baja actividad catalítica de los ancestros y de algunas proteínas actuales hacen suponer que la vía glicolítica no es de principal importancia en las arqueas *Euriarqueotas*, en especial en los *Methanococcales* y *Methanosarcinales*, donde esta vía podría funcionar como un metabolismo secundario, y por ello presentan enzimas bifuncionales y con eficiencias catalíticas bajas. Esto estaría apoyado por ejemplo en que los *Methanococcales* obtienen su energía principalmente a través de la metanogénesis utilizando  $H_2$  y  $CO_2$ , por lo que sería esa vía la que estaría bajo una fuerte presión selectiva en estos organismos. Por otra parte, en los *Thermococcales* las quininas dependientes de ADP presentan eficiencias catalíticas mayores y con enzimas específicas, por lo que en estas *Euriarqueotas* la glicolisis sería de mayor relevancia metabólica. Estudios futuros, tanto en organismos *Thermococcales* los cuales son de interés para la industria en la obtención de enzimas termófilas y por otra parte, estudios en *Methanococcales* y *Methanosarcinales*, de gran interés para la industria biotecnológica y de energía, deberían establecer la relevancia de la glicolisis en el metabolismo general de las arqueas *Euriarqueotas*.

### 3.6 Conclusiones

- Las enzimas específicas actuales de la familia de quinasas ADP-dependientes surgieron de proteínas con actividad bifuncional, lo que concuerda con la hipótesis de estados “generalistas” durante la evolución de enzimas.
- La eficiencia catalítica de las enzimas de esta familia ha ido en aumento a través de la evolución.
- La actividad bifuncional en las enzimas PFKs está dada por la presencia de un ácido glutámico en el sitio de unión del azúcar (E72 en ancPFK-MT).
- La mutación del residuo E72 por alanina en ancPFK-MT fue el cambio más relevante hacia la especialización por fructosa-6P en las enzimas PFKs de *Thermococcales*.

## **4 CAPÍTULO DOS**

**Caracterización experimental de la actividad  
bifuncional en fosfofructoquinasas  
dependientes de ADP provenientes de  
*Methanococcales* y *Methanosarcinales***

#### 4.1 Introducción

Las quinasas de azúcar dependientes de ADP son descritas en literatura como enzimas específicas por glucosa (GKs) o fructosa-6P (PFKs), pero esta visión está sesgada taxonómicamente ya que la gran mayoría de las enzimas caracterizadas pertenecen al reino de las arqueas *Euriarqueotas*, específicamente al orden *Thermococcales* [70]. Como se ha dicho anteriormente, la única enzima del orden *Methanococcal* caracterizada es la enzima de *Methanocaldococcus jannaschii*, la cual es reportada como enzima bifuncional [43] y es vista como la excepción en esta familia de enzimas.

De acuerdo a los resultados mostrados en el capítulo I, en donde encontramos que los ancestros ancPFK-MT y ancPFK-M presentan esta actividad dual y en base a los análisis de rasgos discretos y a los análisis de interacción proteína ligando, las enzimas de los organismos *Methanococcales* y *Methanosarcinales* serían bifuncionales. Con esta idea se propuso caracterizar la enzima de *Methanococcus maripaludis*, un *Methanococcal* lejano filogenéticamente de *M. jannaschii* aunque del mismo orden. Este organismo es mesófilo y es un fijador de nitrógeno además de metanógeno, por lo que el conocimiento del funcionamiento del metabolismo de este organismo es de gran interés biotecnológico. En cuanto a los *Methanosarcinales*, no hay ninguna enzima de este orden descrita en literatura y por lo tanto nos propusimos caracterizar la enzima de *Methanococcoides burtonii*, organismo que es metanógeno y psicrófilo aislado de la Antártida. Además de estudiar la preferencia por sustratos de su PFK, esta proteína es un buen candidato a futuro para estudiar las características fisicoquímicas que permiten la catálisis en este tipo de enzimas a bajas temperaturas.

## **4.2 Hipótesis**

Las quinasas de azúcar dependientes de ADP de organismos *Methanococcales* y *Methanosarcinales* son enzimas bifuncionales capaz de fosforilar glucosa y fructosa-6P.

## **4.3 Objetivo**

Caracterizar la enzima dependiente de ADP del *Methanococcal Methanococcus maripaludis* y la enzima del *Methanosarcinal Methanococcoides burtonni*.

#### ***4.4 Actividad bifuncional fosfofructoquinasa/glucoquinasa en el orden Methanococcales: Caracterización bioquímica de la enzima mesófila de Methanococcus maripaludis***

Hemos expresado, purificado y caracterizado la única enzima homóloga de la familia dependiente de ADP en el organismo mesófilo *Methanococcus maripaludis*. Los resultados muestran que la enzima es bifuncional siendo capaz de fosforilar glucosa y fructosa-6P y que las actividades GK y PFK son altamente reguladas por  $Mg^{2+}$  libre y por el producto AMP. Estudios de modelamiento molecular evidenciaron residuos claves en la unión de los sustratos en tanto que análisis de la reversibilidad de las actividades GK y PFK hacia la formación de glucosa y fructosa-6P, mostraron que sólo la actividad GK es reversible y que podría estar involucrada en la formación de glucosa. Por otra parte, el modelamiento por homología de las enzimas del orden *Methanococcales* indicó que todas las enzimas del orden serían bifuncionales capaces de fosforilar glucosa y fructosa-6P.

A continuación se presentan los resultados de este trabajo publicado en la revista *FEBS Journal*, (2014) 281, 2017-2029. El estudio fue ideado por Victor Castro-Fernandez y Victoria Guixé y el trabajo experimental fue encabezado por Victor Castro-Fernandez en colaboración con Felipe Bravo-Moraga y Alejandra Herrera-Morande.



## Bifunctional ADP-dependent phosphofructokinase/ glucokinase activity in the order *Methanococcales* – biochemical characterization of the mesophilic enzyme from *Methanococcus maripaludis*

Victor Castro-Fernandez, Felipe Bravo-Moraga, Alejandra Herrera-Morande and Victoria Guixé

Departamento de Biología, Facultad de Ciencias, Universidad de Chile, Santiago, Chile

### Keywords

ADP-dependent sugar kinases; bifunctional enzyme; glucokinase; gluconeogenesis; phosphofructokinase

### Correspondence

V. Guixé, Departamento de Biología,  
Facultad de Ciencias, Universidad de Chile,  
Las Palmeras 3425, Ñuñoa, Santiago, Chile,  
7800003.  
Fax: +56 2 227 12983  
Tel: +56 2 297 87335  
E-mail: vguixé@uchile.cl

(Received 26 September 2013, revised 23  
January 2014, accepted 17 February 2014)

doi:10.1111/febs.12757

In some archaea, the phosphorylation of glucose and fructose 6-phosphate (fructose 6P) is carried out by enzymes that are specific for either substrate and that use ADP as phosphoryl donor. In the hyperthermophilic archaeon *Methanocaldococcus jannaschii*, a bifunctional enzyme able to phosphorylate glucose and fructose 6P has been described. To determine whether the ability to phosphorylate both glucose and fructose 6P is a common feature for all enzymes of the order *Methanococcales*, we expressed, purified and characterized the unique homologous protein of the mesophilic archaea *Methanococcus maripaludis*. Assay of the enzyme activity with different sugars, metals and nucleotides allows us to conclude that the enzyme is able to phosphorylate both fructose 6P and glucose in the presence of ADP and a divalent metal cation. Kinetic characterization of the enzyme revealed complex regulation by the free  $Mg^{2+}$  concentration and AMP, with the latter appearing to be a key metabolite. To determine whether this enzyme could have a role in gluconeogenesis, we evaluated the reversibility of both reactions and found that glucokinase activity is reversible, whereas phosphofructokinase activity is not. To determine the important residues for glucose and fructose 6P binding, we modeled the bifunctional phosphofructokinase/glucokinase enzyme from *M. maripaludis* and its interactions with both sugar substrates using protein–ligand docking. Comparison of the active site of the phosphofructokinase/glucokinase enzyme from *M. maripaludis* with the structural models constructed for all the homology sequences present in the order *Methanococcales* shows that all of the ADP-dependent kinases from this order would be able to phosphorylate glucose and fructose 6P, which rules out the current annotation of these enzymes as specific phosphofructokinases.

### Database

Model data are available in the Protein Model Data Base under accession numbers [PM0079106](#), [PM0079107](#), [PM0079108](#), [PM0079109](#), [PM0079110](#), [PM0079111](#), [PM0079112](#), [PM0079113](#), [PM0079114](#), [PM0079115](#) and [PM0079116](#)

### Abbreviations

EM, Embden–Meyerhof; fructose 1,6-bisP, fructose 1,6-bisphosphate; fructose 6P, fructose 6-phosphate; GK, glucokinase; glucose 6P, glucose 6-phosphate; MmPFK/GK, bifunctional phosphofructokinase/glucokinase enzyme from *Methanococcus maripaludis*; PFK, phosphofructokinase.

## Introduction

The archaea domain presents modifications of the classical Embden–Meyerhof (EM) pathway. One of the most striking modifications present in some genera of Euryarchaeota is the phosphorylation of glucose and fructose 6-phosphate (fructose 6P) by ADP as a phosphoryl donor, instead of ATP, as in the classical pathway. These ADP-dependent sugar kinases are able to phosphorylate glucose (ADP-GKS; EC 2.7.1.147) or fructose 6P (ADP-PFK; EC 2.7.1.146) and are homologous between them, although they show no sequence similarity to any ATP-dependent kinase known to date. For this reason, these proteins were proposed as a new family of enzymes [1], and subsequently were classified as belonging to the ribokinase superfamily, given the structural similarity with its members [2]. In the Euryarchaeota kingdom, ADP-dependent sugar kinases have been annotated in the orders: *Pyrococcus*, *Thermococcus*, *Methanosarcina*, *Methanosaeta*, *Methanococcoides*, *Methanococcus*, *Methanocaldococcus* and *Archaeoglobus* [3]. To date, the best characterized enzymes are from the *Thermococcales* group. Glucokinases (GK) from *Thermococcus litoralis* [4] and *Pyrococcus furiosus* [5] shown to be specific for glucose, as well as phosphofructokinases (PFK) from *Pyrococcus horikoshii* [6], *Thermococcus zilligii* [7], *Pyrococcus furiosus* [1] and *Archaeoglobus fulgidus* [8], are specific for fructose 6P. In the order *Methanococcales*, ADP-PFK activity has been detected in cellular extract from *Methanococcus jannaschii*, *Methanococcus (Methanoterris) igneus* and *Methanococcus maripaludis* [9]. Subsequently, the enzyme from the hyperthermophilic methanogenic archaeon *Methanocaldococcus jannaschii* was described as a bifunctional enzyme able to phosphorylate glucose and fructose 6P [10]. Given its dual activity, this enzyme was initially proposed as an ancestral state for all the family, although a subsequent study showed that this enzyme is not an ancestral form [11]. However, the same study did not address whether this dual character is a particular feature of the *M. jannaschii* enzyme or whether it is a common feature for all the enzymes of the order *Methanococcales*. Moreover, in the genome of the mesophilic archaea *M. maripaludis*, there is a unique ADP-dependent homologous protein, which raises the question of whether this homologous has both GK and phosphofructokinase activities.

Recently, a bifunctional fructose 1,6-bisphosphate-aldolase/phosphatase virtually present in all archaeal groups, as well as the deeply branching bacterial lineages, has been reported [12]. It was proposed that this enzyme ensures unidirectional gluconeogenesis under

conditions where the carbon flux does not need to be turned to sugar degradation. Based on its phylogeny and distribution pattern, Say and Fuchs [13] proposed that the bifunctional fructose 1,6-bisphosphatealdolase/phosphatase represents the ancestral gluconeogenic enzyme, supporting an evolutionary scenario where gluconeogenesis preceded glycolysis. On the other hand, the autotrophic methanogens such as *M. maripaludis* grow on a limited range of substrates, including acetate, molecular hydrogen, compounds of 1C such as methylamines and formate, and few alcohols [14]. They are able to store glycogen, but not assimilated sugars, indicating that, besides participating in glycolysis, ADP-dependent kinases could also have a role in glucose synthesis, although this issue has not been evaluated. Anabolism and catabolism of glycogen is of evolutionary interest for comparison with heterotrophic archaea, as well as with respect to the potential applications of these methanogenic organisms in biotechnology.

In the present study, we addressed the expression, purification and biochemical characterization of the unique ADP-dependent homologous protein present in the genome of the mesophilic archaea *M. maripaludis*. Our results demonstrate that this enzyme is bifunctional capable of phosphorylating both glucose and fructose 6P (MmPFK/GK). Considering that less developed organisms, such as those from the order *Methanococcales*, lack the ability to assimilate glucose but are able to store glycogen, the reversibility of the MmPFK/GK glucokinase (GK) activity supports the fact that this enzyme could have a role in glucose synthesis. Characterization of the enzyme revealed a complex regulation by the free  $Mg^{2+}$  concentration and AMP, with the latter appearing to be a key metabolite. On the other hand, molecular modeling of the MmPFK/GK enzyme along with protein–ligand docking with both substrates allows us to determine residues important for substrate binding. Finally, molecular models constructed for all ADP-dependent kinases from the order *Methanococcales* indicate that all of these enzymes would be bifunctional, and thus able to phosphorylate glucose and fructose 6P.

## Results and Discussion

### Phylogenetic tree

To analyze the evolutionary history of the ADP-dependent sugar kinases from the order *Methanococcales*, we inferred a phylogenetic tree with the available sequences from the nonredundant protein database. Characterized homologous PFK enzymes sequences from *Thermococcales* were used as an outgroup to root



the tree (Fig. 1). The tree obtained for the *Methanococcales* group was similar to the one obtained for the complete family [15]. Our inferred tree contains 11 sequences from the order *Methanococcales*; among them are sequences from thermophilic organisms of the *Methanocaldococcaceae* family (genus *Methanocaldococcus* and *Methanotorris*) and the *Methanococcaeae* family (genus *Methanothermococcus*), as well as mesophilic organisms from the genus *Methanococcus* of this latter family.

In the *Methanococcales* group, the enzyme from *M. jannaschii* is the only ADP-dependent sugar kinase characterized that is capable of using glucose and fructose 6P as phosphoryl acceptor substrate. According to the information available in the *Kyoto Encyclopedia of Genes and Genomes* (<http://www.genome.ad.jp/kegg/>), the genome of this organism has only one homologous gene; accordingly, it has been proposed that this enzyme would act as a bifunctional phosphofructokinase/GK enzyme *in vivo* [10]. By contrast, the genome of the *Thermococcales* group has two copies of this gene and presents separate phosphofructokinase and GK activities. To determine whether the phosphorylation of glucose and fructose 6P is carried out exclusively by enzymes of the ADP-dependent sugar kinases family in the modified EM pathway of the orders *Thermococcales* and *Methanococcales*, we searched for ATP dependent hexokinases and phosphofructokinases homologous genes in the *Kyoto Encyclopedia of Genes and Genomes* database. We found no counterpart of these genes; thus, in the *Thermococcales* group, glucose

and fructose 6P phosphorylation is carried out only by specific enzymes of the ADP-dependent sugar kinases family, whereas, in the *Methanococcales* group, both reactions would be carried out by the same enzyme encoded by a unique gene.

#### Biochemical characterization of the enzyme from *M. maripaludis*

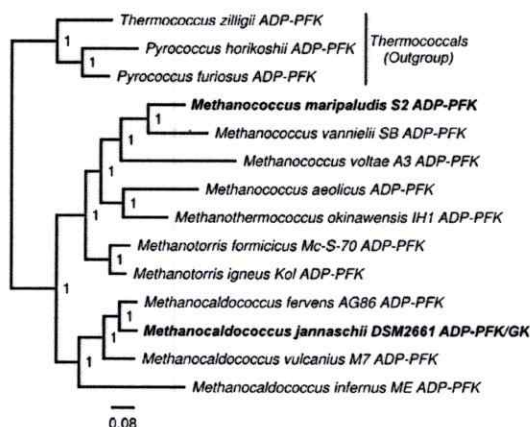
The optimized synthetic gene of the ADP-dependent enzyme from *M. maripaludis* strain S2 (UniProt Protein Database Code: Q6LXQ3) was expressed in *Escherichia coli*. The protein was purified and analyzed by SDS/PAGE (Fig. S1). Analysis of its oligomerization state by gel exclusion chromatography revealed that the enzyme is a monomer, which is usually the oligomeric state found for other proteins of the family, including the *M. jannaschii* enzyme [10].

To analyze the sugar substrate ambiguity of this enzyme, we evaluated 16 different sugars as a phosphoryl acceptor (Fig. 2A). Among these, only glucose and fructose 6P showed ADP consumption. None of the other sugars tested demonstrated a significant consumption of the nucleotide relative to the control without sugar. These results indicate that the mesophilic *M. maripaludis* enzyme is not promiscuous in terms of sugar substrate specificity but, instead, is specific for glucose and fructose 6P. At a sugar concentration of 10 mM, the MmPFK/GK enzyme showed higher activity with glucose compared to that obtained with fructose 6P; with this latter substrate having 75% of the activity measured with glucose.

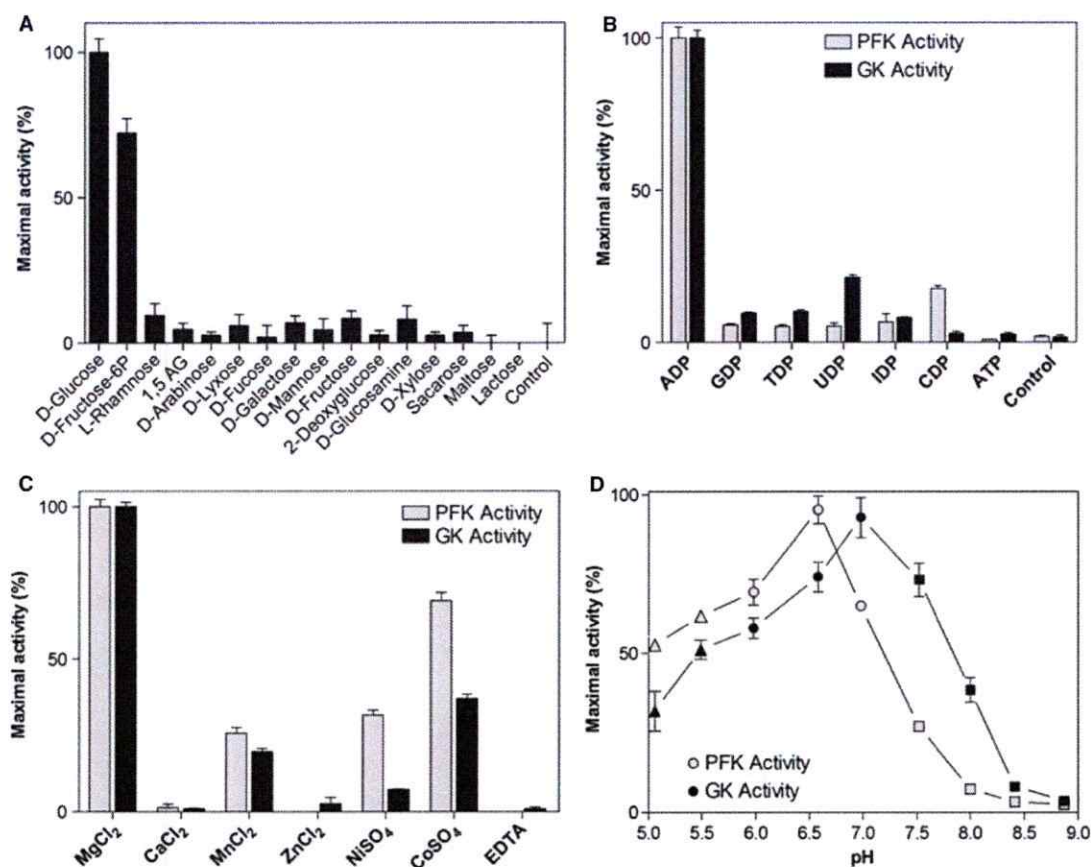
To confirm that this enzyme is a true ADP-dependent kinase, we assayed the enzyme activity in the presence of several nucleotides, using glucose or fructose 6P as a sugar substrate. The highest activity was obtained in the presence of ADP, although phosphoryl group transfer from other diphosphorylated nucleotides was observed, albeit at a significantly slower rate (Fig. 2B). In the presence of ATP, no activity was detected.

Divalent cations were required for MmPFK/GK activity. Among the divalent metal cations tested, the highest activity was observed in the presence of  $Mg^{2+}$ , although, in the presence of  $Co^{2+}$ ,  $Ni^{2+}$  and  $Mn^{2+}$ , significant activity was also measured (Fig. 2C). The presence of EDTA in the reaction mixture completely abolished enzyme activity, indicating that the metal-nucleotide complex is the true substrate for the enzyme reaction.

When investigating the pH at which maximum activity is obtained, we found maximum PFK activity at pH 6.5, whereas GK maximal activity was found at



**Fig. 1.** Phylogenetic tree of ADP-dependent sugar kinases from the order *Methanococcales*. PFK sequences from the order *Thermococcales* were used as an outgroup. The posterior probability of each node is shown in each split. The position of the *M. jannaschii* and *M. maripaludis* sequences is highlighted in bold.



**Fig. 2.** Biochemical characterization of MmPFK/GK. (A) ADP-dependent kinase activity measured as ADP consumption with different sugar substrates (10 mM). Activity is shown as the percentage of activity measured in the presence of glucose. 1,5 AG, 1,5-anhydroglucitol. Control corresponds to activity in the absence of sugar. (B) GK and PFK activity with different nucleotides as a phosphoryl donor. (C) GK and PFK activity in the presence of 1 mM ADP, 2 mM of a different divalent cation and 1 mM glucose or fructose 6P. EDTA 50 mM was used as a control without metal. (D) GK and PFK activity at different pH values; pH 5.0 and 5.5 for acetate/acetic acid; pH 6.0, 6.6 and 7.0 for Pipes/NaOH; pH 7.5, 8.0, 8.5 and 8.9 for Tris-HCl.

pH 7.0 (Fig. 2D). The lower optimum pH obtained in the case of the ADP-PFK activity has been associated with the protonation of a lysine group involved in stabilization of the phosphate group of fructose 6P [11].

#### Kinetic characterization of the MmPFK/GK enzyme

To evaluate the specificity of the MmPFK/GK enzyme with respect to glucose or fructose 6P, we determined the kinetic parameters performed for both sugars in the presence of Mg-ADP (Fig. S2). The activity using fructose 6P as substrate shows hyperbolic behavior, whereas the saturation curve for glucose exhibits

substrate inhibition at high glucose concentrations (> 200 mM). On the other hand, the saturation curves for both PFK and GK activity with Mg-ADP as substrate showed substrate inhibition (Fig. S2). Table 1 shows the kinetic parameters obtained from nonlinear adjustments of the saturation curves. The  $k_{cat}/K_m$  values are  $2.53 \times 10^5$  and  $5.7 \times 10^2 \text{ M}^{-1}\cdot\text{s}^{-1}$  for fructose 6P and glucose, respectively. These results show that, when two competing substrates such as these were present in an equimolar mixture, the rate at which fructose 6P is phosphorylated is 440-fold higher than the glucose phosphorylation rate. This difference in the value of the  $k_{cat}/K_m$  parameter is mainly a result of the dissimilar values of  $K_m$  for both substrates. The  $K_m$  for

**Table 1.** Kinetic parameters of MmPFK/GK.

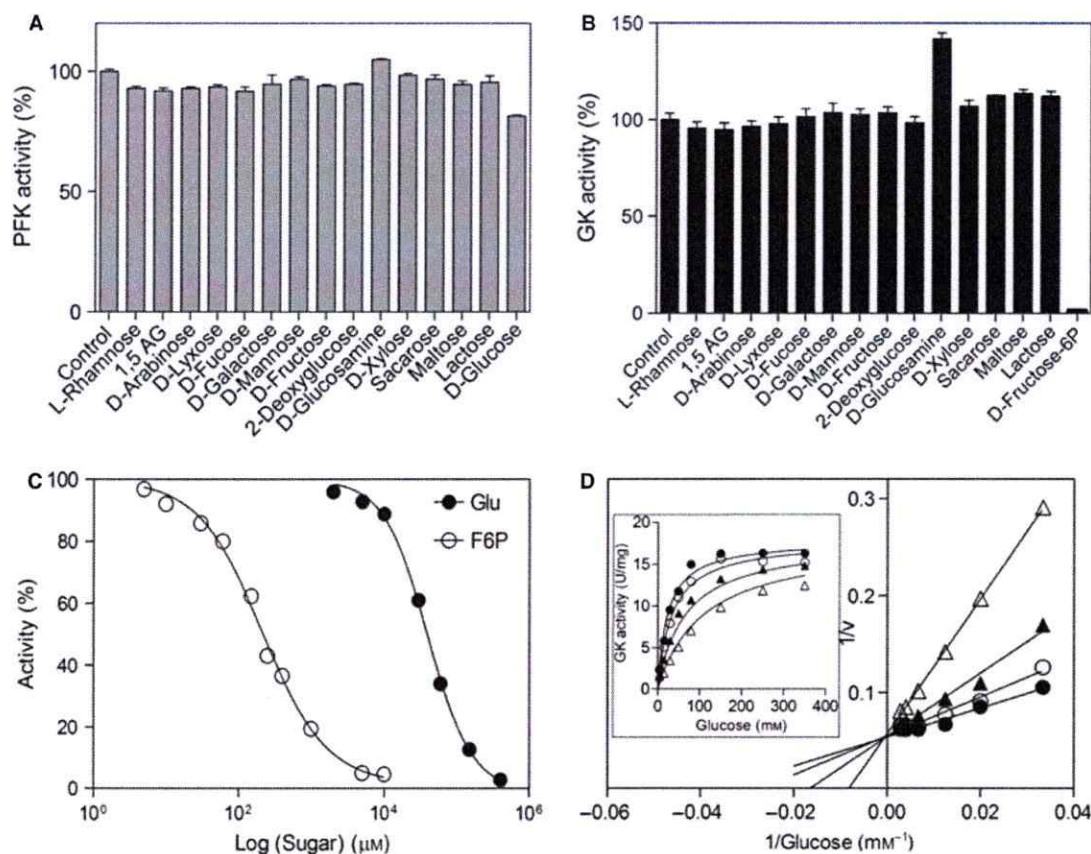
Sugar	$K_m$ (mM)	$k_{cat}$ (s <sup>-1</sup> )	$k_{cat}/K_m$ (M <sup>-1</sup> ·s <sup>-1</sup> )	$K_i$ (mM)
Direct reaction				
Glucose	40 ± 5	23 ± 1	5.75 × 10 <sup>2</sup>	895 ± 205
Fructose 6P	0.065 ± 0.014	16.5 ± 0.6	2.53 × 10 <sup>5</sup>	—
MgADP (GK)	16 ± 6	48 ± 12	3.00 × 10 <sup>3</sup>	25 ± 10
MgADP (PFK)	8 ± 2	30 ± 4	3.75 × 10 <sup>3</sup>	54 ± 21
Reverse reaction				
Glucose 6P	0.623 ± 0.076	0.678 ± 0.015	1.09 × 10 <sup>3</sup>	—
Fructose 1,6-BisP	0.120 ± 0.016	0.45 ± 0.02	3.75 × 10 <sup>3</sup>	—

fructose 6P is in the micromolar range, whereas, for glucose, a  $K_m$  of 40 mM was obtained. Although the enzyme from *M. jannaschii* is also more specific for fructose 6P, its  $k_{cat}/K_m$  ratio is only 60-fold greater for this substrate compared to the value obtained for glucose [10]. Interestingly, despite its high  $K_m$  for glucose, the activity of the MmPFK/GK enzyme with this substrate is higher compared to other sugars, which is mainly a result of its high  $k_{cat}$  value. Noticeably, the  $K_m$  values for ADP in the GK and PFK reactions (16 and 8 mM, respectively) are much higher than those reported for the *M. jannaschii* enzyme (0.032 and 0.63 mM, respectively). Because regulation of the enzyme activity by divalent metal cation has been reported previously [16], we determined the effect of the free Mg<sup>2+</sup> concentration on the kinetic parameters. Indeed, in the presence of the 1 mM free Mg<sup>2+</sup> concentration (instead of the 5 mM Mg<sup>2+</sup> in excess over the nucleotide concentration used in the experiments shown in Table 1), the  $K_m$  values for Mg-ADP in the GK and PFK reaction diminished dramatically, reaching values of 0.3 and 1.2 mM, respectively. A similar scenario is also observed for several enzymes of the ribokinase superfamily whose activity is strongly dependent on the amount of free divalent metal cation present. In some members, this second metal, in addition to the one present in the metal nucleotide complex, acts by inhibiting enzyme activity, whereas, in others, it acts as an activator. In a previous study, we investigated the regulation of specific ADP-dependent GK and PFK enzymes from *Thermococcales* by divalent metal cations and found that only GKs are inhibited in a significant degree by the free metal. For these enzymes, the  $K_m$  values for the metal-nucleotide complex are not significantly affected by the concentration of free metal and so a mechanism was proposed where the binding of a second metal produces a complex with a reduced catalytic constant [16]. However, in the MmPFK/GK enzyme, a different situation is observed; the free Mg<sup>2+</sup> concentration has a profound effect on the  $K_m$  values for MgADP in the GK and PFK reactions, whereas it has no significant effect on the  $V_{max}$

of both reactions. Although the data suggest a competitive inhibition of free Mg<sup>2+</sup> (binding of free metal and MgADP substrate to the same enzyme form), a detailed inhibition mechanism requires additional knowledge of the general kinetic mechanism of the GK and PFK reactions and the dissociation constants for Mg<sup>2+</sup> binding, amongst other data. On the other hand, although the intracellular Mg<sup>2+</sup> concentration in these organisms is not known, even at a concentration of approximately 5 mM of the free cation, the enzyme exhibits significant activity at glucose concentrations between 5 and 10 mM, mainly as a result of its high  $k_{cat}$ . This implies that this activity would be relevant for the *in vivo* function and that the enzyme could be performing both GK and PFK functions *in vivo*.

Although the MmPFK/GK enzyme is unable to use other substrate besides glucose and fructose 6P as substrates, we aimed to further assess the binding specificity by evaluating the ability of the enzyme to accommodate other sugars at the active site. Accordingly, we assayed the ability of different sugar to inhibit the GK and PFK reactions. We employed a set of 14 sugars, besides glucose and fructose 6P, at a fixed concentration of 10 mM for each. For the PFK reaction, none of the sugars tested were able to inhibit the activity significantly (Fig. 3A) with the only exception of glucose, which inhibits enzyme activity by approximately 20%, as would be expected if both sugar substrates bind to the same active site. For the GK reaction, the results are almost the same, except for glucosamine, which results in approximately 40% activation (Fig. 3B). The activation effect is concentration dependent and is not an artefact of the assay (data not shown). In this case, strong inhibition of the GK reaction was observed in the presence of fructose 6P, which can be explained by the low  $K_m$  of the enzyme for the phosphorylated sugar, which is present at a concentration of 10 mM, as well as by the high  $K_m$  for glucose measured under this experimental condition (approximately 5 mM free Mg<sup>2+</sup>). A more detailed inspection of the inhibitory range of both sugars allowed us to determine an IC<sub>50</sub> value of 220 μM for





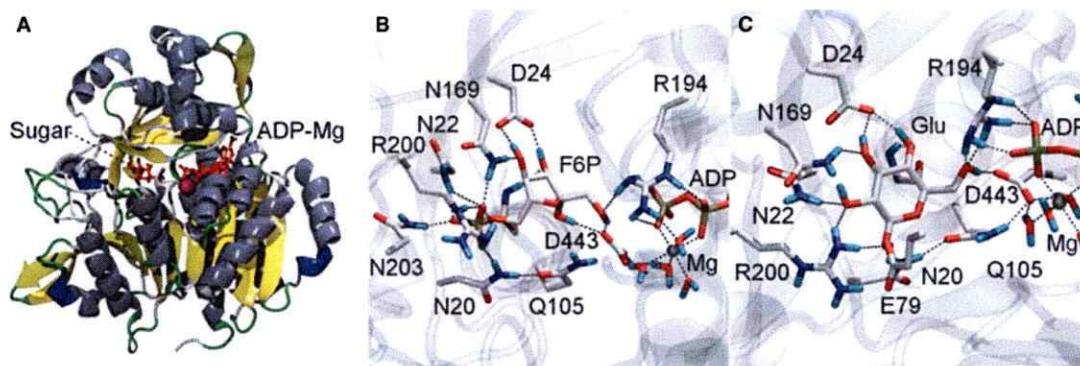
**Fig. 3.** Inhibition of MmPFK/GK by sugars. (A) Inhibition of PFK activity by different sugars at 10 mM, activity without inhibitor was used as a control, 1,5 AG, 1,5-anhydroglucitol. Only glucose presented significant inhibition (20% of inhibition). (B) Inhibition of GK activity by different sugars at 10 mM; in the presence of fructose 6P, only 2% of residual activity was detected. Interestingly, in the presence of glucosamine, 40% activation was observed. (C) Dose-response inhibition of GK activity (white circles) by fructose 6P (F6P) and inhibition of PFK activity (black circles) by glucose (Glu). The IC<sub>50</sub> for fructose 6P and glucose was 220 μM and 40 mM, respectively. (D) Double-reciprocal plot for competitive inhibition of GK activity by fructose 6P. Fructose 6P concentrations were: 0 mM (black circles), 0.1 mM (white circles), 0.4 mM (black triangles) and 0.8 mM (white triangles). Inset: global fit to a competitive inhibition mechanism, a K<sub>i</sub> of 226 μM for fructose 6P was determined. Error bars in (A) and (B) indicate the SE of three independent measurements.

fructose 6P in the GK reaction and an IC<sub>50</sub> of 40 mM for glucose in the PFK reaction (Fig. 3C). Binding of both substrates to the same active site was demonstrated by the competitive inhibition of fructose 6P when glucose was used as the variable substrate (Fig. 3D), where a K<sub>i</sub> value of 226 μM was obtained.

#### Molecular modeling of MmPFK/GK

To study the structural determinants of the sugar specificity, the structure of the ADP-dependent PFK/GK from *M. maripaludis* was modeled using homology modeling (Table S1). With this model, we performed docking assays with fructose 6P and glucose, and then

the Mg-ADP complex was mounted onto the resulting docking structures. The whole structures were subjected to an energy minimization process (Fig. 4A). For ADP, inspection of residues present within 5 Å away from the nucleotide identified: K76, P77, Q105, K175, E176, Q191, S192, N193, R194, I196, E270, D297, E298, E300, H345, T346, M347, Y348, I350, A376, S377, A380, S429, R430, I431, V432, E433, N434, P435, T438, G440, L441, G442, D443, T444 and I445. Among these M347, I431 and L441 create a hydrophobic pocket around the adenine group; R194 makes a hydrogen bond with α and β phosphates; carbonyl and NH groups from V432 peptide bond make a hydrogen bond with the NH<sub>2</sub> group of C6 and the



**Fig. 4.** Protein-ligand interaction modeling results. (A) Homology model of MmPFK/GK; sugar and nucleotide binding sites are indicated. Substrates are shown as red sticks and  $Mg^{2+}$  is shown as a magenta sphere. (B)  $\beta$ -D-fructose 6P docked in MmPFK/GK enzyme after the energy minimization procedure. (C)  $\beta$ -D-glucose docked in MmPFK/GK enzyme after the energy minimization procedure. Hydrogen bonds are shown as a dotted line in (B) and (C).

N1 atom of adenine; the carbonyl group from peptide bond of H345 makes hydrogen bonds with the C3 hydroxyl of the ribose ring and the NH groups from peptide bonds of G440 and L441; and G442 and D443 make hydrogen bonds with the  $\beta$  phosphate of ADP, whereas both phosphates ( $\alpha$  and  $\beta$ ) are coordinated to a  $Mg^{2+}$  ion. On the other hand, the waters that coordinate the  $Mg^{2+}$  ion interact with the carbonyl group of the peptide bond F271, as well as the side chains of E270, E300, H345 and D443 (Fig. S3). E270 and E300 corresponds to residues present in the HXE and NXXE motifs, respectively, which have been described as divalent cation binding sites in the ADP-dependent sugar kinases family [16].

For glucose, analysis of residues that have at least one atom within an area of 5 Å away from the sugar identified the residues; N20, N22, D24, E79, G103, G104, Q105, I108, N169, R170, I171, R194, I196, A198, R200, G230, Q232, A272, V439, G440 and D443. For fructose 6P, the residues were essentially the same, excluding E79, R170, I196, A198, Q232 and A272, and adding V21, K167 and N203.

The active site of MmPFK/GK modeled with fructose 6P (Fig. 4B) shows that the phosphate group of fructose 6P is stabilized by hydrogen bonds established through the side chains of residues N20, N22, N169, R200 and N203. On the other hand, D24 interacts with the hydroxyl groups of C3 and C4 of the furanose ring, whereas the NH group of the peptide bond of G104 interacts with the C3 hydroxyl, the N169 interacts with the C4 hydroxyl, the R194 interacts with the C1 hydroxyl and D443 interacts with the C1 and C2 hydroxyls (Fig. 4B). Residues R194 and D443 also interact with the C1 hydroxyl phosphoryl acceptor

group. The catalytic mechanism of the ribokinase superfamily members is highly conserved, where an aspartic acid (or cysteine) acts as the catalytic base that subtracts the hydroxyl hydrogen, creating an alkoxide ion that is stabilized by an arginine (or lysine) residue, which in turns allows the nucleophilic attack to  $\beta$  (or  $\gamma$ ) phosphate of the Mg-ADP (or Mg-ATP) substrate [3]. Site-directed mutagenesis of the aspartic residue in several members of ribokinase superfamily members provokes a dramatic loss of activity [3]. Also, site-directed mutagenesis of the positive residue noted above for the ADP-dependent PFK from *P. horikoshii* (R185) causes a dramatic decrease of activity [6]. In MmPFK/GK, these residues correspond to R194 and D443, which make appropriate interactions with the substrates with respect to their roles in the catalytic mechanism proposed for the whole superfamily.

Residue side chains involved in the stabilization of glucose through hydrogen bonds are N20 and E79 with the C1 hydroxyl group, as well as N22 with the C2 hydroxyl group, whereas R200 forms a hydrogen bond with the C1 and C2 hydroxyls groups; N169 forms a bond with the C2 and C3 hydroxyls groups; D24 forms a bond with the C3 and C4 hydroxyls groups; and the NH group from peptide bond of G104 forms a bond with the C4 hydroxyl. R194 and D443 established an interaction with the C6 hydroxyl group, which is the phosphate group acceptor (Fig. 4C). As in fructose 6P, R194 and D443 enable glucose to accommodate an appropriate orientation for catalysis. On the other hand, the results of molecular modeling show that both sugars are bound to the enzyme by essentially the same residues, except for N203, which established an interaction only when the



substrate was fructose 6P, and E79, which interacts only with glucose. However, analysis of the whole family sequences shows that E79 is present in specific GKs and also that N203 is conserved in all specific GKs from *Methanosarcinales* and *Thermococcales* but is absent in PFKs from *Methanosarcinales* and *Thermococcales*. This suggests that the role of these residues in substrate specificity is unclear and that more thorough evolutionary studies are needed to demonstrate the contribution of these or other residues to substrate specificity in the whole family.

To determine whether active site glucose interactions are preserved throughout the order *Methanococcales*, we constructed homology models of all of the ADP-dependent sugar kinases available for this order (Table S2), except for the enzyme from *M. jannaschii*, which has already been studied *in vitro* [10] and *in silico* [11]. We performed a structural superposition of each of the nine models constructed in the presence of glucose with the one built for the *M. maripaludis* enzyme (Fig. S4). We found that residues present in a 5-Å cut-off area from glucose are completely conserved in the order *Methanococcales*, which supports the idea that all of the proteins of this order would be able to bind glucose besides fructose 6P, as was noted up to now.

#### Role of MmPFK/GK in glucose synthesis

Previous studies have proposed that the EM pathway has an gluconeogenic origin [7,13], which is consistent with the theory of an chemoautrophic origin of metabolism [17]. For example, less developed organisms, such as those from the order *Methanococcales*, lack the ability to assimilate glucose but are able to store glycogen [18]. Metabolic pathways in higher organisms operate through the degradation of organic compounds and CO<sub>2</sub> release, whereas most primitive organisms work as CO<sub>2</sub> fixers for compound formation [19]. Therefore, we investigated the possible role of the MmPFK/GK enzyme in glucose formation by analyzing its reverse catalysis for the PFK and GK activity. Unexpectedly, in contrast to that found for the direct reaction, the  $K_m$  for glucose 6-phosphate (glucose 6P) is 623 μM (Table 1) instead of the value of 40 mM determined for glucose. However, for the reverse reaction, the  $k_{cat}$  value is 34-fold lower than that obtained for the direct reaction. Analysis of the  $k_{cat}/K_m$  ratios shows that the glucose dephosphorylation reaction is two-fold more efficient than its phosphorylation. Even though this difference is marginal, it indicates that the MmPFK/GK enzyme can catalyze glucose formation. Because *M. maripaludis*, similar to other methanogens, is able to synthesize glycogen intracellularly and degrade it under

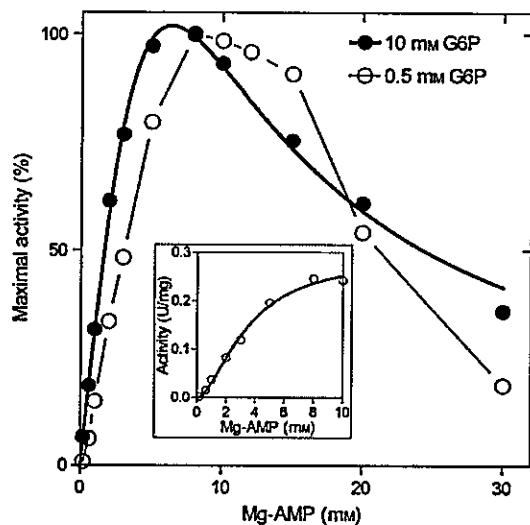
starvation, the possibility has been raised regarding the production of methane from glycogen or D-glucose by methanogen itself [10]. In the absence of other annotated genes, the reversible GK activity of this bifunctional enzyme could be responsible for glucose formation in *M. maripaludis*.

To gain insights into the operation of both the direct and reverse GK reactions, we evaluated the effect of the nucleotide product. For the reverse reaction, ADP presents a strong inhibition of enzyme activity with an IC<sub>50</sub> value of 80 μM when the reaction was assayed in the presence of 1 mM AMP, 6 mM MgCl<sub>2</sub> and 10 mM glucose 6P. On the other hand, in the direct GK reaction, AMP shows an interesting behavior; at concentrations < 5 mM, it activates glucose phosphorylating activity, whereas, at higher concentrations, it presents the expected product inhibition (not shown). Also, the complex effect of AMP is observed when acting as substrate in the reverse GK reaction, where a strong substrate inhibition effect is observed ( $K_i$  of 0.4 mM at 10 mM of glucose 6P). Unexpectedly, when the same experiment was performed at a glucose 6P concentration of 0.5 mM, a sigmoidal behavior was observed, along with an even more pronounced inhibition (Fig. 5). These results highlight the role of AMP as the key metabolite responsible for the operation of this enzyme either in the glycolytic or gluconeogenic flux direction.

The reverse catalysis for PFK activity (Table 1) showed a  $K_m$  for fructose 1,6-bisphosphate (fructose 1,6-bisP) similar to that obtained for fructose 6P, although with a  $k_{cat}$  value 36-fold lower than that determined for the direct reaction, where the efficiency of phosphatase activity is 67-fold lower than the kinase activity. The irreversibility of the kinase activity is in agreement with the presence of a gene coding a fructose 1,6-bisP/aldolase bifunctional enzyme in the *M. maripaludis* genome, which has been described as an enzyme that catalyzes the dephosphorylation of fructose 1,6-bisP, and operates only toward gluconeogenesis [17]. Although the calculated  $\Delta G^\circ$  for each of the ADP-PFK and ADP-GK reactions is a rather large negative value with a large equilibrium constant, similar to those determined for ATP [20] (Table 2), the values were obtained under standard conditions. In the cell, the mass action ratio can be modified substantially the flux direction and, consequently, in the absence of information about the intracellular concentration of the substrates and products of the reactions involved, the predictive value of these constants in relation to flux direction is limited.

These results show that, besides the ADP-dependent sugar kinase from *M. jannaschii*, the enzyme from





**Fig. 5.** Mg-AMP saturation curves for GK reverse activity. GK reverse activity in the presence of approximately 5 mM of free  $Mg^{2+}$  and different glucose 6P concentrations. At 10 mM of glucose 6P (G6P), a strong substrate inhibition was observed; fitting the data to a substrate inhibition equation gives a  $K_i$  of 0.4 mM. At 0.5 mM of glucose 6P, a more complex behavior was observed; below 10 mM of MgAMP, the saturation curve presents a sigmoidal behavior with a Hill coefficient of 1.7 (inset); over this concentration, a strong inhibition effect is observed.

**Table 2.** Standard free energy change of GK and PFK reactions.

Reaction	$\Delta G^\circ$ (kJ·mol <sup>-1</sup> )	$K_{eq}$ at 25 °C
ADP		
ADP-GK <sup>a</sup>	-19.0	$2.14 \times 10^3$
ADP-PFK <sup>a</sup>	-16.5	$7.80 \times 10^2$
ATP		
ATP-GK <sup>b</sup>	-16.7	$8.45 \times 10^2$
ATP-PFK <sup>b</sup>	-14.2	$3.08 \times 10^2$

<sup>a</sup>See supporting information (Fig. S5). <sup>b</sup>Jencks [20].

*M. maripaludis* is also able to phosphorylate glucose and fructose 6P. Moreover, by molecular modeling of all of the ADP-dependent family enzymes from the order *Methanococcales*, we showed that they are all bifunctional and able to phosphorylate both substrates. These results agree with the presence of only one gene of the ADP-dependent family in the genome of these organisms. By contrast, the genome of *Thermococcales* has two copies of the gene, each coding for specific GK or PFK enzymes, which are also present in the order *Methanosarcinal* [11,15]. Future studies should investigate the relationship between the presence of only one copy of this gene and the evolutionary history of the

family, as well as its role in gluconeogenesis, because, according to the high conservation of the active site residues in enzymes of the order *Methanococcales*, bifunctionality is a feature shared by all its members. Moreover, the effect of the free  $Mg^{2+}$  concentration, as well as the effect of glucosamine and AMP, suggests a complex regulation of the activity of this archaeal kinase, which has not been recognized previously and could be a common feature for all the enzymes of the order *Methanococcales*.

## Experimental procedures

### Alignment and Bayesian inference of phylogeny

The protein BLAST web server was used to extract all of the sequences of the order *Methanococcales* present in the non-redundant protein database sequence using a PSI-BLAST algorithm with three iterations and ADP-PFK/GK from *M. jannaschii* as a template (data available up to December 2012). Multiple sequence alignment was constructed based on three-dimensional and secondary structure constraints using PROMALS3D [21] and then misaligned positions were corrected by visual inspection in MULTISEQ from VMD [22]. The Bayesian inference of phylogeny was performed using MRBAYES, version 3.1.2 [23]. For the analysis, we used Cprev as the fixed rate model, which showed a posterior probability of 1.0 in the mixed model and  $\gamma$ -shaped rate variation across the site with a proportion of the invariable sites. The number of generations was set to  $1 \times 10^6$ , and the sample frequency to 100, with two runs with four chains for each run and a temperature parameter of 0.2. The mean SD of the split frequencies was  $< 1 \times 10^{-4}$ . The consensus tree was calculated from 150 002 sample trees.

### Expression and purification of recombinant ADP-PFK/GK homolog from *M. maripaludis*

Mm-ADP-PFK/GK cDNA (Genscript, Piscataway, NJ, USA), optimized for expression in *E. coli*, was directly cloned into the modified pET-28b vector in the restriction sites *NdeI* and *BamHI* and its sequence was verified by DNA sequencing. This vector inserted six histidines and a cleavage site for the TEV protease into the C-terminal of the protein. *E. coli* BL21(DE3) pLysS was transformed with pET-Mmpfk/gk and grown in LB medium containing  $35 \mu\text{g}\cdot\text{mL}^{-1}$  kanamycin and  $35 \mu\text{g}\cdot\text{mL}^{-1}$  chloramphenicol at 37 °C until  $D_{600}$  of approximately 0.4 was reached. Expression of the recombinant protein was induced with 1 mM of isopropyl thio- $\beta$ -D-galactoside overnight. Cells were harvested by centrifugation, suspended in binding buffer (50 mM Tris-HCl, pH 7.6, 500 mM NaCl, 20 mM imidazole and 5 mM  $MgCl_2$ ) and disrupted by sonication. After centrifugation (18 514 g for 30 min), the soluble fraction was

loaded onto a Ni<sup>2+</sup>-NTA affinity column (HisTrap HP; GE Healthcare, Little Chalfont, UK). Protein was eluted with a linear gradient between 20 and 500 mM imidazole and fractions with enzyme activity were pooled. The protein was dialyzed against 25 mM Pipes (pH 6.5) and 5 mM MgCl<sub>2</sub>, and then glycerol was added to 50% v/v followed by storage at -20 °C. Enzyme purity was analyzed by SDS/PAGE stained with coomassie blue.

#### Determination of enzyme activity

PFK activity was assayed spectrophotometrically as described previously [24]. Briefly, enzyme preparation was mixed with reaction buffer containing 25 mM Pipes buffer (pH 6.5), 25 mM MgCl<sub>2</sub>, 2 mM fructose 6P, 0.2 mM NADH, 20 mM ADP, 1.96 U of  $\alpha$ -glycerophosphate dehydrogenase, 19.6 U of triosephosphate isomerase and 0.52 U of aldolase in a final volume of 0.25 mL (standard assay).

GK activity was measured spectrophotometrically by monitoring NAD<sup>+</sup> reduction at 340 nm, coupled with glucose 6P oxidation at 30 °C. A standard assay was carried out in a final volume of 0.25 mL in the presence of 25 mM Hepes (pH 7.8), 0.5 mM NAD<sup>+</sup>, 2–5 U of glucose 6P dehydrogenase from *Leuconostoc mesenteroides*, 200 mM glucose, 25 mM MgCl<sub>2</sub> and 20 mM ADP (standard assay).

Reverse activities (GK and PFK) were measured by monitoring the appearance of ADP with a coupling assay containing pyruvate kinase/lactate dehydrogenase in a final volume of 0.25 mL. Briefly, enzyme preparation was mixed with reaction buffer containing 25 mM Hepes (pH 7.0), 0.4 U of pyruvate kinase, 1.2 U of lactate dehydrogenase, 0.3 mM phosphoenolpyruvate, 125 mM KCl, 0.2 mM NADH, 15 mM MgCl<sub>2</sub>, 10 mM AMP and 10 mM glucose 6P or fructose 1,6-bisP.

Inhibition of the reverse GK activity by ADP was measured by monitoring glucose appearance using the glucose oxidase/peroxidase system from BioSystem (Barcelona, Spain). The assay mixture contained: 20 mM phosphate (pH 7.5), 1 mM phenol, > 2 U·mL<sup>-1</sup> glucose oxidase, > 0.2 U·mL<sup>-1</sup> peroxidase and 0.1 mM 4-aminopyrine. Reduction of 4-aminopyrine to quinoneimine was monitored at 500 nm.

All enzyme activities were determined at 30 °C using a UV/visible Synergy 2 spectrophotometer (BioTek, Winooski, VT, USA) and 96-well plates (model 269620; Nunc, Rochester, NY, USA). An extinction coefficient of 6.22 mm<sup>-1</sup>·cm<sup>-1</sup> for NADH and a path length correction at 1 cm was used for specific activity (U·mg<sup>-1</sup>) determination.

#### Reagent concentrations

Glucose and fructose 6P were titrated with the respective assays using ADP-dependent glucokinase from *T. litoralis* and ADP-dependent PFK from *P. horikoshii*, respectively; ADP was titrated with a GK assay using ADP-GK from

*T. litoralis*, glucose 6P was titrated with glucose 6P dehydrogenase from *L. mesenteroides* and NAD<sup>+</sup>, the titration was performed with a co-substrate concentration at least 20-fold over the compound titrated. Protein concentration was determined using the Bio-Rad protein assay (Bio-Rad, Hercules, CA, USA) and BSA Standard (Thermo Fisher Scientific Inc., Waltham, MA, USA) as a reference. Substrate titration and protein concentrations were measured in a photodiode array spectrophotometer (8453; Hewlett Packard/Agilent, Santa Clara, CA, USA).

#### Substrate specificity

Phosphoryl acceptor specificity of MmPFK/GK was determined by measuring the consumption of ADP in a discontinuous assay. The reaction mixture contained 25 mM Tris-HCl (pH 7.0), 20 mM MgCl<sub>2</sub>, 25 mM ADP, 10  $\mu$ g of MmPFK/GK and 10 mM of each sugar (17 sugars tested) in a final volume of 0.2 mL. The mixture was incubated for 15 min at 30 °C, the reaction was stopped with perchloric acid at a final concentration of 3% v/v, placed on ice for 5 min, and then the mixture was neutralized with a final concentration of NaHCO<sub>3</sub> of 0.5 M. Next, 5  $\mu$ L of reaction mixture was titrated with 495  $\mu$ L of ADP titration mix containing 0.8 U of pyruvate kinase, 2.4 U of lactate dehydrogenase, 0.3 mM phosphoenolpyruvate, 125 mM KCl, 25 mM Tris-HCl (pH 7.0), 5 mM MgCl<sub>2</sub> and 0.2 mM NADH at 37 °C. ADP consumption with different sugars was calculated using the reaction mixture without sugars as a reference. Activity was reported as a percentage of ADP consumption obtained with glucose in 15 min.

Activity with different nucleotides was determined by measuring PFK and GK activity as described above, substituting ADP for GDP, IDP, CDP, TDP, UDP and ATP at 10 mM for each nucleotide in the presence of 15 mM MgCl<sub>2</sub>. The activity is reported as percentage of that obtained with ADP. The divalent cation preference was tested similarly; MgCl<sub>2</sub> was replaced by CaCl<sub>2</sub>, MnCl<sub>2</sub>, ZnCl<sub>2</sub>, NiSO<sub>4</sub> and CoSO<sub>4</sub> at a concentration of 2 mM in the presence of 1 mM of ADP. Metal ion dependent catalysis was evaluated by adding 50 mM EDTA to the assay in the presence of MgCl<sub>2</sub>. To discard a counterion effect, the assay was performed with MgSO<sub>4</sub>, instead of MgCl<sub>2</sub>, obtaining the same GK and PFK activity under both conditions (data not shown).

#### pH dependence

pH dependence activity was determined by measuring the PFK and GK activity spectrophotometrically as described above, substituting Pipes and Hepes buffer: pH 5.0 and 5.5 were attained with acetate/acetic acid buffer; pH 6.0, 6.6 and 7.0 were attained with Pipes buffer; pH 7.5, 8.0, 8.5 and 8.9 were attained with Tris-HCl buffer. The buffer concentration was 50 mM.

### Kinetic parameters

For GK and PFK assays, kinetic parameters were determined at 30 °C, varying the concentration of substrates at the saturating co-substrate concentration. In each case, 5 mM MgCl<sub>2</sub> in excess over the nucleotide concentration (ADP or AMP) was used (approximately 5 mM free Mg<sup>2+</sup> concentration). For kinetic parameters determinations at 1 mM free Mg<sup>2+</sup>, a *K<sub>d</sub>* of 676 μM for the Mg-ADP complex was considered; ADP protonation was neglected considering the pH of the assay. Initial velocity studies were performed twice in triplicate. Data were analyzed using nonlinear regression with GRAPHPAD PRISM, version 5.0 (GraphPad Software Inc., San Diego, CA, USA) and fitted to the Michaelis-Menten, substrate inhibition, dose-response inhibition or competitive inhibition equations.

### Molecular mass determination

The molecular mass of the enzyme was determined by gel filtration on a TSK gel column G3000SWXL (7.8 mm × 30 cm) (Tosoh, Tokyo, Japan), equilibrated with column buffer (100 mM Tris-HCl, pH 7.0, 100 mM NaCl and 2 mM MgCl<sub>2</sub>) and calibrated with Thyroglobulin (669 kDa), immunoglobulin G (150 kDa), myoglobin (17.6 kDa) and vitamin B<sub>12</sub> (1.3 kDa).

### Molecular modeling and docking

Recently, we have shown that ligands modified the major and minor domain angle in a ADP-dependent sugar kinase family member and that these conformational changes are kinetically relevant [15]. For this reason, we modeled the MmPFK/GK enzyme in the closed conformation using a modified protocol as described previously [11].

Briefly, the *P. horikoshii* PFK structure (Protein Data Bank Code: [3DRW](#)) has a 43% identity with MmPFK/GK, although this structure is in an open conformation. The only structure in the closed conformation is GK from *Pyrococcus furiosus* (Protein Data Bank Code: [1UA4](#)), which shares only 27% sequence identity with MmPFK. Then, the 3DRW structure was split off into its large and small domains and both of them were structurally aligned with the [1UA4](#) structure. Next, the closed structure and the two fragments were used as templates for modeling the enzyme in the closed conformation. Fifty models were constructed with MODELLER, version 8 [25]. From these, the best 10 potential DOPE models were chosen, and quality was evaluated with PROSA2003 [26], PROCHECK [27] and VERIFY3D [28] (Table S1). The same procedure was performed for each of the remaining sequences of the *Methanococcales* group, except that only 10 models were evaluated (Table S2).

Docking assays with glucose and fructose 6P were performed with AUTODOCKVINA, version 1.0 [29], with protonation states according to the pH values used in the kinetic

experiments. The protonation state of the ionizable residues was calculated using the web server H++ [30]. After this procedure, partial charges were derived with the Gasteiger method using AUTODOCKTOOL [31]. Residues N169 and R200 were free to move when the ligand was glucose, whereas, for fructose 6P, E79 and R200 were set as flexible side chains. Docking results with a lower interaction energy and the phosphoryl acceptor hydroxyl oriented towards the GXGD motif were selected because aspartic acid is considered as the catalytic base for phosphate group transfer in all ribokinase superfamily members [3]. Next, the Mg-ADP complex was mounted onto the resulting docking structures according to the procedure described by Merino *et al.* [16] and then both structures were subjected to energy minimization with NAMD, version 2.7 [32] using the CHARMM27 [33] force field. Proteins were placed in a water box, the system was neutralized with NaCl to a final concentration of 0.1 M, and 10 000 iterations were performed without restrictions.

### Acknowledgements

This work was supported by Fondo Nacional de Desarrollo Científico y Tecnológico (Fondecyt, Chile) Grant 1110137 and CONICYT scholarship 24121448.

### References

- 1 Tuininga JE, Verhees CH, van der Oost J, Kengen SW, Stams AJ & de Vos WM (1999) Molecular and biochemical characterization of the ADP-dependent phosphofructokinase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J Biol Chem* **274**, 21023–21028.
- 2 Ito S, Fushinobu S, Yoshioka I, Koga S, Matsuzawa H & Wakagi T (2001) Structural basis for the ADP-specificity of a novel glucokinase from a hyperthermophilic archaeon. *Structure* **9**, 205–214.
- 3 Guixé V & Merino F (2009) The ADP-dependent sugar kinase family: kinetic and evolutionary aspects. *IUBMB Life* **61**, 753–761.
- 4 Koga S, Yoshioka I, Sakuraba H, Takahashi M, Sakasegawa S, Shimizu S & 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**, 1079–1085.
- 5 Kengen SW, Tuininga JE, de Bok FA, Stams AJ & de Vos WM (1995) Purification and characterization of a novel ADP-dependent glucokinase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J Biol Chem* **270**, 30453–30457.
- 6 Currie MA, Merino F, Skarina T, Wong AHY, Singer A, Brown G, Savchenko A, Caniguier A, Guixé V, Yakunin AF *et al.* (2009) ADP-dependent

- 6-phosphofructokinase from *Pyrococcus horikoshii* OT3: structure determination and biochemical characterization of PH1645. *J Biol Chem* **284**, 22664–22671.
- 7 Ronimus RS & Morgan HW (2003) Distribution and phylogenies of enzymes of the Embden–Meyerhof–Parnas pathway from archaea and hyperthermophilic bacteria support a gluconeogenic origin of metabolism. *Archaea* **1**, 199–221.
  - 8 Hansen T & 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**, 29–35.
  - 9 Verhees CH, Tuininga JE, Kengen SWM, Stams AJM, Van Der Oost J & De Vos WM (2001) ADP-dependent phosphofructokinases in mesophilic and thermophilic methanogenic archaea. *J Bacteriol* **183**, 7145–7153.
  - 10 Sakuraba H, Yoshioka I, Koga S, Takahashi M, Kitahama Y, Satomura T, Kawakami R & Ohshima T (2002) ADP-dependent glucokinase/phosphofructokinase, a novel bifunctional enzyme from the hyperthermophilic archaeon *Methanococcus jannaschii*. *J Biol Chem* **277**, 12495–12498.
  - 11 Merino F & 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.
  - 12 Sato T & Atomi H (2011) Novel metabolic pathways in Archaea. *Curr Opin Microbiol* **14**, 307–314.
  - 13 Say RF & Fuchs G (2010) Fructose 1,6-bisphosphate aldolase/phosphatase may be an ancestral gluconeogenic enzyme. *Nature* **464**, 1077–1081.
  - 14 Jones WJ, Nagle DP Jr & Whitman WB (1987) Methanogens and the diversity of archaeobacteria. *Microbiol Rev* **51**, 135–177.
  - 15 Rivas-Pardo JA, Herrera-Morande A, Castro-Fernandez V, Fernandez FJ, Vega MC & Guixé V (2013) Crystal structure, SAXS and kinetic mechanism of hyperthermophilic ADP-dependent glucokinase from *Thermococcus litoralis* reveal a conserved mechanism for catalysis. *PLoS One* **8**, e66687.
  - 16 Merino F, Rivas-Pardo JA, Caniuguir A, García I & Guixé V (2012) Catalytic and regulatory roles of divalent metal cations on the phosphoryl-transfer mechanism of ADP-dependent sugar kinases from hyperthermophilic archaea. *Biochimie* **94**, 516–524.
  - 17 Wächtershäuser G (1988) Before enzymes and templates: theory of surface metabolism. *Microbiol Rev* **52**, 452–484.
  - 18 König H, Nusser E & Stetter KO (1985) Glycogen in *Methanobolus* and *Methanococcus*. *FEMS Microbiol Lett* **28**, 265–269.
  - 19 Berg IA, Kockelkorn D, Ramos-Vera WH, Say RF, Zarzycki J, Hügler M, Alber BE & Fuchs G (2010) Autotrophic carbon fixation in archaea. *Nat Rev Microbiol* **8**, 447–460.
  - 20 Jencks WP (1976) Physical and Chemical Data. In *Handbook of Biochemistry and Molecular Biology*, 3rd edn (Fasman G, ed.), pp. 296–304. CRC Press, Cleveland, OH.
  - 21 Pei J, Kim B-H & Grishin NV (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res* **36**, 2295–2300.
  - 22 Roberts E, Eargle J, Wright D & Luthey-Schulten Z (2006) MultiSeq: unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics* **7**, 382.
  - 23 Ronquist F & Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.
  - 24 Kengen SW, De Bok FA, Van Loo ND, Dijkema C, Stams AJ & De Vos WM (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**, 17537–17541.
  - 25 Sali A & Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**, 779–815.
  - 26 Sippl MJ (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins Struct Funct Genet* **17**, 355–362.
  - 27 Laskowski RA, MacArthur MW, Moss DS & Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **26**, 283–291.
  - 28 Bowie JU, Luthy R & Eisenberg D (1990) Fold into a known three-dimensional structure. *Science*, **253**, 164–170.
  - 29 Trott O & Olson AJ (2009) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **31**, 455–461.
  - 30 Gordon JC, Myers JB, Folta T, Shoja V, Heath LS & Onufriev A (2005) H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res* **33**, W368–W371.
  - 31 Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS & Olson AJ (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* **30**, 2785–2791.
  - 32 Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L & Schulten K (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781–1802.
  - 33 MacKerell AD, Bashford D, Bellott Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha

S, Joseph-McCarthy D *et al.* (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102, 3586–3616.

### Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site:

**Fig. S1.** SDS/PAGE of the MmPFK/GK purification.

**Fig. S2.** Kinetic characterization of MmPFK/GK.

**Fig. S3.** ADP interactions at the active site of the MmPFK/GK homology model.

**Fig. S4.** Comparison of the ADP-dependent sugar kinases homology models for sequences of the order *Methanococcales* with the *M. maripaludis* homology model.

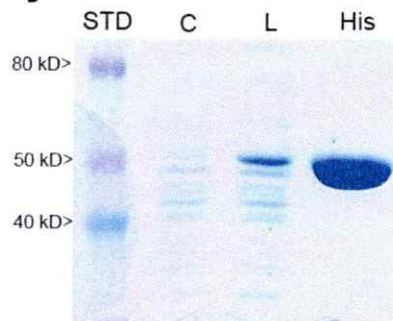
**Fig. S5.** Estimation of standard free energy change and equilibrium constants of the ADP-GK and ADP-PFK reactions.

**Table S1.** Evaluation of the MmPFK/GK homology model.

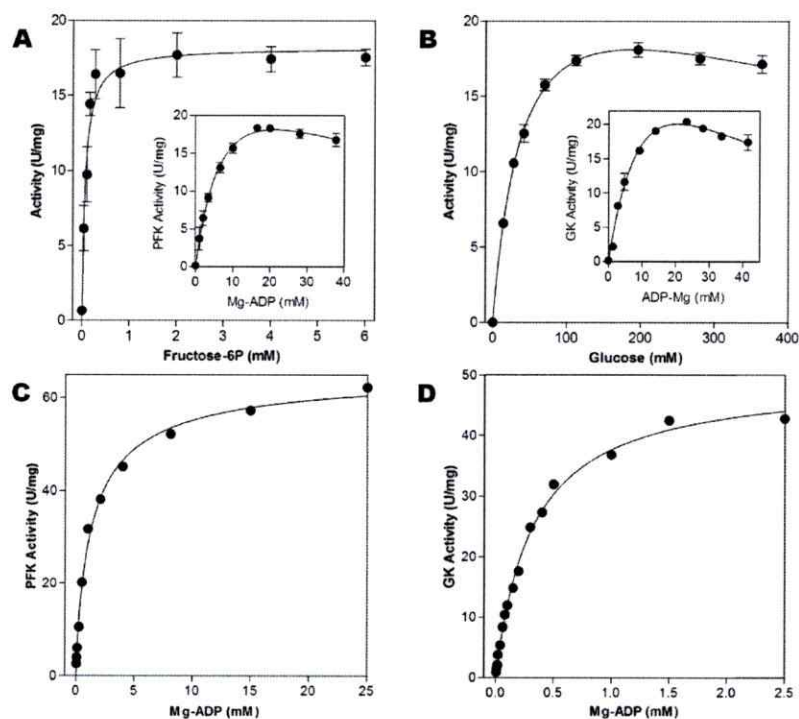
**Table S2.** Evaluation of the ADP-PFK/GK homology models for sequences of the order *Methanococcales*.



**Supplementary material: DOI: 10.1111/febs.12757**



**Figure S1. SDS-PAGE of the MmPFK/GK purification**, STD: molecular weight standards, C: fraction not retained in the HisTrap column, L: soluble fraction of total lysate, His: purified protein after HisTrap column. 20  $\mu$ g of protein were loaded in each lane.



**Figure S2. Kinetic characterization of MmPFK/GK.** **A.** PFK activity at different fructose-6P concentrations, inset; saturation curve for MgADP in the presence of 2 mM fructose-6P and 5 mM excess of  $MgCl_2$  over the ADP concentration. **B.** GK activity at different glucose concentration, inset; saturation curve for MgADP in presence of 200 mM glucose and 5 mM excess of  $MgCl_2$  over the ADP concentration. Curves in A and B were fit to substrate inhibition model. **C** and **D.** MgADP saturation curve for PFK and GK activity respectively at 1 mM free  $Mg^{2+}$ . Both curves were fit to the Michaelis-Menten model. Differences in apparent maximal activities between different free  $Mg^{2+}$  concentrations are due to the effect of substrate inhibition observed in A and B.

Table S1. Evaluation of the MmPFK/GK homology model

Model	DOPE Potencial	PROCHECK (Región of Ramachandran Plot)				VERIFY- 3D Score>0.2	PROSA Z-Score
		Most favoured	Additional allowed	Generously allowed	Disallo wed		
Mm37	-54582	391	24	4	5	82,54%	-9,34
Mm24	-54575	393	26	3	2	82,11%	-9,29
Mm00	-54559	389	32	2	1	84,91%	-9,17
Mm45	-54443	389	30	1	4	84,27%	-9,17
Mm04	<b>-54442</b>	<b>396</b>	<b>24</b>	<b>3</b>	<b>1</b>	<b>84,48%</b>	<b>-9,43</b>
Mm38	-54432	384	34	5	1	80,60%	-9,36
Mm30	-54407	388	31	4	1	77,59%	-9,06
Mm23	-54318	392	27	3	2	86,42%	-9,06
Mm46	-54308	390	29	3	2	83,19%	-9,49
Mm05	-54265	385	31	5	3	80,39%	-9,04

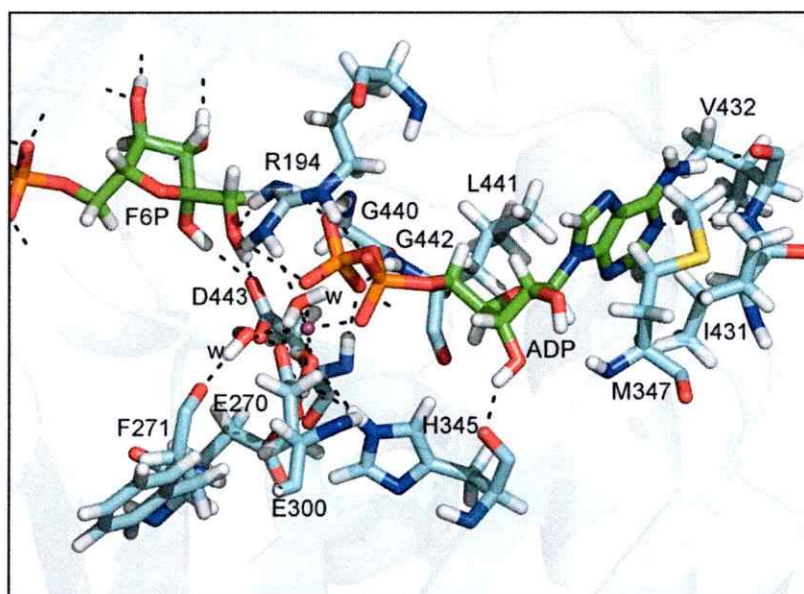


Figure S3. ADP interactions at the active site of the MmPFK/GK homology model. Carbon atom of substrate are shown in green and protein residues in cyan,  $Mg^{2+}$  is show as a magenta sphere and w correspond to coordination water.

Table S2. Evaluation of the ADP-PFK/GK homology models for sequences  
of the order *Methanococcales*

Model	DOPE Potencial	PROCHECK (N° of residues in region of Ramachandran Plot)				VERIFY- 3D Score>0. 2
		Most favoured	Additional allowed	Generously allowed	Disallowe d	
aeolicu.pdb	-55862	401	29	3	0	85.08
fervens.pdb	-54815	389	27	4	2	94.34
formicicus.pdb	-56930	388	25	5	1	87.07
igneus.pdb	-54357	392	28	1	1	87.53
infernus.pdb	-55271	378	30	3	1	89.06
okinawens.pdb	-56736	390	40	3	1	93.89
vanniellii.pdb	-55956	377	39	6	1	89.06
voltae.pdb	-55761	423	37	5	2	67.00
vulcanius.pdb	-55400	395	32	3	1	87.58

Figure S4. Comparison of the ADP-dependent sugar kinases homology models for sequences of the order *Methanococcales* with the *M. maripaludis* homology model.

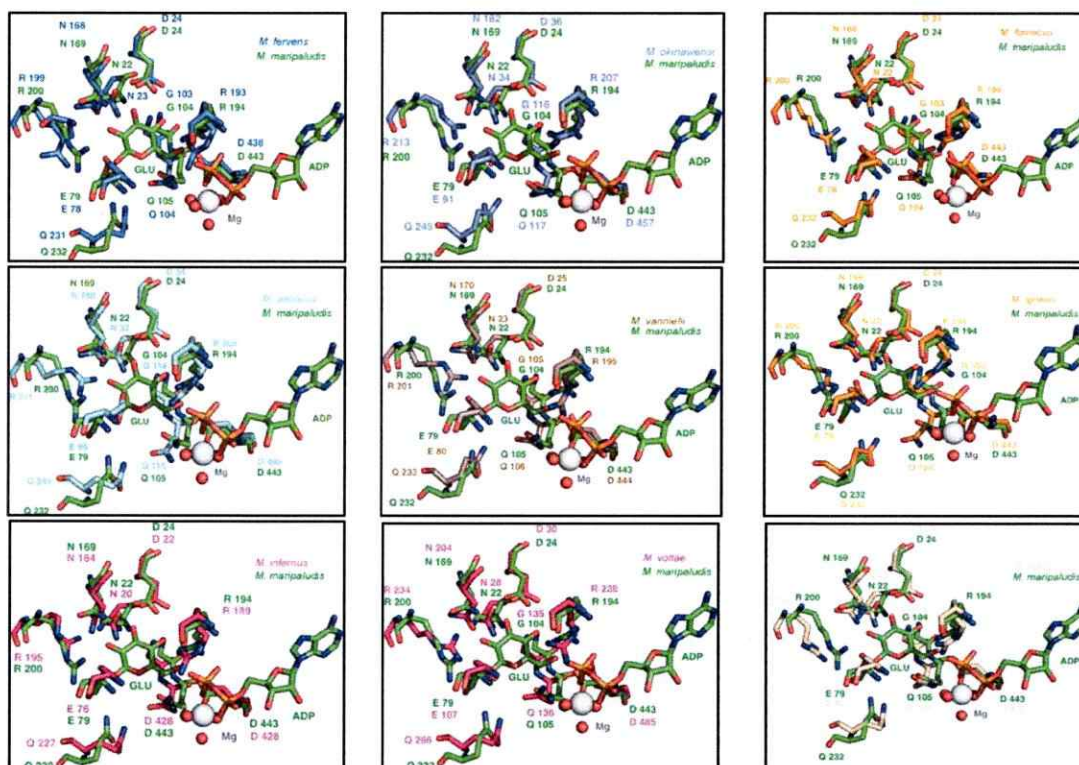
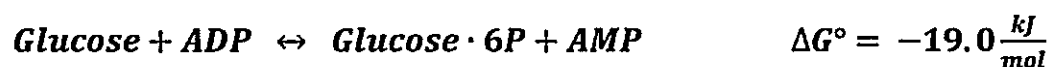
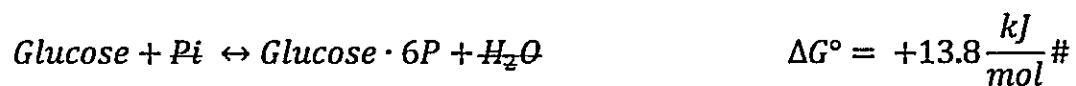
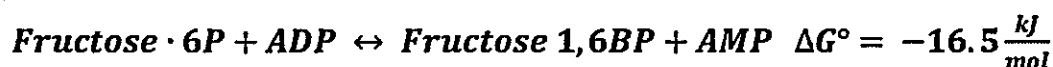
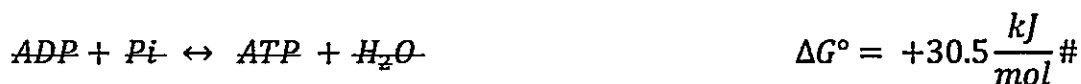
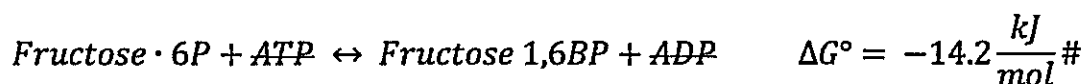




Figure S5. Estimation of standard free energy change and equilibrium constants of the ADP-GK and ADP-PFK reactions



$$\text{ADP} - \text{GK at } 25^\circ\text{C}, \quad K_{eq} = 2.14 \times 10^3$$



$$\text{ADP} - \text{PFK at } 25^\circ\text{C}, \quad K_{eq} = 7.80 \times 10^2$$

# Data taken from: Jencks W P (1976) Handbook of Biochemistry and Molecular Biology, (Fasman GD, ed) 3rd edn. vol I, pp.296-304, CRC Press, Cleveland, Ohio.

#### ***4.5 Análisis del uso de glucosa y fructosa-6P como sustratos en la enzima psicrófila de Methanococcoides burtonii***

En el árbol de la familia, el orden *Methanosarcina* contiene organismos termófilos, mesófilos y psicrófilos y presenta tanto glucoquinasas como fosfofructoquinasas dependientes de ADP. Interesantemente, *Methanococcoides burtonii* presenta las dos secuencias (glucoquinasa y fosfofructoquinasa) de la familia en su genoma (Uniprot: Q12WC0 y Q12WB9). Sin embargo, la secuencia Q12WC0, identificada como glucoquinasa, presenta una remoción en el extremo C-terminal el cual es fundamental para la actividad catalítica, por lo que esta secuencia sería un pseudogen con un producto no funcional, mientras que la otra copia podría complementar ambas actividades si fuera una enzima bifuncional.

Para explorar esta posibilidad, sintetizamos el gen de la proteína optimizado para su expresión en *E. coli* (código Uniprot: Q12WB9) presente en el genoma de *M. burtonii* cepa DSM6242 (Figura 1). La proteína fue expresada en *E. coli*, purificada y hasta el momento hemos caracterizado el uso de ella por glucosa y fructosa-6P.

La enzima presentó actividad con glucosa y fructosa-6P. Las constantes cinéticas muestran que la enzima es 4 órdenes de magnitud más preferente por fructosa-6P respecto a glucosa (Tabla 8), aunque con una actividad glucoquinasa significativa. Estos resultados confirman que las enzimas del orden *Methanosarcinal* son capaces de fosforilar ambos sustratos al igual que las enzimas de los *Methanococcales*.

**Tabla 8 Constantes cinéticas para MbPFK a 25°C**

<b>Fructosa-6P</b>		<b>Glucosa</b>		<b>Fructosa-6P</b>	<b>Glucosa</b>	<b>Rasgo</b>
$K_m(\mu\text{M})$	$k_{\text{cat}}(\text{s}^{-1})$	$K_m(\text{mM})$	$k_{\text{cat}}(\text{s}^{-1})$	$k_{\text{cat}} / K_m$ ( $\text{M}^{-1}\text{s}^{-1}$ )	$k_{\text{cat}} / K_m$ ( $\text{M}^{-1}\text{s}^{-1}$ )	
23±8	7,9±0,1	24±3	0,91±0,03	3,4 x10 <sup>5</sup>	3,8x10 <sup>1</sup>	Bifuncional

#### 4.6 *Discusión*

La bifuncionalidad de las enzimas PFKs de *Methanococcales* es imprescindible para tener una glicolisis completa en estos organismos, ya que no hay evidencia genómica ni experimental de otras quinasas que puedan fosforilar glucosa. Diferente es el caso de los organismos *Methanosarcinales*, los cuales presentan dos copias de quinasas dependientes de ADP, una copia presenta claramente características de glucoquinasa en su secuencia, mientras el otro gen presenta características de PFK. En este último trabajo hemos demostrado que la enzima de *M. burtonnii* es bifuncional. Además, de acuerdo a la alta conservación del sitio activo del resto de las enzimas PFKs del grupo, todos los organismos de éste presentarían una redundancia en cuanto a la fosforilación de glucosa; en el caso particular de *M. burtonnii* sería esta redundancia la que permitiría el flujo a través de la glicólisis, ya que este organismo posee una gran deleción en el extremo C-terminal del gen de la glucoquinasa lo que produciría una proteína no funcional.

El estudiar experimentalmente la especificidad por sustrato en esta familia nos permite asignar ciertos determinantes estructurales a determinadas especificidades y estos determinantes estructurales los podemos transformar a elementos de la secuencia primarias mediante un alineamiento múltiple de secuencias. Esto permite predecir la especificidad que tendrá una enzima basándonos no solo en identidad de secuencias o cercanía filogenética, como asignan las actividades enzimáticas los algoritmos cuando se ingresan los marcos de lectura abiertos en las bases de datos, como por ejemplo luego de la secuenciación de un genoma. De acuerdo a los resultados obtenidos en este trabajo y utilizando la numeración de la enzima *Mm*PFK/GK podemos encontrar los siguientes

motivos relevantes para la especificidad por el sustrato azúcar; motivo 1: <sup>75</sup>G-K-(P/A)-X-(E/A)<sup>79</sup>, motivo 2: <sup>168</sup>(I/V)-(N/H)-(Y/R)-I-(F/Y)-E<sup>173</sup> y motivo 3 <sup>194</sup>R-F-I-X-X-X(R/D)<sup>200</sup>, en negrita y subrayado se encuentran los residuos que interaccionan directamente con los sustratos y que están directamente relacionados a la especificidad. Así, podemos describir que el residuo E<sup>79</sup> del primer motivo está en todas las enzimas que unen glucosa, mientras en las enzimas que no la unen (específicas por fructosa-6P) en esta posición se encuentra un residuo de A<sup>79</sup>. En el segundo motivo, el residuo N<sup>169</sup> se encuentra en todas las enzimas que unen fructosa-6P, mientras que en las enzimas que solo unen glucosa este residuo es reemplazado por una H<sup>169</sup>. Por último, el residuo R<sup>200</sup> del tercer motivo es el que permite la unión de fructosa-6P, ya que este residuo interacciona directamente con el fosfato del sustrato, pero este residuo no restringe la unión de glucosa, ya que se encuentra en todas las enzimas bifuncionales. La ausencia de esta arginina y su reemplazo por un D<sup>200</sup> en las enzimas glucoquinasas explicaría la ausencia de actividad PFK en las glucoquinasas específicas de *Thermococcales*. De acuerdo a esta información podemos identificar a las enzimas glucoquinasas específicas como las que poseen en los motivos descritos los siguientes residuos; E<sup>79</sup>, H<sup>169</sup> y D<sup>200</sup>, mientras que en las enzimas bifuncionales se encuentran los residuos E<sup>79</sup>, N<sup>179</sup> y R<sup>200</sup>, siendo la única diferencia de estas últimas con las enzimas específicas por fructosa-6P el residuo E79. En las PFKs específicas de *Thermococcales* se encuentran los residuos A<sup>79</sup>, N<sup>179</sup> y R<sup>200</sup>

#### 4.7 Conclusiones

- Las enzimas de organismos *Methanococcales* y *Methanosarcinales* son bifuncionales y presentan una mayor especificidad por fructosa-6P respecto a glucosa, aunque ambas actividades serían relevantes *in vivo*.
- Considerando la reversibilidad de la actividad GK de *Mm*PFK/GK esta enzima podría estar involucrada en la síntesis de glucosa en los *Methanococcales*.
- La enzima *Mm*PFK es altamente regulada por el  $Mg^{2+}$  libre y por su producto AMP.
- La especificidad por sustrato en la familia se puede describir con tres motivos de secuencia, los que permiten la identificación de glucoquinasas específicas, fosfofructoquinasas específicas y enzimas bifuncionales.



## **5 CAPÍTULO TRES**

**Estructura, estabilidad térmica y cambios  
conformacionales durante la evolución de la  
familia de quinasas de azúcar dependientes de  
ADP en arqueas.**

## *5.1 Introducción*

Desde que se comenzaron a determinar estructuras de proteínas comenzó a quedar claro que a través de la evolución la estructura de las proteínas es mucho más conservada que la secuencia. De hecho, hay numerosos ejemplos de proteínas con estructuras muy conservadas entre las que no se puede detectar similitud de secuencia [88], como es el caso de la superfamilia Riboquinasa, donde la estructura de las quinasa dependientes de ADP es muy similar a la estructura de la familia Riboquinasa, pero no es posible encontrar similitud utilizando solo las secuencias [70]. A pesar de la alta conservación estructural que muestran las proteínas homologas, muchas de ellas muestran características estructurales que las diferencian, las cuales pueden ser inserción, delección o sustitución de elementos de estructura secundaria, permutaciones de estos elementos o re arreglos de hojas  $\beta$  y serían este tipo de cambios los que permitirían la evolución de los plegamientos en las proteínas. Por otra parte, hay pocos estudios respecto a la conservación de los cambios conformacionales importantes para la función a través de la evolución y a la correlación entre las distintas conformaciones que puede adoptar una proteína, su conservación estructural y el mecanismo de funcionamiento como por ejemplo el orden de entrada de los sustratos en una enzima.

Hasta la fecha solo dos metodologías permiten determinar estructuras a nivel atómico, por una parte está la resonancia magnética nuclear, la cual tiene como limitante la elevada concentración de proteína en solución y la cantidad de residuos que puede resolver, actualmente se utiliza comúnmente con proteínas con menos de 200 aminoácidos. Por otro lado, está la difracción de rayos X, la cual necesita tener la



muestra en un arreglo periódico en forma de cristal, lo cual limita enormemente el tipo de proteínas que se pueden estudiar con esta metodología [89]. La difracción de rayos X también se puede utilizar con proteínas en solución para obtener información a baja resolución de la superficie de la proteína y de radios de giros, esta metodología se denomina dispersión de rayos X a bajo ángulo o SAXS (del inglés *Small-angle X-ray scattering*). También existen otras técnicas que permiten obtener información estructural desde proteínas en solución aunque a un nivel mucho menos detallado como el dicroísmo circular (DC).

El dicroísmo circular es una técnica espectroscópica que permite obtener información estructural de proteínas en solución. Esta técnica se basa en la absorción diferencial de la luz circularmente polarizada, la que puede descomponerse en dos componentes circularmente polarizados; uno componente izquierdo o L (del inglés *Left*) y un componente derecho o R (del inglés *Right*). Un cromóforo que sea ópticamente activo (quiral) absorberá más de un componente respecto a la otro ( $A_R < A_L$  o  $A_R > A_L$  donde A es absorbancia), causando una diferencia de absorbancia entre los dos componentes ( $\Delta A = A_R - A_L$ ) y provocando que la luz neta (R+L) presente elipticidad ( $\theta$ ). Comúnmente los espectropolarímetros miden la diferencias de absorbancia entre los componentes L y R ( $\Delta A = A_R - A_L$ ), pero por razones históricas la unidad en que se reporta es la elipticidad ( $\theta$ ) en grados. Esta elipticidad está dada por  $\theta = \tan^{-1}(b/a)$ , donde b y a son los ejes menor y mayor de la elipse resultante de la absorción diferencial de los componentes R y L, la relación entre  $\Delta A$  y  $\theta$  está dada por  $\theta = 32,98 \cdot \Delta A$  [90].

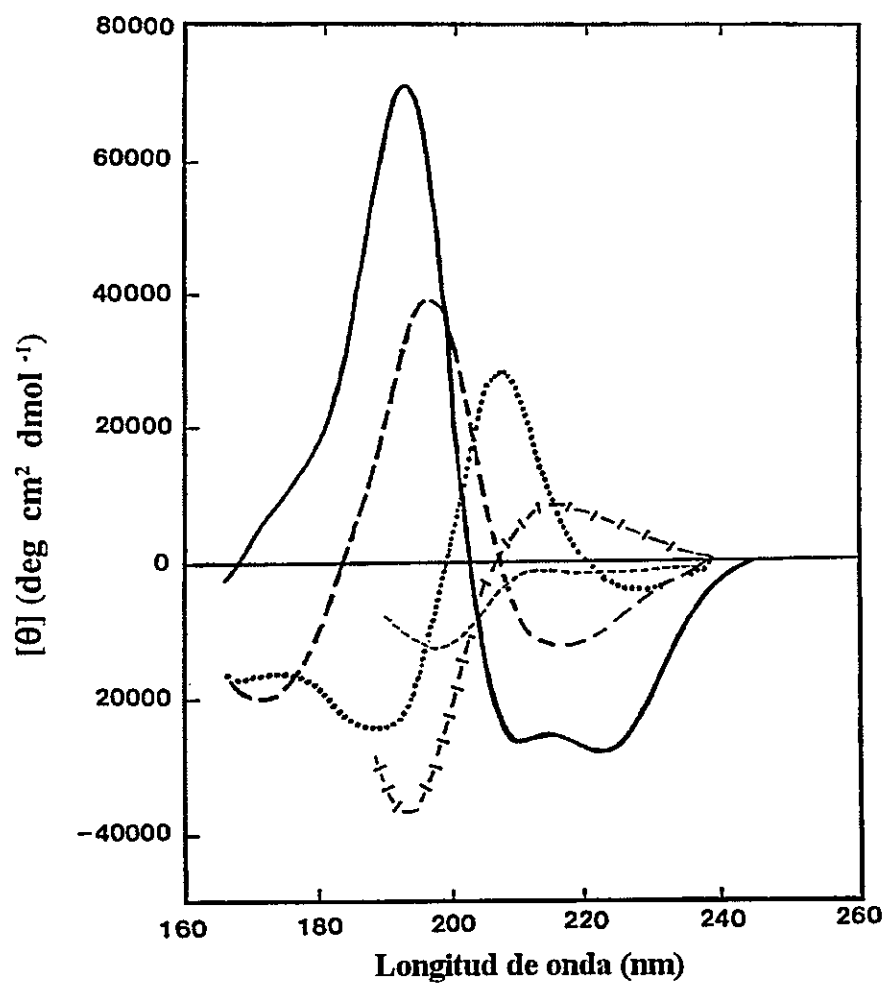
El espectro de dicroísmo es obtenido cuando la señal de dicroísmo es medida en función de la longitud de onda y dependiendo de la región espectral analizada se puede obtener información complementaria; así en la región del UV lejano (bajo los 240 nm) se extrae información de la estructura secundaria que adoptan los enlaces peptídicos, la región del UV cercano (260-320 nm) entrega información del ambiente de los aminoácidos aromáticos, es decir de la estructura terciaria de las proteínas y por otra parte la absorción alrededor de los 260 nm arroja información del ambiente de los enlaces disulfuros.

El análisis de los espectros de DC en el UV lejano es el de uso más común en proteínas. Esta región permite analizar la composición de estructura secundaria que adoptan los enlaces peptídicos, pudiéndose analizar los porcentajes de hélices, hojas y vueltas o estructuras al azar. La absorción en esta región está dada principalmente por los enlaces peptídicos o amidas en donde los electrones de este enlace presentan dos transiciones principales; una débil y amplia en torno a los 220 nm dada por la transición de los electrones entre los orbitales  $n \rightarrow \pi^*$  y otra más intensa alrededor de los 190 nm dada por la transición entre los orbitales  $\pi \rightarrow \pi^*$ [90].

Varios métodos empíricos han sido desarrollados para estimar cuantitativamente el contenido de estructura secundaria de proteínas a partir de espectros de CD en el UV lejano. El principio básico de estos métodos se basa en que un espectro de DC ( $E_\lambda$ ) puede ser expresado como una combinación lineal de los espectros de los componentes individuales de estructura secundaria encontrados en proteínas ( $S_{k\lambda}$ ) donde  $k$  puede ser hélice, hojas u otro elemento de estructura secundaria (**Figura 16**). Así el espectro de

una proteína que contiene distintos elementos de estructura secundaria se puede expresar como  $E_{\lambda} = \sum f_k \cdot S_{k\lambda}$ , donde  $f_k$  es la fracción del componente de estructura secundaria  $k$ . Los métodos más avanzados utilizan además bases de datos de espectros de DC de proteínas con estructura tridimensional conocida como referencia y con algoritmos que le permiten optimizar las variables para minimizar las diferencias entre los espectros calculados y los experimentales (RMSD) [91] proporcionando así una estimación más exacta de los porcentajes de cada componente.

En la reconstrucción de la filogenia de esta familia (Figura 3) se observan organismos con gran diversidad de nichos ecológicos. En particular, es interesante observar las diferentes temperatura a las cuales estos organismos se desarrollan y por ende las diferentes temperaturas en la que sus proteínas trabajan. Las quinasas de azúcar dependientes de ADP se encuentran en arqueas del orden *Thermococcal* las cuales son todas hipertermófilas, en organismos del orden *Methanococcales* que presenta organismos termófilos (género *Methanocaldococcus*, *Methanotorris* y *Methanothermococcus*) y organismos mesófilos (género *Methanococcus*) y arqueas del orden *Methanosarcina* que presenta organismos psicrófilos, mesófilos y termófilos. Dada esta diversidad, es interesante evaluar cómo ha evolucionado la termoestabilidad de las proteínas de esta familia, ya que se han reportado reconstrucciones ancestrales de proteínas que presentan elevada estabilidad frente a la temperatura, respecto a las proteínas actuales mesófilas [92,93].



**Figura 16. Espectros de DC para diferentes tipos de estructuras secundarias encontradas en proteínas.**

Hélice  $\alpha$  (línea continua), hoja  $\beta$  antiparalela (línea con guiones largos), vueltas  $\beta$  tipo I (línea de puntos), hélice  $3_1$  extendida o hélice poli (Pro) II (línea de guiones cruzados), estructura irregular (línea de guiones cortos). Adaptada de [90].

Las estructuras determinadas experimentalmente en la familia corresponden a GKs de *P. furiosus*, *P. horikoshi*, *T. litoralis*, y a la enzima PFK de *P. horikoshii*. Estas estructuras se encuentran cristalizadas en distintas condiciones que no son directamente comparables. Por ejemplo, la GK de *T. litoralis* se encuentra cristalizada con ADP [81], pero los autores no encontraron diferencias en relación a la posición de los dominios mayor y menor cuando determinaron la estructura a partir de cristales lavados con solución de cristalización libre de ADP. Por otra parte Tsuge y col cuando determinaron la estructura de la GK de *P. horikoshii* en forma apo, encontraron una conformación más abierta dada por una gran diferencia en la posición del dominio menor de la PhGK cuando fue comparada con la estructura TlGK-ADP [94]. Posteriormente se determinó la estructura de la GK de *Pyrococcus furiosus* en presencia de un análogo de ADP(ADP $\beta$ S) y glucosa [40], la que se encuentra en una conformación más cerrada que las anteriores, por lo que apoya la hipótesis de un cambio conformacional inducido por ligando. Dado lo anterior es importante determinar si los cambios que muestran las estructuras cristalográficas son importantes para la catálisis y analizar a través de la evolución, la conservación de estos cambios conformacionales en la familia.

## **5.2 Hipótesis**

La estructura de los ancestros y proteínas actuales se encuentra conservada, al igual que los cambios conformacionales entre estados abierto/cerrado los cuales son relevantes para la catálisis en la familia de quinasas dependientes de ADP. Además, las proteínas ancestrales de esta familia fueron proteínas termófilas.

## **5.3 Objetivos**

Determinar las fracciones de elementos de estructura secundaria y la termoestabilidad entre proteínas actuales y ancestros.

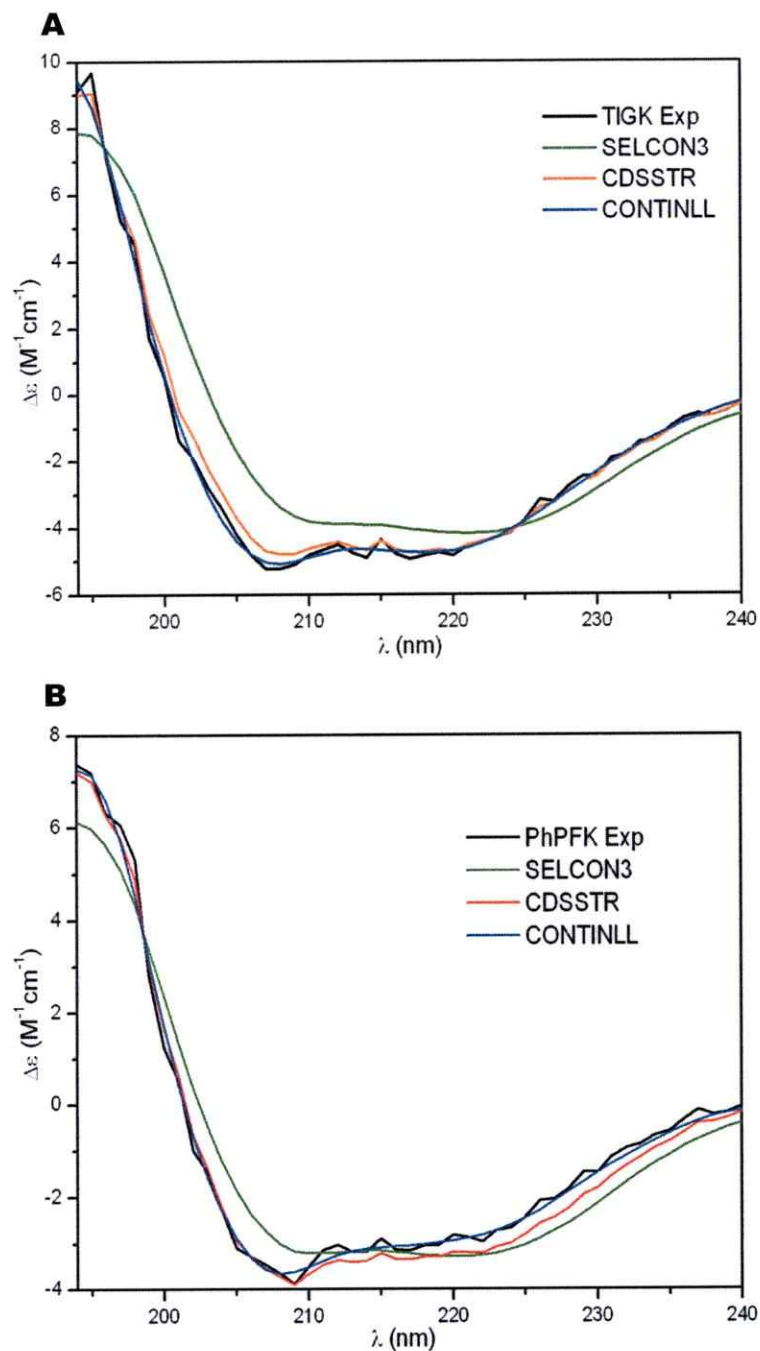
Determinar estructuras cristalográficas en distintas conformaciones y analizar la correlación con los mecanismos cinéticos de las enzimas.

## 5.4 Resultados

### 5.4.1 Análisis del contenido de estructura secundaria a través de la evolución en la familia de quinasas dependientes de ADP mediante dicroísmo circular

Para analizar la conservación de la estructura secundaria durante la evolución hemos obtenidos los espectros de dicroísmo circular y calculado los porcentajes de estructura secundaria para las proteínas actuales *TIGK*, *PhPFK* y *MjPFK/GK* y para los ancestros ancGK, ancPFK-MT y ancPFK-T. Hasta la fecha sólo se ha reportado la estructura tridimensional de *TIGK* y *PhPFK*. Para determinar el porcentaje de estructura secundaria de las proteínas se utilizó el paquete de programas CDpro [91] el cual incluye tres programas distintos para el análisis de espectros de dicroísmo circular; SELCON3, CDSSTR y CONTINLL (**Figura 17**). Para evaluar el programa que entrega los mejores resultados para esta familia de proteínas, se utilizó como control la asignación de estructura secundaria del programa DSSP a las estructuras cristalográficas de *TIGK*(1GC5) y *PhPFK*(3DRW) (**Tabla 9**).

De acuerdo a las estructuras cristalográficas, los programas CDSSTR y CONTINLL muestran las mejores asignaciones de porcentaje de estructuras secundarias. Estos programas también muestran menores diferencias entre los espectros calculados y experimentales, de acuerdo a los RMSD. Entre los dos programas, los resultados de CDSSTR tienen una mayor exactitud al estimar los porcentajes de estructura secundaria, a pesar que los RMSD entre los espectros experimentales y calculados sea levemente superior comparado con los obtenidos con CONTINLL. Para las otras proteínas, solo se consideró el cálculo de porcentajes de estructura secundaria hecho por CDSSTR.



**Figura 17. Comparación entre espectro experimental de TIGK y PhPFK y los espectros calculados por los programas SELCON3, CDSSTR y CONTINLL.**  
 Espectros de DC para TIGK (A) y PhPFK (B), la línea negra representa el espectro experimental luego de la remoción del ruido mediante el filtro FFT.

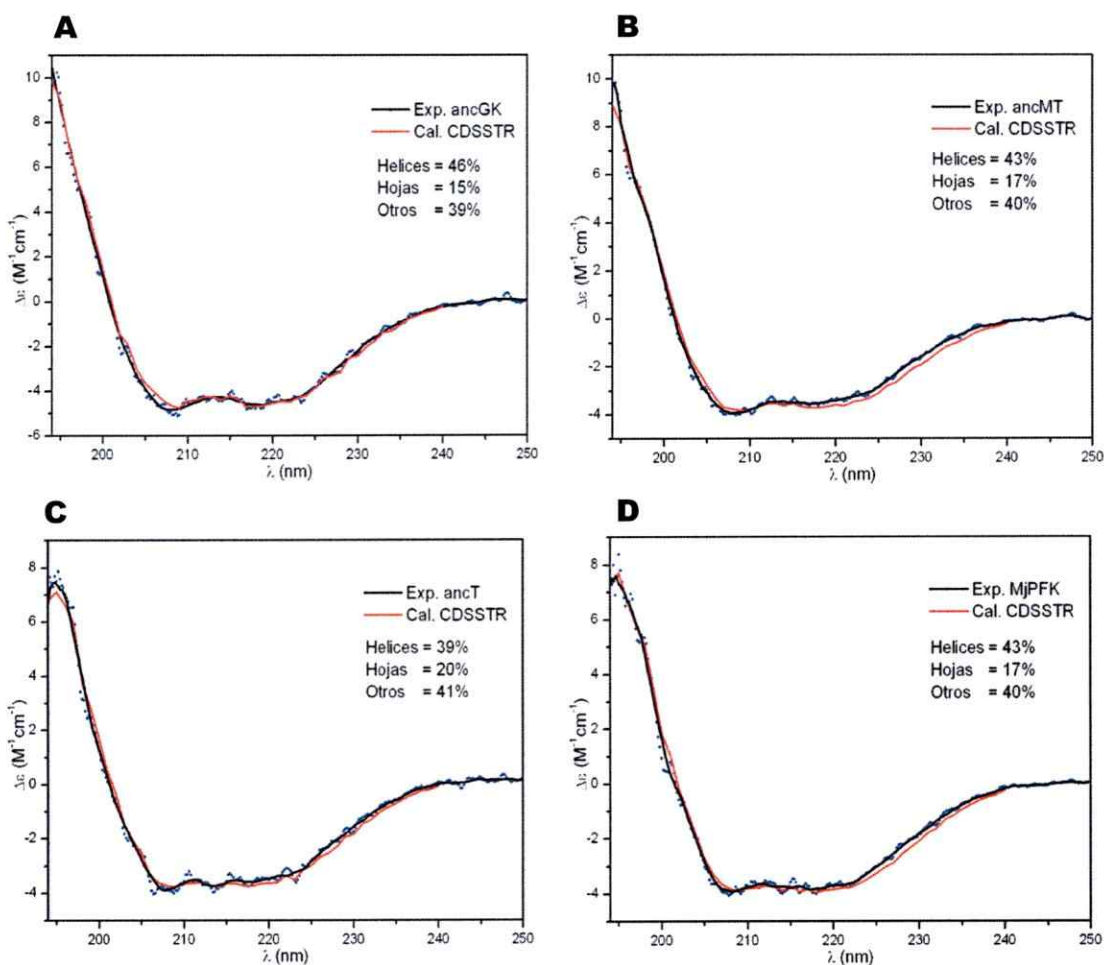


**Tabla 9. Porcentajes de estructura secundaria a partir de la estructura cristalográfica y espectros de DC**

Tipo	Estructura <i>TIGK 1GC5</i> (DSSP)	SELCON3	CDSSTR	CONTINLL
<b>Hélices</b>	47%	45%	<b>47%</b>	42%
<b>Hojas</b>	19%	13%	<b>17%</b>	15%
<b>Otras</b>	34%	42%	<b>36%</b>	43%
<b>RMSD (Exp-Cal)</b>	---	1,474	<b>0,326</b>	0,260

Tipo	Estructura <i>PhPFK 3DRW</i> (DSSP)	SELCON3	CDSSTR	CONTINLL
<b>Hélices</b>	43%	37%	<b>44%</b>	36%
<b>Hojas</b>	19%	14%	<b>15%</b>	18%
<b>Otras</b>	38%	49%	<b>41%</b>	46%
<b>RMSD (Exp-Cal)</b>	---	0,700	<b>0,278</b>	0,200



**Figura 18. Espectros de dicroísmo circular para las proteínas sin estructura cristalográfica.**

Espectros de dicroísmo circular experimentales (Exp: línea negra) y ajustes calculados con el programa CDSSTR (Cal. línea naranja). A, ancGK. B, ancPFK-MT. C, ancPFK-T y D, *MjPFK*.

**Tabla 10. Comparación de porcentajes de estructuras secundarias calculadas con CDSSTR**

	% Hélice	% Hojas	% Otros
ancPFK-MT	43	17	40
ancPFK-T	39	20	41
<i>PhPFK</i>	44	15	41
<i>MjPFK</i>	43	17	40
ancGK	46	15	39
<i>TIGK</i>	47	17	36

Los resultados de los espectros de CD para proteínas actuales y ancestrales muestran que el contenido de estructura secundaria es altamente conservado entre todas las proteínas evaluadas, lo que revela que las proteínas ancestrales inferidas presentan proporciones de elementos de estructura secundaria muy similares a los de las proteínas actuales. Esto indica que el plegamiento de las enzimas ancestrales es similar al de las proteínas actuales y que la conservación de su estructura se ha mantenido través de la evolución, en acuerdo con lo propuesto por las teorías de evolución de proteínas [20].

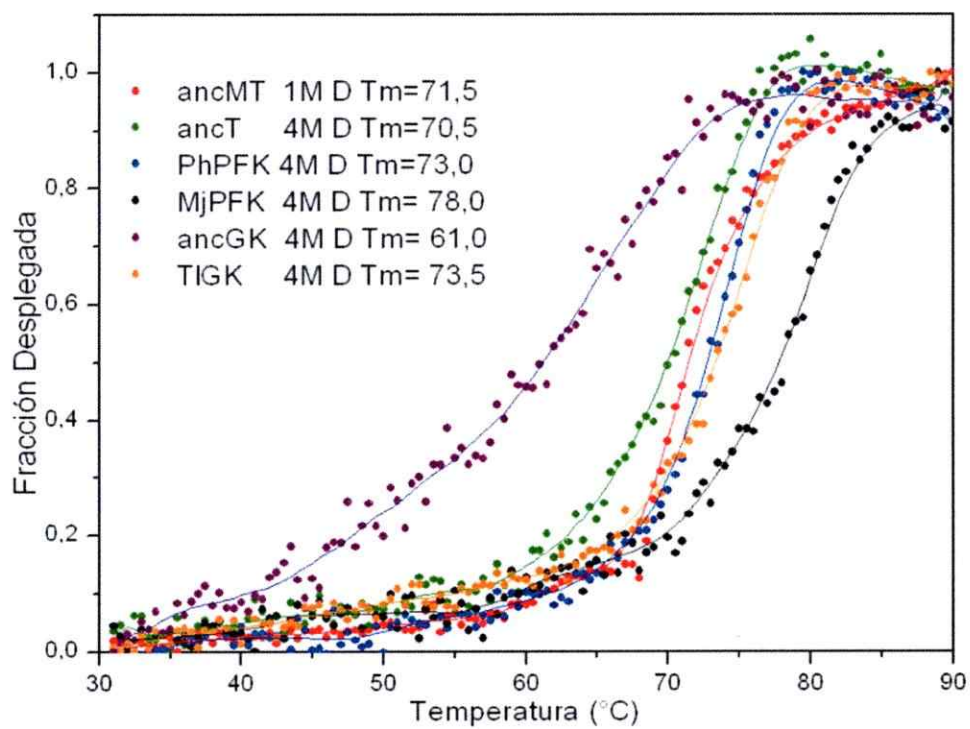
#### 5.4.2 Análisis de la estabilidad térmica de las proteínas a través de la evolución de la familia

Para determinar la temperatura media de desplegamiento ( $T_m$ ) (ensayos fuera del equilibrio) se realizaron rampas de temperatura siguiendo la elipticidad a 220 nm (Figura 19). Dado que las proteínas hipertermófilas se despliegan sobre los 100 °C, agregamos un agente caotrópico (cloruro de guanidinio) para observar el desplegamiento bajo los 100 °C.

Las  $T_m$  muestran que en el grupo de las PFKs, el ancPFK-MT es la proteína menos estable. Aun cuando no se observa despelgamiento de esta proteína hasta los 90 °C en ausencia de cloruro de guanidinio, sólo es necesaria una concentración de 1M del agente para lograr desplegamiento bajo los 90 °C. Le sigue en estabilidad el ancestro ancPFK-T; en tanto que las proteínas termófilas actuales (*Ph*PFK y *Mj*PFK/GK) mostraron ser las más estables frente a la temperatura. En el grupo de las GKs, el

ancestro ancGK también mostró ser menos estable que la proteína termófila actual (*TIGK*).

Aunque todos los ancestros mostraron ser termófilos, estos son menos resistentes a la denaturación térmica que las proteínas termófilas actuales, lo que indicaría que la resistencia al desplegamiento térmico ha ido en aumento hacia las formas termófilas actuales.



**Figura 19. Curvas de desplegamiento térmico.**

Desplegamiento seguido por microscopio circular a 220nm. En el inserto de la figura se indica la T<sub>m</sub> para cada proteína y la concentración de denaturante (D: cloruro de guanidinio) utilizada. Velocidad de 60 °C/hora.

#### 5.4.3 Conservación de las interacciones involucradas en la transición abierta/cerrada en la glucoquinasa dependiente de ADP de *Thermococcus litoralis* en la familia de quinasas ADP dependiente

Con el fin de determinar si los cambios conformacionales evidenciados por las estructuras cristalográficas son relevantes para la catálisis y estudiar la conservación de estos cambios conformacionales en la familia, se estudió la TIGK mediante cinética enzimática, cristalografía de rayos X, dispersión de rayos X de bajo ángulo en solución (SAXS) y se realizó además un análisis evolutivo de las interacciones putativas que permitirían el cambio de conformaciones.

A continuación se presenta el trabajo publicado en la revista *PlosOne* (2013), 8, e66687. El estudio fue ideado por Jaime A. Rivas-Pardo y Victoria Guixé, el trabajo fue encabezado por Jaime A. Rivas-Pardo quien determinó el mecanismo cinético (ayudado por Alejandra Herrera-Morande y Victor Castro-Fernández) y realizó los estudios de SAXS. Alejandra Herrera-Morande, Francisco J. Fernandez y M. Cristina Vega determinaron las estructuras de la TIGK en forma apo y holo y Victor Castro-Fernandez realizó el análisis de las estructuras en cuanto a los potenciales electrostático, los cluster de residuos y su conservación a través de la evolución de la familia.



# Crystal Structure, SAXS and Kinetic Mechanism of Hyperthermophilic ADP-Dependent Glucokinase from *Thermococcus litoralis* Reveal a Conserved Mechanism for Catalysis

Jaime Andrés Rivas-Pardo<sup>1</sup>, Alejandra Herrera-Morande<sup>1,2</sup>, Víctor Castro-Fernandez<sup>1</sup>, Francisco J. Fernandez<sup>2</sup>, M. Cristina Vega<sup>2\*</sup>, Victoria Guixé<sup>1\*</sup>

<sup>1</sup> Departamento de Biología, Facultad de Ciencias, Universidad de Chile, Santiago, Chile, <sup>2</sup> Centro de Investigaciones Biológicas (CIB-CSIC), Madrid, España

## Abstract

ADP-dependent glucokinases represent a unique family of kinases that belong to the ribokinase superfamily, being present mainly in hyperthermophilic archaea. For these enzymes there is no agreement about the magnitude of the structural transitions associated with ligand binding and whether they are meaningful to the function of the enzyme. We used the ADP-dependent glucokinase from *Thermococcus litoralis* as a model to investigate the conformational changes observed in X-ray crystallographic structures upon substrate binding and to compare them with those determined in solution in order to understand their interplay with the glucokinase function. Initial velocity studies indicate that catalysis follows a sequential ordered mechanism that correlates with the structural transitions experienced by the enzyme in solution and in the crystal state. The combined data allowed us to resolve the open-closed conformational transition that accounts for the complete reaction cycle and to identify the corresponding clusters of aminoacids residues responsible for it. These results provide molecular bases for a general mechanism conserved across the ADP-dependent kinase family.

**Citation:** Rivas-Pardo JA, Herrera-Morande A, Castro-Fernandez V, Fernandez FJ, Vega MC, et al. (2013) Crystal Structure, SAXS and Kinetic Mechanism of Hyperthermophilic ADP-Dependent Glucokinase from *Thermococcus litoralis* Reveal a Conserved Mechanism for Catalysis. PLoS ONE 8(6): e66687. doi:10.1371/journal.pone.0066687

**Editor:** Beata G. Vertessy, Institute of Enzymology of the Hungarian Academy of Science, Hungary

**Received:** January 29, 2013; **Accepted:** May 10, 2013; **Published:** June 20, 2013

**Copyright:** © 2013 Rivas-Pardo et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by Fondo Nacional de Desarrollo Científico y Tecnológico FONDECYT grant 1110137 to V.G., Spanish Ministry of Science and Innovation grants PET2008\_0101, BIO2009-11184 and BFU2010-22260-C02-02 and the EC project ComplexINC (Framework Programme 7 (FP7) under grant agreement no. 279039) to M.C.V. The SAXS measurements (SAXS1 and SAXS2 beamlines) were supported by LNLS (Laboratório Nacional de Luz Síncrotron), Sao Paulo, Brazil, to V.G. and J.A.R.-P. We also thank the Departamento de Postgrado y Postítulo, Universidad de Chile by the support given to J.A.R.-P. to accomplish the SAXS studies in Sao Paulo. J.A.R.-P. and V.C.-F. are PhD fellowship from Comisión Nacional de Investigación Científica y Tecnológica de Chile. A.H.-M. acknowledges the support of the PhD program in Molecular Biotechnology of the Universitat de Barcelona. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: cvega@cib.csic.es (MCV); vguixe@uchile.cl (VG)

## Introduction

The Embden–Meyerhof pathway is the most common route for the degradation of glucose. Although this metabolic pathway is generally highly conserved between different organisms, several archaea of the *Euryarchaeota* have evolved major modifications with only four of the classical enzymes present in the canonical pathway. One of these major differences is the presence of ADP-dependent glucokinases (GKs) and phosphofructokinases (PFKs), instead of the classical ATP-dependent kinases [1,2]. Although some authors have attributed the presence of ADP-dependent enzymes to a matter of metabolic adaptation to high temperatures and to starvation conditions, several facts indicate that the presence of these proteins in the central metabolism of archaea is not related to the hyperthermophilic life style [3]. Even though these ADP-dependent kinases show no sequence similarity to ATP-dependent enzymes known to date, the determination of their three dimensional structures allowed their reliable classification as members of the ribokinase superfamily. Structurally, these enzymes share a

common Rossmann-like fold characterized by a  $\alpha/\beta/\alpha$  topology, which constitutes the large domain. In addition to this core ribokinase-like fold; other members of this superfamily have an extra small domain, which in the case of the ADP-dependent kinases is formed by a five stranded  $\beta$ -sheet with some  $\alpha$ -helical insertions, the active site lying between the two domains. The small domain, which could function as an active-site lid to protect substrates from hydrolysis, has been proposed as a phylogenetic marker for the evolution of this superfamily [4].

To date the crystal structures of three ADP-dependent GKs are known [5–7]. These structures come from different organisms and were crystallized under experimental conditions that are not directly comparable. For example, Ito et al. [5] reported the first structure of an ADP-dependent kinase, the glucokinase from *T. litoralis* (TIGK) complexed with ADP. These authors found no evidence for a conformational change induced by ADP by comparing the ADP complex with the structure of the apo-TIGK, prepared by soaking a holo crystal in ADP-free solution, thereby raising the question as to whether an induced fit mechanism was required for catalysis. Later, Tsuge et al. [6]

determined the crystal structure of apo-GK from *Pyrococcus horikoshii* (PhGK) at 2.0 Å resolution; when the structure of apo-PhGK was superimposed onto that of TIGK·ADP, a displacement by >5 Å of the small domain was observed, that led the authors to suggest that this large conformational change could take place during catalysis. Finally, the crystal structure of an ADP-GK from *Pyrococcus furiosus* (PiGK), complexed with ADPβS and glucose was reported, and the authors stressed the importance of glucose binding to achieve the closed conformation. They referred to the apo and ADP complexed forms as more open and open conformations [7].

Binding of ligands to proteins can trigger conformational changes, and conformational changes of proteins can increase binding affinity. The interplay between conformational changes and ligand binding is usually described in terms of induced fit or conformational selection models. The classical view corresponds to the work of Koshland and establishes an *induced fit* mechanism whereby ligand binding provokes a conformational change in the protein [8]. The other view, which has become more popular in recent years, suggests that the protein exists as an ensemble of conformations, a fraction of which is able to recognize and bind a particular ligand, favored by the intrinsic dynamics of the protein (*conformational selection*) [9,10]. However, as reviewed by Vertessy and Orosz [11], this concept was first described by the Hungarian biochemist F. B. Straub, who coined the term "fluctuation Fit" to hypothesize that proteins may exist in equilibrium between several different conformations. Also, other less used terms like "conformational selectivity", "stabilization of conformational ensembles", "population shift", "selected fit" and "pre-existing equilibrium" have been used to describe this phenomenon. To date, many studies show an even more complex interplay between intrinsic dynamics and ligand-induced motions [12,13]. Recently, Vogt and Di Cera have offered an explanation for the apparent infrequency reporting cases of conformational selection in the literature [14].

Since the crystal structures provide only static pictures defining snapshots along the catalytic cycle while necessarily missing about the dynamics and timing of the domain movements that occur in solution, it is necessary to assess whether the different conformations observed in the crystalline state are relevant to catalysis and related to the conformational behavior of the enzyme in solution. To address these questions for ADP-dependent TIGK, we employed a multidisciplinary approach combining the determination of the kinetic mechanism in order to evaluate the catalytically relevant enzyme forms to which substrate are bound, small angle X-ray scattering (SAXS) as a sensitive probe of the solution conformation in the absence and presence of ligands, and finally X-ray crystal structure determinations of the enzyme in the apo form and as a ternary complex form; this afforded the direct comparison of the structures of the same enzyme in all ligand states.

Our results provide insights into determinants of the structural changes that accompany ligand binding both in solution and in the crystalline state. We determine the binding order of each substrate to the active site in order to correlate the conformational changes that the enzyme experiences on the basis of its kinetic mechanism, with those observed by SAXS and X-ray crystallography. Detailed inspection of conserved clusters of residues involved in the open-closed conformational transition lead us to propose a general mechanism for the open-closed conformational transition that is a hallmark throughout the ADP-dependent kinases family.

## Results

### Kinetic Mechanism

To elucidate the order of addition of substrates and release of products, steady-state kinetics approaches are followed to measure initial reaction rates and to test the nature of inhibition by product [15,16]. In order to ascertain if the kinetic mechanism proceeds through a substituted-enzyme mechanism (ping-pong mechanism) or ternary complex formation (sequential mechanism), we performed initial velocity studies. TIGK presents a hyperbolic saturation curve for both substrates, Mg·ADP and D-glucose (Figure 1A and B). When Mg·ADP was varied at fixed D-glucose concentration an increase in the initial velocity was observed as D-glucose concentration was raised. The same trend was obtained when D-glucose was varied at fixed concentrations of Mg·ADP. In both cases, when data were analyzed using double reciprocal plots a family of lines intersecting above the abscissa was observed (Figure 1C and D). The observed intersection of lines is characteristic of a sequential mechanism where a ternary central complex must form before products are released. Table 1 summarizes the kinetic constants assuming a bi-substrate Bi-Bi model under steady state conditions, obtained from secondary plots of the slope and intercepts of double reciprocal plots.

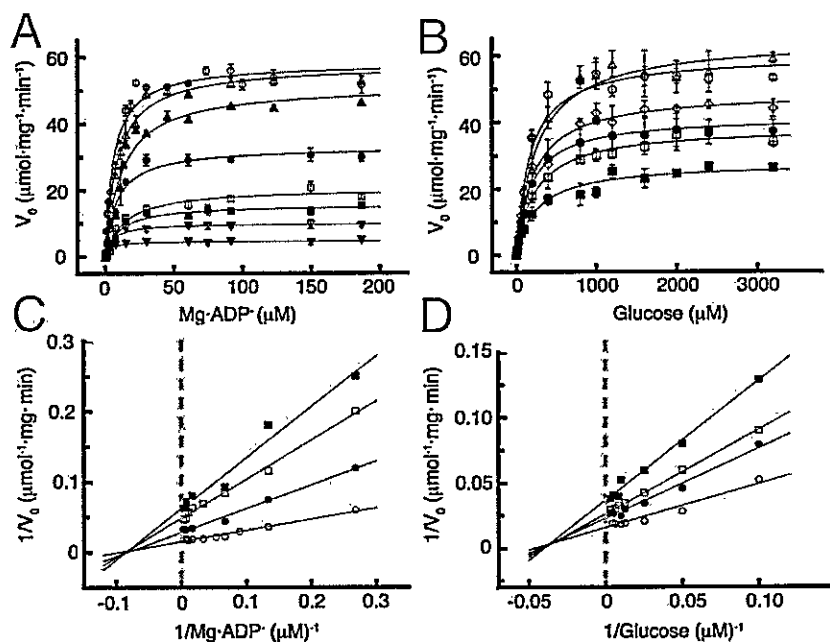
To distinguish whether substrate binding to the active site was random or ordered, we performed product inhibition studies. Our results indicated that Mg·AMP behaves as a competitive inhibitor of Mg·ADP and as a mixed type inhibitor with respect to D-glucose; the other product, D-glucose-6-P presents a mixed type inhibition versus either Mg·ADP or D-glucose (Figure 2). These patterns of inhibition are consistent with an ordered sequential mechanism in which Mg·ADP is the first substrate to bind to the catalytic site and Mg·AMP the last product to be released. Replotting of the slopes and intercepts of the double reciprocal plots obtained by product inhibition showed a linear dependence of the slopes and intercepts with the inhibitor concentration (not shown) indicating the existence of a single product inhibition without the formation of dead end complexes. Our results fully agree with a reaction pathway involving the formation of a ternary complex as a compulsory step, which strongly support an ordered sequential. Table S1 summarizes the inhibition constants obtained from fitting the data to an ordered sequential mechanism, where substrate A is Mg·ADP and B is D-glucose (Figure 3).

### Ligand-Induced Structural Transitions Observed by SAXS

In order to correlate catalytically relevant binding of substrates determined through the kinetic studies with the putative structural transitions induced by them, and to understand their interplay with the GK function we used solution small X-ray scattering (SAXS). This technique provides a sensitive probe of the macromolecular conformation in solution and has been applied to answer questions about conformation and conformational changes in several molecular systems [17].

Scattering curves for TIGK in solution demonstrated that the molecule changes its shape upon binding of the substrates, especially at angles  $q < 0.15 \text{ \AA}^{-1}$ . The shape of the scattering curves for the apo form as well as in the presence of D-glucose were almost identical, the only difference between them being a decrease in the dispersion of the data at angles  $q > 0.2 \text{ \AA}^{-1}$  when D-glucose was present (Figure 4A). However, in the presence of Mg·ADP the scattering curve exhibited a dramatic change. The dispersion decreased and the slope of the scattering curve became clearly different, especially at angles  $q < 0.1 \text{ \AA}^{-1}$  and at high angles  $q > 0.2 \text{ \AA}^{-1}$  (Figure 4A). To obtain structural information of the ternary complex we used ADPβS as a nonhydrolyzable





**Figure 1. Initial velocity patterns for TIGK with D-glucose and MgADP as variable substrates.** (A) Saturation curves for Mg-ADP at different constant concentrations of D-glucose. The fixed initial D-glucose concentrations were: 30  $\mu\text{M}$  ( $\nabla$ ), 50  $\mu\text{M}$  ( $\nabla$ ), 100  $\mu\text{M}$  ( $\blacksquare$ ), 150  $\mu\text{M}$  ( $\square$ ), 200  $\mu\text{M}$  ( $\bullet$ ), 250  $\mu\text{M}$  ( $\blacktriangle$ ), 500 ( $\Delta$ ) and 1000  $\mu\text{M}$  ( $\circ$ ). (B) Saturation curves for D-glucose at different constant MgADP concentrations; 10  $\mu\text{M}$  ( $\blacksquare$ ), 30  $\mu\text{M}$  ( $\square$ ), 50  $\mu\text{M}$  ( $\bullet$ ), 100  $\mu\text{M}$  ( $\diamond$ ), 300  $\mu\text{M}$  ( $\Delta$ ) and 1000  $\mu\text{M}$  ( $\circ$ )MgADP. (C) Double reciprocal plots of curves shown in A. For clarity only four concentrations are depicted: 100, 150, 200 and 1000  $\mu\text{M}$  D-glucose. (D) Double reciprocal plots of curves shown in B. For clarity only four concentrations are shown: 10, 30, 50 and 1000  $\mu\text{M}$  Mg-ADP.  
doi:10.1371/journal.pone.0066687.g001

ADP analog, which binds to the active site but cannot participate as a substrate for the transfer reaction (data not shown). The scattering curve of the Mg $\cdot$ ADP $\beta$ S-D-glucose complex, analogous to the ternary complex, had an even lower dispersion and the slope of the curve at high angles was similar to the one obtained for Mg $\cdot$ ADP, whereas at intermediate angles ( $q \approx 0.1 \text{ \AA}^{-1}$ ) a slope difference was observed (Figure 4A).

Analysis of the pair distance distribution functions  $P(r)$  revealed the conformational changes triggered by ligands. The radius of gyration ( $R_g$ ) of the apo-enzyme was 29.0  $\text{\AA}$  while in the presence of Mg-ADP was reduced to 25.9  $\text{\AA}$ , confirming the formation of the E $\cdot$ A complex (Figure 4B). However, in the presence of D-glucose the measured  $R_g$  was 30.0  $\text{\AA}$  (Figure 4B), which is almost identical to the one obtained for the apo-enzyme (Table 2), indicating the absence of complex formation in the presence of

glucose only. Other experimental approaches support the ligand binding events that were derived by kinetic and SAXS experiments. For example, isothermal titration calorimetry experiments demonstrated binding of MgADP to the enzyme in the absence of D-glucose, with dissociation constant close to the  $K_m$  value for this substrate (data not shown). Using the same strategy, we were unable to detect the formation of a TIGK-D-glucose complex. In the ternary complex,  $R_g$  further decreased to 24.0  $\text{\AA}$  (Table 2 and Figure 4B), indicating that the total conformational change was attained only after E $\cdot$ A $\cdot$ B complex formation. Similarly, fluorescent measurements following the Tryptophan signal from the TIGK shows an increase in intensity upon binding of glucose only in the presence of a ADP analog (AlF $_3$ -AMP) (data not shown). Thus, SAXS data suggest that changes in enzyme compactness recapitulated the sequential substrate binding events; binding of both substrates to the active site triggered a full domain closure involving a 5.0  $\text{\AA}$  reduction in  $R_g$  (Figure 4B and Table 2).

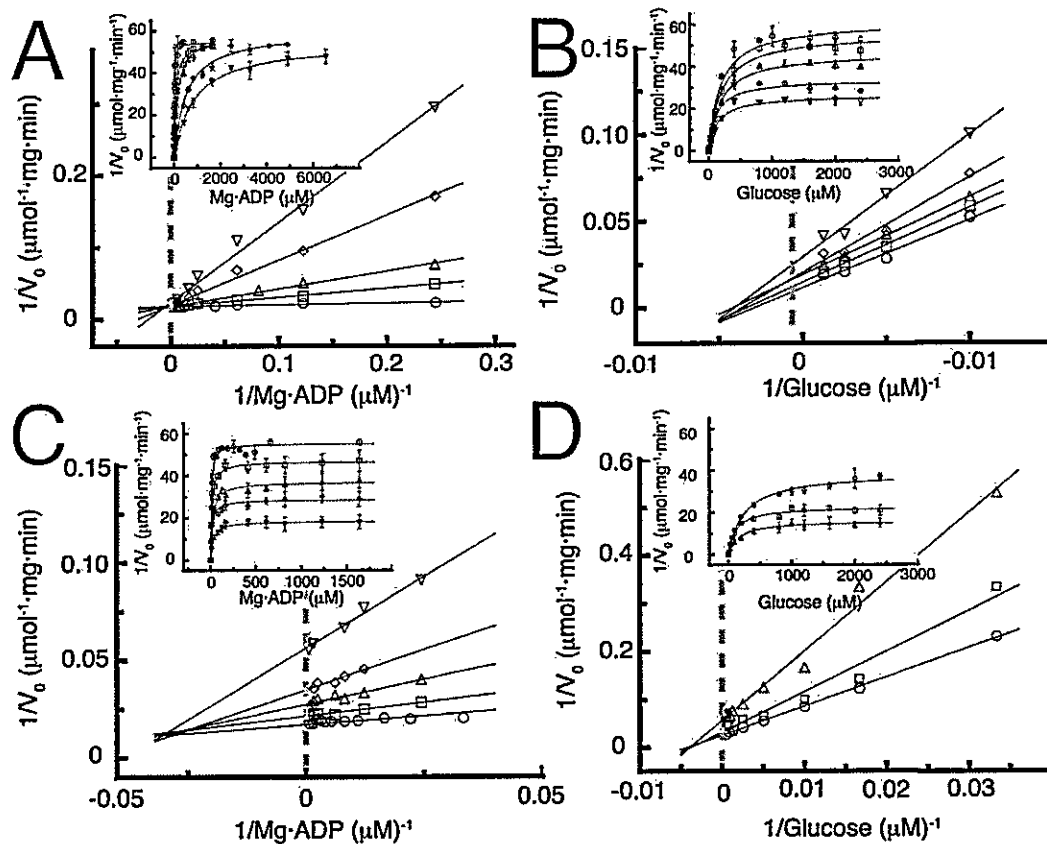
*Ab initio* models built using the  $P(r)$  curves for each condition suggest that the relative position of the mass centers of the two domains changes as a function of the ligands added to the solution. For example, in the model built for the apo and the Mg $\cdot$ ADP conditions, the active-site cleft between the small and large domains is respectively widely or moderately open (Figure 5A and 5B), whereas in the presence of Mg $\cdot$ ADP $\beta$ S-D-glucose, the small domain has moved toward the large domain to occupy a position that occludes the active site, suggesting that the enzyme has transitioned toward a fully closed state (Figure 5C).

**Table 1. Kinetic parameters for ADP-dependent glucokinase from *T. litoralis*.**

	$K_m$ $^2$ ( $\mu\text{M}$ )	$V_{max}$ $^2$ ( $\mu\text{mol}\cdot\text{mg}^{-1}\cdot\text{min}^{-1}$ )
MgADP	8.6	67.8
D-glucose	218.5	67.8

$^2$ Values obtained from fitting the data to the linear expression of bi-substrate ordered sequential model.

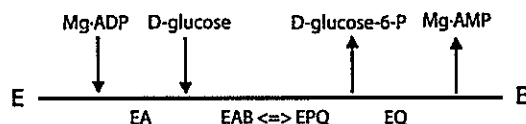
doi:10.1371/journal.pone.0066687.t001



**Figure 2. Product inhibition patterns for TIGK.** (A) Inhibition by Mg-AMP with Mg-ADP as variable substrate. Measurements were assayed at fixed Mg-AMP concentrations: 100  $\mu\text{M}$  ( $\square$ ), 200  $\mu\text{M}$  ( $\Delta$ ), 600  $\mu\text{M}$  ( $\diamond$ ) and 1000  $\mu\text{M}$  ( $\nabla$ ). Control curve in absence of product ( $\circ$ ) was also included in the graph. (B) Inhibition by Mg-AMP with D-glucose as variable substrate. Measurements were assayed at fixed concentrations of Mg-AMP: 435  $\mu\text{M}$  ( $\square$ ), 858  $\mu\text{M}$  ( $\Delta$ ), 1738  $\mu\text{M}$  ( $\diamond$ ) and 3657  $\mu\text{M}$  ( $\nabla$ ). Control curve in the absence of product ( $\circ$ ) was also included in the graph. (C) Product inhibition by glucose-6-P with Mg-ADP as variable substrate. Measurements were assayed at fixed glucose-6-P concentrations: 100  $\mu\text{M}$  ( $\square$ ), 500  $\mu\text{M}$  ( $\Delta$ ), 1000  $\mu\text{M}$  ( $\diamond$ ) and 2000  $\mu\text{M}$  ( $\nabla$ ). Control without the product was included ( $\circ$ ). (D) Product inhibition by D-glucose-6-P with D-glucose as variable substrate. Measurements were assayed at fixed glucose-6-P concentrations: 500  $\mu\text{M}$  ( $\square$ ) and 2000  $\mu\text{M}$  ( $\Delta$ ). Control curve without the presence of product was included ( $\circ$ ). Inset of all figures show the non-linear fit of the total data.  
doi:10.1371/journal.pone.0066687.g002

### Structure Determination of the apo and holo forms of TIGK

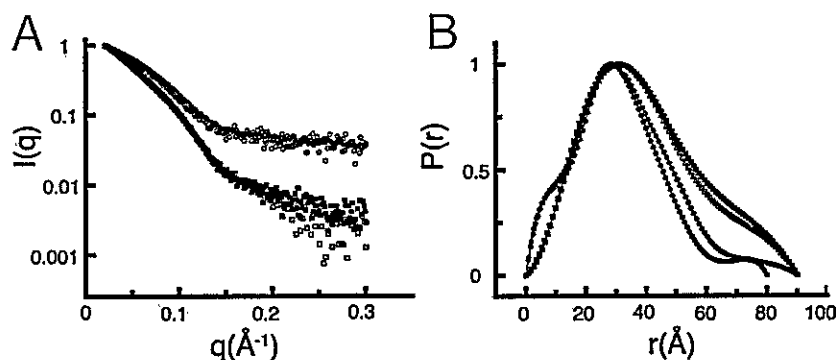
To gain insight into the structural basis of TIGK catalysis, binding of substrates and associated conformational changes, we sought to obtain high-resolution crystal structures of this enzyme under different conditions. We determined the crystal structures of



**Figure 3. Sequential ordered Bi-Bi mechanisms for the reaction catalyzed by TIGK.** Arrows indicate the entry and release of substrate and products.  
doi:10.1371/journal.pone.0066687.g003

TIGK in its apo form (absence of ligands) and holo form, in the presence of D-glucose and the nonhydrolyzable ADP analog ADP $\beta$ S (ternary complex, ADP $\beta$ S-D-glucose). Both structures were obtained by molecular replacement, and refined to 2.05 and 2.58  $\text{\AA}$  resolution, respectively. X-ray data collection and refinement statistics for each structure are summarized in Table 3.

The overall structure of TIGK in both states (apo-enzyme and ternary complex) shows the same two domains organization described previously [5]; the large domain harbors a Rossmann fold ( $\alpha/\beta/\alpha$ ) architecture with a central twelve stranded  $\beta$ -sheet surrounded by thirteen  $\alpha$ -helices and three  $3_{10}$  helices, and the small inserted domain consisting of five  $\beta$ -strands and four  $\alpha$ -helices. The active site is located in a cleft between the two domains. The insertion of the small domain between strand 2 and helix  $\beta$  of the large domain at the farthest end of the active site cavity provides a physical basis for ligand-mediated communication between the two domains (Figure 6).



**Figure 4. Scattering curves and pair distance distribution functions  $P(r)$ .** (A) Scattering patterns for the different conditions explored: apo-enzyme ( $\square$ ), enzyme-D-glucose ( $\blacksquare$ ), enzyme-Mg-ADP ( $\circ$ ) and in the presence of Mg-ADP $\beta$ S and glucose ( $\bullet$ ). The curves were normalized to unity at their maximum value for comparison purposes. (B)  $P(r)$  graphs for each condition calculated by Fourier transformation using GNOM (28). The graphs were normalized to unity at their maximum value for comparison purposes. doi:10.1371/journal.pone.0066687.g004

The spatial disposition of the two domains was different when comparing the structures of the apo-enzyme and the ternary complex (Figure 6). In the absence of ligands, the mass centers of the small and large domain were separated by 29.4 Å, but when Mg-ADP $\beta$ S and D-glucose were simultaneously bound to the active site, the two domains underwent conformational changes causing the inter-domain distance to shrink by nearly 3 Å, that to 26.5 Å (Table 2). In the structure of TIGK·Mg-ADP reported previously [5], the distance between domains is 28.1 Å. Therefore, after binding of the first substrate (Mg-ADP) the distance between domains shrank by only 1.3 Å, whereas in the ternary complex, the inter-domain distance diminishes by an additional 1.6 Å, to complete a 2.9 Å closure movement (Figure 6C).

The conformational changes that occur upon sequential substrate binding can be explained by an almost pure rotation (or a rotation plus a translation) facilitated by residues in the flexible inter-domain connection. Structural changes that accompany binding of the first substrate (Mg-ADP) include a small relative rotation (4.8°) between the two domains plus additional movements that do not completely occlude the active site cavity. Accordingly, TIGK·Mg-ADP structure has been previously described as an “open” structure [5]. In contrast, the crystal structure of the ternary complex TIGK·Mg-ADP $\beta$ S·D-glucose exhibits an 11.8° relative rotation with respect to the apo-TIGK

structure and corresponds to a fully closed conformation that entraps both substrates in a Michaelis-Menten-like complex. This series of rotations recapitulate the sequential binding of substrates: binding of Mg-ADP leading to a 4.8° rotation, the second substrate causing a 7.2° rotation (TIGK·Mg-ADP $\beta$ S·D-glucose versus TIGK·Mg-ADP). Thus, the total rotation to complete the active site closure is 11.7° (TIGK·Mg-ADP $\beta$ S·D-glucose versus apo TIGK) (Figure S1). All these translational and rotational rearrangements are accompanied by several charge distribution rearrangements at the inner space of the active site, which facilitate domain closure (Figure S2).

A full description of the interactions present at both binding sites (Mg-ADP and D-glucose) is illustrated in Supporting Information (Figure S3 and Figure S4).

Inspection of the TIGK·Mg-ADP $\beta$ S·D-glucose structure reveals five clusters of interactions formed by residues from the large and the small domain that could mediate the transition from the open to the closed conformation (see below). Three of these clusters involve attractive interactions governed by hydrogen (H) bonds, a fourth one is an electrostatic attraction involving a cation- $\pi$  interaction, and the fifth cluster, which consists of an intrinsically destabilizing interaction (an electrostatic repulsion), could play a regulatory role during domain opening by counteracting the net attractive interaction of the other four clusters (Figure 7).

Cluster 1 is formed by Glu188 (from the small domain) and residues Thr446 and Val447 (both from the large domain). In the open conformation, these residues are too far apart to interact, but when Mg-ADP $\beta$ S·D-glucose is bound to the active site, the greater proximity between these residues afforded by the conformational change facilitates the formation of two H-bonds between the Glu188 side chain and the hydroxyl group of Thr446 and two other H-bonds with the backbone amide groups of Thr446 and Val447 (Figure 7, cluster 1).

Cluster 2 involves Arg202 (small domain) and Tyr354 (large domain). In the ternary complex, the guanidinium group of Arg202, relocates on top of the Tyr354 phenol side chain (large domain), establishing a new  $\pi$ - $\pi$  cation interaction. The distance between the two side-chain groups is reduced from 12.0 Å in the apo-enzyme to only 3.4 Å in the ternary complex (Figure 7, cluster 2).

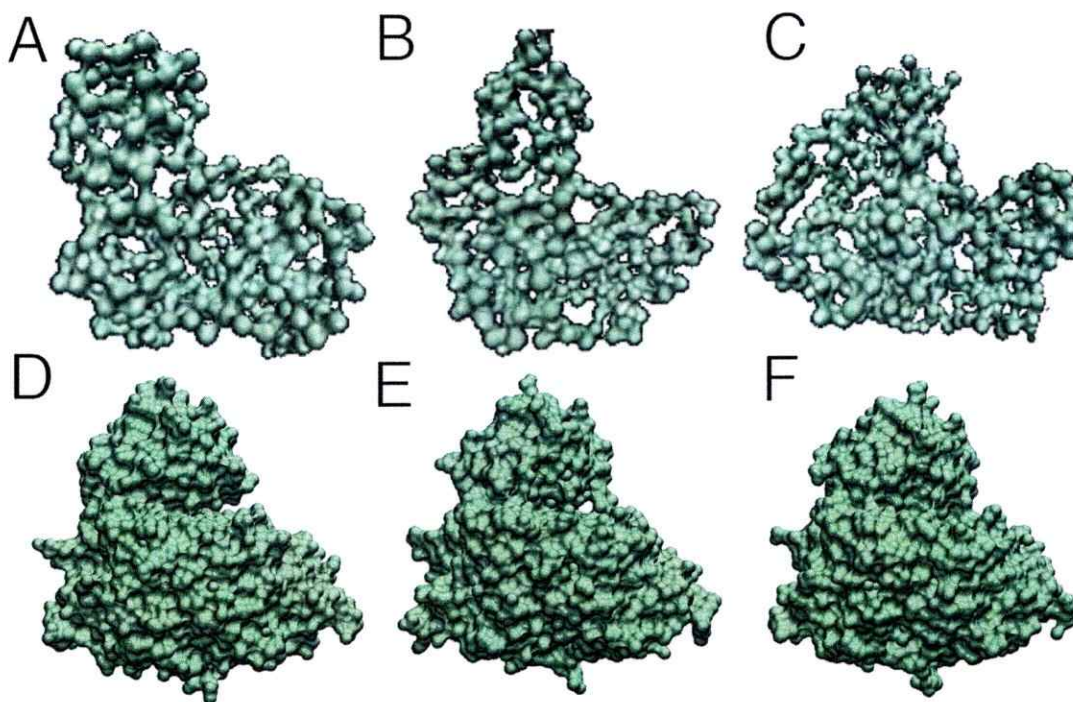
In Cluster 3, Arg117 from the small domain only can interact in the closed conformation with residues Gly386 and Ser445, both

**Table 2. TIGK radius of gyration ( $R_g$ ) and distance between the small and large domains under different conditions.**

	Apo-enzyme	D-Glucose	Mg-ADP	Mg-ADP $\beta$ S-Glc
$R_g^s$ (Å)	29.0 $\pm$ 0.1	30.0 $\pm$ 0.1	25.9 $\pm$ 0.1	24.0 $\pm$ 0.1
$R_g^t$ (Å)	24.1	N.A.	23.7	23.1
$D^1$ (Å)	29.4	N.A.	28.1	26.5

$R_g^s$  radius of gyration calculated using probability density function graphs.  $R_g^t$  theoretical radius of gyration calculated with the software CRYSOLOG [48], using crystallographic data.  $D^1$  distance between the small and large domain calculate from crystal structures. For the apo-enzyme and the ternary ADP $\beta$ S·D-glucose complex, we used the PDB structures solved in this work, whereas for the ADP complex the coordinates available from the PDB (1GCS) were used. N.A., not applicable.

doi:10.1371/journal.pone.0066687.t002



**Figure 5. Comparison between SAXS envelope models and crystals structures.** (A) Enzyme model obtained from the SAXS data in the absence of substrates; (B) in the presence of Mg-ADP and (C) in the presence of Mg-ADP and D-glucose. Every model was built using GASBOR with no symmetry constraints. (D), (E), (F) Surface representation of the enzyme's structures in the absence of substrates; in the presence of ADP (PDB 1GCS) and in the presence of Mg-ADPβ5 and D-glucose, respectively. doi:10.1371/journal.pone.0066687.g005

from the large domain. In the small domain, Tyr46 establishes an H bond with Glu115, which in turn accepts two H-bonds from Arg117. Therefore, the spatial location and polarity of Arg117 is highly influenced by the interactions that it established. On the other hand, Lys382 belonging to the large domain is located at 2.9 Å from Arg117 (in the ternary complex), forming three H-bonds with residues from the same domain: Gly386, Ser445, and Val447, which could influence the position and charge distribution of Lys382. Considering charge, distance and the microenvironment that surround Arg117 and Lys382 side chains, the net nature of this interaction is likely to be attractive (Figure 7, cluster 3).

Cluster 4 includes Ala441 and Ser442 from the large domain and Arg191 from the small domain. In the open conformation the side chain of Arg191 is too far away from Ala441 and Ser442 to interact, whereas in the closed conformation (ternary complex), Arg191 can donate two new H-bonds: one to the carbonyl group from Ala441 and another to the hydroxyl group from of Ser442 side chain (Figure 7, cluster 4).

Cluster 5 is defined by interactions between residues Lys74 and Lys246 located in the small and large domain, respectively. In the apo-enzyme the distance between amine groups is >8.7 Å, whereas in the ternary complex this distance decrease to 4.0 Å, close enough to establish a repulsive interaction between the side chains of these Lys residues. We hypothesize that this repulsive interaction that is formed only in the ternary complex could have a regulatory function by favoring domain opening as soon as the

reaction products abandon the active site (Figure 7, repulsive interaction).

#### Conserved Interactions Involved in the Open-closed Transition in the ADP-dependent Kinase Family

To evaluate whether the existence of clusters of residues involved in attractive and repulsive interactions between both domains corresponds to a conserved strategy among the ADP-dependent sugar kinase family, we analyzed the evolutionary profile of this family obtained by Bayesian inference of phylogeny the interacting residues are conserved. In the profile obtained (Figure 8A) all the groups represented in the tree correspond to the following archaeal orders: ADP-GKs from *Thermococcales* (blue), ADP-PFKs from *Thermococcales* (pink), ADP-PFKs from *Methanococcales* (purple), ADP-PFKs from *Methanosarcinales* (green), ADP-GKs from *Methanosarcinales* (cyan); except ADP-GK from Eukarya (gray) which was used as outgroup. Analysis of both the sequences employed to construct the tree and the equivalent structural positions of the interactions involved in the opening/closing domain revealed that some interactions are highly conserved (Figure 8B and 8C). For example, in cluster 1 residues Glu188, Thr446 and Val447 are fully conserved across the whole family. On the other hand, residues involved in cation- $\pi$  interactions (cluster 2) are conserved in all ADP-dependent glucokinases from hyperthermophilic organisms belonging to the *Thermococcales* group. Residues from cluster 3, which participate in the attractive interactions between Arg117 and Ser445/Gly386, are conserved

**Table 3. Crystallographic data statistics and refinement.**

	Apo-TIGK	Holo-TIGK (Mg-ADP/S-D-glucose)
<b>Data collection</b>		
Beamline	ID23-2	ID14-4
Wavelength (Å)	0.8726	0.9794
Space group	<i>P</i> <sub>3</sub> ,21	<i>P</i> <sub>3</sub> ,21
<b>Cell dimensions</b>		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	109.1, 109.1, 129.6	106.5, 106.5, 130.1
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 120	90, 90, 120
Resolution (Å)	40.0–2.05 (2.10–2.05)*	35.0–2.58 (2.65–2.58)*
<i>R</i> <sub>merge</sub>	0.089 (0.564)	0.093 (0.672)
<i>I</i> / <i>σ</i> <i>I</i>	19.4 (3.7)	18.3 (3.5)
Total reflections	687,053 (72,183)	234,319 (33,911)
Unique reflections	56,090 (7710)	27,361 (1992)
Completeness	94.7 (99.2)	99.9 (100.0)
Redundancy	9.4 (12.2)	7.4 (7.4)
<b>Refinement</b>		
Resolution (Å)	39.3–2.05 (2.10–2.05)	34.9–2.58 (2.65–2.58)
No. reflections	55,935 (3529)	25,931 (1402)
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.17/0.21	0.17/0.22
<b>No. atoms</b>		
Proteins	3854	3810
ADP/PS	–	23
D-glucose	–	13
Water	293	163
<b>B-factors (Å<sup>2</sup>)</b>		
Protein	34.6	50.5
AMP	–	50.2
D-glucose	–	49.9
Water	44.8	47.6
<b>R.m.s. deviations</b>		
Bond lengths (Å)	0.023	0.009
Bond angles (°)	2.058	1.360
<b>Ramachandran plot</b>		
Favored	457(98.3%)	450(97.0%)
Allowed	7(1.5%)	11(2.6%)
Outliers	1(0.2%)	2(0.4%)

doi:10.1371/journal.pone.0066687.t003

in all GKs and PFKs from thermophilic archaea (*Thermococcales* and *Methanococcales*). Interestingly, in the context of their repulsive interaction, Arg117 and Lys382 are conserved in all the enzymes of the thermophilic organisms analyzed in our study. Residues of cluster 4 and 5 are only present in some sequences and therefore are less well conserved.

## Discussion

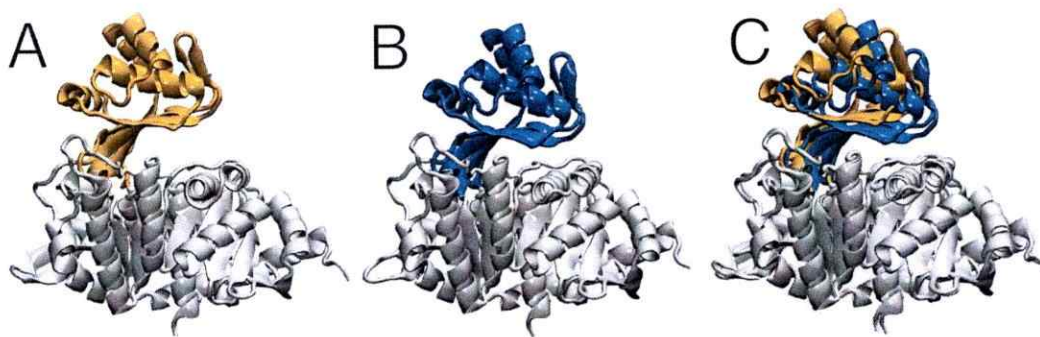
Understanding the mechanism of molecular recognition and in particular the characteristics of the response of a macromolecule to binding of a small molecule binding represents a challenging issue of modern structural biology. Our studies concerning the kinetic properties of ADP-GK in solution, along with the SAXS and X-

ray crystallographic data provide important new perspectives on the conformational changes and dynamics of TIGK during catalysis, which are also relevant for other ADP-dependent hyperthermophilic kinases. The results support the idea that the enzyme is conformationally flexible and sensitive to specific substrate binding events. When the enzyme is in ligand-free state (apo), the large and small domains are distant from each other and accordingly there are no discernible interactions between residues from either domain. In contrast, for the ternary TIGK·Mg·ADP·D-glucose complex to become properly assembled, the two domains of the holo enzyme are rearranged by a twisting and closing motion that leads to a fully closed, catalytically competent conformation.

Comparison of the molecular envelope calculated from the SAXS data with those obtained from the crystal structures allows us to identify differences between the relative positions of the small and large domains in different ligation states (Figure 5). For the apo-enzyme, the SAXS *ab initio* model suggests a more open conformation than the one found in the crystal. Differences are also found when comparing the ternary complex condition; in this case the SAXS envelope suggests a more closed conformation than the one observed in the crystalline structure. Thus, although the crystal structure indicates that the enzyme undergoes domain closure (Figure 6C), the SAXS results suggest that this conformational change is even more dramatic in solution; the apo form is more open and the holo form is more closed than the ones obtained by X-ray diffraction (Figure 5A, 5C and Table 2). In fact, the difference between the theoretical *R*<sub>g</sub> calculated for the crystal structures of the apo and holo enzymes is only 1 Å, whereas the *R*<sub>g</sub> calculated from the *P*(*r*) indicate a domain closure of 5 Å (Table 2). This situation could reflect the larger conformational freedom of the enzyme in solution compared to the crystalline state. In spite of this, the most striking structural aspect about nucleotide and D-glucose binding to the active site is that both binding events indicate a sequential conformational change, in complete agreement with the orderly entry of substrates determined from the kinetic mechanism. Mg-ADP triggers the first structural change (semi-closed conformation), which favors the oncoming of the small domain toward the large domain, followed by the entry of D-glucose, which in turn leads to ternary complex formation and total domain closure. This sequential conformational change is a strong suggestion for an induced-fit mechanism [8]. In our case, movement toward the catalytically active conformation results in the occlusion of the active site. It has been proposed that this mechanism would favor catalysis by precluding solvent from competing with the catalyzed reaction but domain closure could also impair catalysis by restricting binding of the second substrate to this enzyme form. The latter possibility raises the question as to whether such a situation is even compatible with a conformational selection model since, even if the enzyme samples this semi-closed state in solution, steric occlusion of the active-site cleft might impede substrate binding. Moreover, Sullivan and Holyoak [18] have proposed that enzymes with lid-gated active sites must operate by an induced fit mechanism instead of conformational selection, and Okazaki and Takada [12] reported that stronger and long-range interactions favor induced fit, whereas shorter-range interactions favor conformational selection. However, although the preceding evidence supports an induced fit mechanism, it has been stressed that both mechanisms are likely to play important roles in molecular recognition and that these models are not mutually exclusive.

For example, even if initial binding were achieved by a conformational selection mechanism, it is likely that further changes in protein structure and energy landscape must occur in





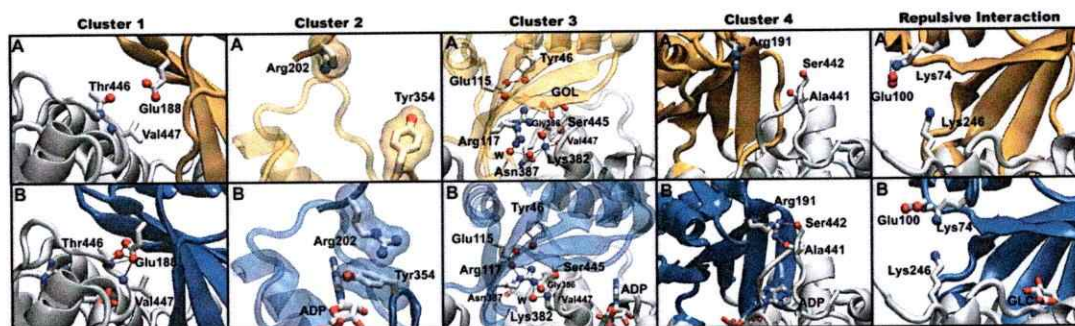
**Figure 6. Structures of the *T. litoralis* glucokinase in the apo form and in the Mg-ADP $\beta$ S-D-glucose ternary complex.** (A) Ribbon representation of the enzyme in the absence of substrate. The large domain is shown in white, whereas the small domain is shown in yellow. (B) Ribbon representation of the enzyme in the presence of ADP $\beta$ S and D-glucose. The large domain is shown in white and the small domain is shown in blue. (C) Structural alignment between the apo and holo (ADP $\beta$ S-D-glucose) forms. doi:10.1371/journal.pone.0066687.g006

order to achieve optimal intermolecular interactions, which undoubtedly constitutes an induced fit process. The co-existence of both mechanisms is reported to exist in maltose binding protein [19] whereas the induced fit mechanism has been proposed for other mesophilic enzymes belonging to the superfamily, such as ribokinase and adenosine kinase [20–22]. In the case of ribokinase, it has been suggested that domain closing is a prerequisite for nucleotide binding [20] whilst Ito *et al* [6] propose that access to the nucleotide binding site of TIGK is more difficult after domain closing and propose that ADP binds weakly to the open conformation with little or absent conformational change, as opposed to what was described in the present study.

Another important point to be considered is how the enzyme coordinates its structural elements to assume the closed conformation and how, once this conformation is stabilized. The

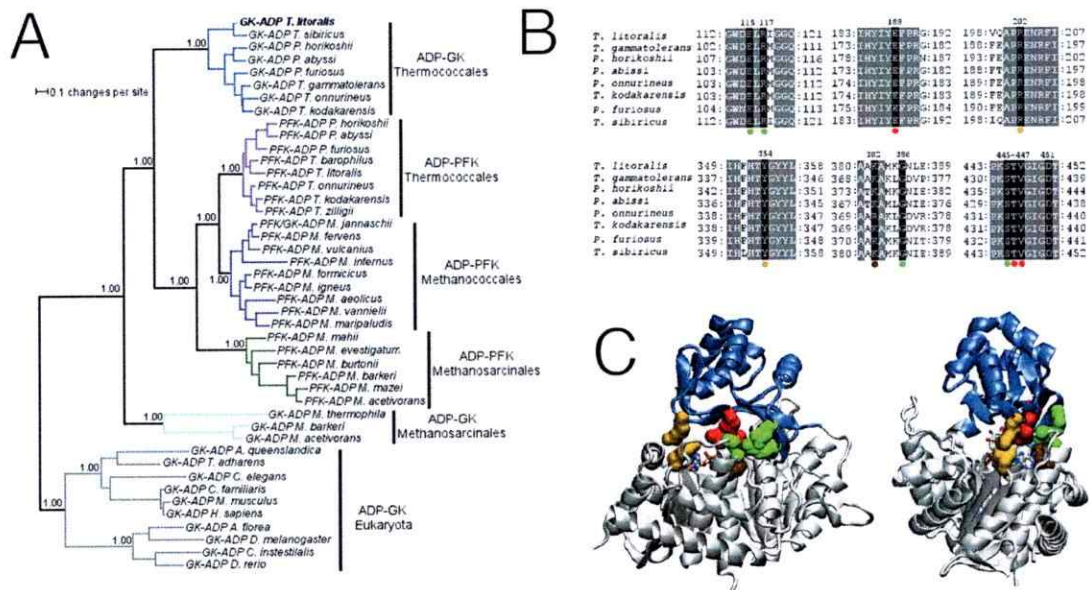
interaction between the large and small domain (via specific residue clusters), could represent a successfully strategy to achieved this goal. All the clusters described communicate both domains by non-covalent bonds and some of them are conserved in other enzyme homologs. Conservation analysis of cluster 1 shows that Glu188, from the small domain, and Thr446 and Val447, both in the large domain, are conserved across the whole family. This feature allows us to infer that net attractive interaction network between the GK large and small domains has been conserved during evolution in thermophilic and mesophilic enzymes of archaea as well as in the eukaryote domain, thereby implying that this interaction must already be present in the last common ancestor of the whole family.

On the other hand, the cation- $\pi$  interaction between the side chain of Arg202 and Tyr354 (cluster 2) is a very sophisticated



**Figure 7. Interactions clusters between the small and large domain that mediate the transition from the open to the closed conformation.** (A) Structure of the enzyme in the apo-form and (B) structure of the ternary TIGK-Mg-ADP $\beta$ S-D-glucose complex. The small domain was colored in yellow (apo-enzyme) or blue (ternary complex) for clarity purposes. Side chains of residues are represented as sticks. The atoms involved in H-bonds are represented as spheres. The nucleotide (ADP), the cosolvent, glycerol (GOL) and water (W) are also shown. Cluster 1 achieves communication between the small and large domain through Glu188, Thr446 and Val447. Cluster 2 residues contribute to stabilize the ADP-induced conformational change. In the apo-enzyme, Arg202 of the large domain is over the active-site pocket and the side chain of Tyr354 is oriented toward the outside of the pocket. In contrast, in the TIGK-Mg-ADP $\beta$ S-D-glucose ternary complex Arg202 and Tyr354 form a stacking interaction afforded by the rotation of their side-chains so as to allow a cation- $\pi$  interaction. Cluster 3's interaction network is rearranged upon ADP binding through the formation of H-bonds between the large and small domain that generates a net attraction. In cluster 4, Arg191 from the small domain H-bonds Ala441 and Ser442 from the large domain. The only net repulsive interaction is provided by cluster 5, involving Lys74 and Lys246; in the apo-enzyme Glu100 is H-bonded to Lys74 thus neutralizing its net charge, but in the ternary complex this stabilizing interaction is absent and Lys74 comes closer to Lys246 despite the associated unfavorable energy barrier. doi:10.1371/journal.pone.0066687.g007





**Figure 8. Conservation of cluster residues involved in domain opening/closure across the ADP-dependent sugar kinases family. (A)** Consensus phylogenetic tree for ADP-dependent sugar kinase family determined by Bayesian inference. Groups of enzymes from archaea are shown in color: GK from *Thermococcales* (blue), PFK from *Thermococcales* (pink), PFK from *Methanococcales* (purple) PFK from *Methanosarcinales* (dark green) and GK from *Methanosarcinales* (cyan). The group of enzymes from eukaryotic organisms that were used as an outgroup to establish the tree root is shown in gray. The posterior probability of some interesting groups is shown in its respective node. **(B)** Multiple sequence alignment of glucokinases from the *Thermococcales* group. Residues involved in clusters described in the texts are indicated by dots; cluster 1, red (Glu188-Thr446/Val447); cluster 2, yellow (Arg202-Tyr354); cluster 3, green (Arg117/Glu115-Gly386/Ser445) and brown (Lys382-Arg117). **(C)** Amino acids involved in clusters 1–3 are shown as spheres and colored as in B.  
doi:10.1371/journal.pone.0066687.g008

strategy dedicated to stabilize the closed conformation (Figure 7). Noticeably, as in cluster 1, these interactions are not exclusive of the TIGK enzyme, rather they are a common feature present in all the hyperthermophilic ADP-dependent GKs from the *Thermococcales* group while absent in mesophilic glucokinases (Figure 8). This class of interaction was studied by Minoux and Chipot [23], using a molecular dynamics approach in 1718 proteins. Their results show that  $\pi$ -system interactions (cation- $\pi$  and  $\pi$ -stacking interaction) are usually established between arginine and tyrosine residues. Moreover, the distances between the interacting  $\pi$ -systems calculated *in silico* by these authors (3.6 Å) are very similar to the ones seen in the crystal structure of the ternary complex of the TIGK (3.4Å) (Figure 7 cluster 2). In another study, Gromiha *et al.* [24] suggested that in thermophilic proteins the  $\pi$ -system interaction frequently employs Tyr residues, instead of other aromatic residues such as Phe or Trp.

In the last few years there is an increasing number of reports that highlight the relevance of cation- $\pi$  interactions in biological systems. These interactions have been demonstrated to have a substantial impact on protein structure as well as in catalysis and organic synthesis. For example, using a multidisciplinary approach Pecsai *et al.* [25] showed that elimination of the aromatic interaction had a little effect on substrate binding but a major impact in the catalytic efficiency of dUTPase. Based on the similar position of these aromatic interactions in various nucleotide-hydrolyzing enzymes, the authors proposed that this kind of interactions could be a general component of the enzymatic catalysis of phosphate ester hydrolysis. The biological relevance of cation- $\pi$  interactions

are illustrated, among others examples, by the critical role that they play in nicotine addiction, their prevalence at protein-protein interfaces, and their contribution to the understanding of thermo or psychro stabilizing interactions [26,27].

In the case of the cluster 3, residues involved in the attractive interaction between Arg117 from the small domain and residues Ser445 and Gly386 from the large domain (Figure 7) are conserved in all the thermophilic PFKs and GKs, but not in the mesophilic enzymes. The same scenario is seen with the repulsive interaction present in the same cluster, formed by residues Arg117 and Lys382, where the interaction is conserved in all thermophilic proteins of the family but not in the mesophilic members.

Finally, cluster 5 involves repulsive interactions between two lysine residues; none of them are conserved in any homolog considered in our analysis. In particular, this cluster seems to be exclusive of TIGK. Theoretical prediction of the protonation state of these lysine residues, using the PDB2PQR server implemented for electrostatics calculations [28], indicate that both residues should be positively charged at the pH employed for catalysis and therefore repel one another, facilitating domain opening. Interactions of this nature were deeply studied by Vondráček *et al.* [29] who, using an *in silico* approach, analyzed the lysine-lysine and arginine-arginine ion pair. The authors compared their results with analysis of structural databases, indicating that there is a strong repulsive effect between the ammonium groups ( $\text{NH}_4^+$ ) of lysine. This effect is due to charged ammonium ion and its almost spherical geometry, which favors repulsive Coulomb potential.

In summary, the closed conformation could be stabilized by at least four sets of attractive interactions between the small and large domain: one cation- $\pi$  (cluster 2) and three H-bonds clusters (clusters 1, 3 and 4), which in total form 7 formally attractive interactions (considering only two H-bonds for Glu188). Cluster 5 contributes one repulsive interaction between the large and small domain and could therefore facilitate domain opening and product release. In total, 6 H-bonds and 1 charge-charge interaction support domain closing whereas 1 electrostatic repulsive interaction favors the opening. Preliminary molecular dynamics studies used to evaluate the behavior of each cluster commented above, confirmed the attractive nature for the cluster 1, 2 and 4 and the repulsive character of clusters 3 and 5.

A similar scenario has been described for phosphoglycerate kinase [30]. This enzyme would work as a spring, whereby the entry of the substrate favors domain closing which, in turn, results in the exposure of a hydrophobic patch that quickly triggers domain opening. Also, an H-bond network that stabilizes domain closure has been described for this enzyme [31].

Here we present an integrated approach that determines the conformational changes experienced by a hyperthermophilic enzyme (TIGK) in solution and in the crystalline state and their correlation with the kinetic mechanism. The results support an induced-fit mechanism although the participation of a conformational selection mechanism, at least in some reaction steps, cannot be ruled out. Also, we demonstrate the presence of conserved cluster of residues that could be involved in the stabilization of the closed conformation, as well as another one that could account for domain opening. Some of these interactions are fully conserved in the whole family of ADP-dependent kinases, hence constituting the basis for a general mechanism involved in catalysis, which may prove general for many other hyperthermophilic enzymes.

## Materials and Methods

### Purification of TIGK

The enzyme was purified as described by Merino *et al.*, [32]. Briefly, the cells from 1L of LB broth were harvested by centrifugation and disrupted by sonication. Enzyme purification consisted basically in a thermal shock followed by  $(\text{NH}_4)_2\text{SO}_4$  precipitation and two-chromatographic steps (hydrophobic and ionic exchange chromatography). Every step during purification was performed in 50 mM Tris-HCl pH 7.8, and 5 mM  $\text{MgCl}_2$ . Prior to any kinetic or SAXS measurement, the buffer was exchanged to 50 mM Hepes-NaOH pH 7.8 and 5 mM  $\text{MgCl}_2$ .

### Kinetic Experiments

Enzyme activity was measured spectrophotometrically by following  $\text{NAD}^+$  reduction at 340 nm coupled with D-glucose-6-phosphate oxidation at 40°C. Standard assay was carried out in a final volume of 0.7 mL in the presence of 25 mM Hepes pH 7.8, 0.5 mM  $\text{NAD}^+$ , 2–5 U of D-glucose-6-P dehydrogenase and the indicated concentrations of substrates (ADP and glucose). In all initial velocity studies performed, the concentration of free  $\text{Mg}^{2+}$  was kept constant.

The same spectrophotometric method was used to determine product inhibition kinetics with  $\text{Mg}\cdot\text{AMP}$ , whereas for glucose-6-P a discontinuous assay was employed whereby the disappearance of  $\text{Mg}\cdot\text{ADP}$  is quantified by the formation of pyruvate and its subsequent oxidation to lactate. The reaction mixture for the discontinuous assay contains 0.2 mM NADH, 100 mM KCl, 4 U of pyruvate kinase and 4 U of lactate dehydrogenase, in 1 mL of reaction.

### Kinetic Data Analysis

The data from initial velocity experiments were analyzed using non-linear hyperbolic plot and double reciprocal plots, both considering a kinetic model of a Bi-Bi ordered sequential mechanism (Equation 1). The kinetics constants were obtained from secondary plots of the slopes and intercepts of double reciprocal plots. For product inhibition experiments, the inhibition pattern and inhibition constant were analyzed by plotting the slopes and intercepts of the primary double reciprocal plots against product concentration. The Marquardt-Levenberg algorithm was used for the non-linear and linear adjustment of the kinetic parameters. The weighting system used for our fit was r2, giving equal weight to each value included in the kinetic measurements. Kinetic constants calculated by global fit using the program Visual Enzymics (SoftZymics) (2010) and by linear regression of the reciprocal plots, were essentially the same.

The nomenclature used in the initial velocity expression is according to Cleland [33].

Equation 1

$$\frac{v}{V_{\max}} = \frac{[A][B]}{K_{ia}K_{mb} + K_{mb}[A] + K_{ma}[B] + [A][B]}$$

### SAXS Measurements and Data Processing

All scattering data were collected in the beamline D11A-SAXS1 [34], in the National Synchrotron Light Laboratory (Campinas, SP, Brazil). A wavelength  $\lambda$  1.488 Å and a sample-to-detector distance of 1000 mm were used. The scattering data were recorded using the photon-counting Pilatus detector. The magnitude of the scattering vector, defined as  $q = 4\pi\sin\theta/\lambda$  (where  $2\theta$  is the scattering angle) was ranged  $0.001 \text{ nm}^{-1} < q < 0.3 \text{ Å}^{-1}$ . For every experimental condition the temperature was maintained at 40°C and samples were irradiated for 3 or 5 min, collecting several spectra to monitor radiation damage and beam stability. The buffer scattering was subtracted for each condition and the resulting curve was analyzed. For each condition employed, several protein concentrations were assayed in order to ascertain the influence of protein aggregation in the scattering of the sample. TIGK at  $1\text{--}6 \text{ mg}\cdot\text{mL}^{-1}$  were used, observing significant changes at low angle only when protein concentration exceeded  $5 \text{ mg}\cdot\text{mL}^{-1}$ ; hence further analyses were restricted to the scattering data from protein samples at  $4 \text{ mg}\cdot\text{mL}^{-1}$ . Radii of gyration were measured using the Guinier approximation considering angles where  $q < 1.3/R_g$ . We also evaluated the  $R_g$  by the pair distance distribution  $P(r)$  calculated with the software GNOM [35]. *Ab initio* models of the enzyme were built using the software GASBOR [36].

### Crystallization, Data Collection and Structure Solution

TIGK was concentrated to  $15 \text{ mg}\cdot\text{mL}^{-1}$  in storage buffer. Sitting-drop vapor-diffusion crystallization experiments were set up by mixing 1  $\mu\text{L}$  protein solution and 1  $\mu\text{L}$  crystallization condition over 75  $\mu\text{L}$  mother liquor. Crystals of apo-TIGK were obtained in 14% (w/v) polyethylene glycol (PEG) 6000, 0.2 M  $\text{LiSO}_4$ , 0.1 M sodium citrate pH 3.6 and 5 mM dithiothreitol (DTT). Crystals of TIGK-Mg-ADP $\beta\text{S}$ -D-glucose were obtained in 1.5 M sodium citrate pH 5.2, 5 mM DTT, 5 mM  $\text{MgCl}_2$ , 6 mM ADP $\beta\text{S}$  and 30 mM D-glucose. Both apo-TIGK and TIGK-Mg-ADP $\beta\text{S}$ -D-glucose crystals were cryo-protected by direct addition of 20% (v/v) sterile glycerol and flash-frozen in liquid nitrogen.



Complete and redundant X-ray diffraction data sets were collected at the ID23-2 (apo-TIGK, to a resolution of 2.05 Å) and ID14-4 (TIGK·Mg·ADPβS·D-glucose-, to a resolution of 2.58 Å) beamlines at the European Synchrotron Radiation Facility (ESRF, Grenoble, France). Data were indexed and integrated with XDS [37] and scaled and merged with SCALA [38] from the CCP4 software package [39]. Crystallographic data processing and merging statistics are summarized in Table 3. The structures of TIGK·Mg·ADPβS·D-glucose and apo-TIGK were phased by molecular replacement with PHASER 2.5 [40] using the semi-open structure of TIGK·Mg·ADP determined previously (PDB 1GC5) [5]. Minimally biased maps of the molecular replacement solutions calculated before refinement showed conspicuous density differences between the search model and the structure of TIGK·Mg·ADPβS·D-glucose consistent with the presence of bound substrates and accompanied by a significant inter-domain movement. ADPβS was modeled as AMP since electron density for the β-phosphate was missing in the electron density maps. The structures were iteratively built using Coot [41] and refined with REFMAC5 [42]. Upon convergence, the  $R_{work}/R_{free}$  for the apo-TIGK and TIGK·Mg·ADPβS·D-glucose structures were 0.17/0.21 and 0.17/0.22, respectively, and both had good geometry and stereochemistry as assessed by MolProbity [43]. Refinement and validation statistics are summarized in Table 3. All figures of protein structures were prepared with VMD [44].

#### Accession Numbers

The X-ray crystal structures have been deposited with the Protein Data Bank with accession codes 4B8R (apo-TIGK) and 4B8S (TIGK·Mg·ADPβS·D-glucose).

#### Alignment and Bayesian Inference of Phylogeny

The protein sequences were extracted using the protein BLAST Server from non-redundant protein sequence database (nr) using the PSI-BLAST algorithm with 5 iterations, using ADP-GK from *T. litoralis*, ADP-PFK from *P. horokostii* and ADP-GK from *H. sapiens* as templates. The multiple sequence alignment (MSA) was constructed based on three-dimensional and secondary structure constraints using Promals3D [45], and then misaligned positions were corrected manually. To build an evolutionary profile of the family, redundant sequences in eukaryotes were removed using

QR Sequence tool of Multiseq in VMD [46] with a PID 40%. Finally gaps positions were removed using Seqverter and the MSA was prepared for Bayesian inference of phylogeny with Mr. Bayes version 3.1.2 [47]. For the analysis we used Cprev as the fixed rate model, which showed a posterior probability of 1.0 in the mixed model and gamma-shaped rate variation across site with a proportion of invariable site. The number of generations was set to  $1 \times 10^6$ , samplefreq to 100, with 2 run with 4 chains per run and temperature parameter set to 0.2. The average standard deviation of the split frequencies was less than 0.01 upon convergence. The consensus tree was calculated from 15,002 samples trees.

#### Supporting Information

**Figure S1 Structural alignment of TIGK structures.**  
(DOCX)

**Figure S2 Representation of Molecular Electrostatic Potentials.**  
(DOCX)

**Figure S3 TIGK substrate contacts.**  
(DOCX)

**Figure S4 Nucleotide and sugar binding sites.**  
(DOCX)

**Table S1 Product inhibition parameters.**  
(DOCX)

#### Acknowledgments

The authors thank the European Synchrotron Radiation Facility (ESRF), Grenoble, France, for access to the ID14-4 and ID23-2 beamlines and SAXS1 and SAXS2 beamlines from LNLS, Sao Paulo, Brazil. We also thank Dr. Takayoshi Wakagi (University of Tokyo) from whom the expression vector containing ADP-dependent glucokinase was obtained.

#### Author Contributions

Conceived and designed the experiments: JAR-P VG. Performed the experiments: JAR-P AH-M VC-F. Analyzed the data: JAR-P AHM VC-F FJF MCV VG. Contributed reagents/materials/analysis tools: JAR-P MCV VG. Wrote the paper: JAR-P VG.

#### References

- Sakuraba H, Ohshima T (2002) Novel energy metabolism in anaerobic hyperthermophilic archaea: a modified Embden-Meyerhof pathway. *J. Biosci Bioeng* 93: 441–448.
- Sakuraba H, Goda S, Ohshima T (2004) Unique sugar metabolism and novel enzymes of hyperthermophilic archaea. *Chem Rec* 3: 281–287.
- Merino F, Guixé V (2003) The ADP-dependent sugar kinase family: kinetic and evolution aspects. *IUBMB Life* 61: 753–761.
- Zhang CY, Dougherty M, Ealick SE (2004) Crystal Structure of an aminimidazole riboside kinase from *Salmonella enterica*: implications for the evolution of the ribokinase superfamily. *Structure* 12: 1809–1821.
- Ito S, Fushinobu S, Yoshioka I, Koga S, Matsuzawa H, et al. (2001) Structural basis for the ADP-specificity of a novel glucokinase from a hyperthermophilic archaeon. *Structure* 9: 205–214.
- Tsuge H, Sakuraba H, Koba T, Kujime A, Katunuma N, et al. (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: 2456–2463.
- Ito S, Fushinobu S, Jeong JJ, Yoshida I, Koga S, et al. (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* 33: 871–883.
- Koshland DE (1958) Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc Nat Acad Sci (USA)* 44: 98–104.
- Berger C, Weber-Bornhauser S, Eggenberger J, Hanes J, Pluckhun A, et al. (1999) Antigen recognition by conformational selection. *FEBS Letters* 450: 149–153.
- Kumar S, Ma B, Tsai CJ, Sinha N, Nussinov R (2000) Folding and binding cascades: Dynamic landscapes and population shifts. *Protein Sci* 9: 10–19.
- Vertessy B, Oroz F (2010) From “fluctuation fit” to “conformational selection”: Evolution, rediscovery, and integration of a concept. *Bioessays* 33: 30–34.
- Okazaki K, Takada S (2008) Dynamic energy landscape view of coupled binding and protein conformational change: Induced-fit versus population-shift mechanisms. *Proc Natl Acad Sci (USA)* 105: 11182–11187.
- Tobi D, Bahar I (2005) Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc Natl Acad Sci (USA)* 102: 18908–18913.
- Vogt AD, Di Cera E (2012) Conformational selection or induced fit? A critical appraisal of the kinetic mechanism. *Biochemistry* 51: 5894–5902.
- Cleland W (1963) The kinetics of enzyme catalyzed reactions with two or more substrate or products, I, II and III. *Biochim. Biophys. Acta* 67: 104–137, 173–187, 188–196.
- Cornish-Bowden A (2011) *Fundamentals of enzyme kinetics* Wiley-Blackwell ed. Fourth edition.
- Rice LM, Montalosa EA, Agard DA (2008) The lattice as allosteric effector: structural studies of αβ- and γ-tubulin clarify the role of GTP in microtubule assembly. *Proc Natl Aca. Sci (USA)* 105: 5378–5383.
- Sullivan SM, Holyoak T (2008) Enzymes with lid-gated active sites must operate by an induced fit mechanism instead of conformational selection. *Proc Natl Acad Sci (USA)* 105: 13829–13834.
- Tang C, Schwieters CD, Clore GM (2007) Open-to-closed transition in apo maltose binding protein observed by paramagnetic NMR. *Nature* 449: 1078–1082.

20. Sigrell JA, Cameron AD, Mowbray SL (1999) Induced fit on sugar binding ribokinase. *J Mol Biol* 290: 1009–1018.
21. Reddy MC, Palaninathan SK, Shetty NS, Owen JL, Watson MD, et al. (2007) High resolution crystal structures of *Mycobacterium tuberculosis* adenosine kinase. *J Biol Chem* 282: 27334–27342.
22. Selumacher MA, Scott DM, Mathews LJ, Ealick SE, Roos DS, et al. (2000). Crystal structures of *Toxoplasma gondii* adenosine kinase reveal a novel catalytic mechanism and product binding. *J Mol Biol* 296: 549–567.
23. Minoux H, Chipot C (1999) Cation- $\pi$  interactions in proteins: can simple models provide an accurate description? *J Am Chem Soc* 121: 10366–10372.
24. Gromiha MM, Thomas S, Santhosh C (2002) Role of cation-p interactions to the stability of thermophilic proteins. *Prog Biochem Biotechnol* 32: 355–362.
25. Pecsí I, Leveles I, Harmat V, Vertessy BG, Toth J (2010) Aromatic stacking between nucleobase and enzyme promotes phosphate ester hydrolysis in dUTPase. *Nucleic Acids Research* 38: 7179–7186.
26. Dougherty DA (2012) The Cation- $\pi$  Interaction. *Accounts of Chemical Research*.
27. Mahadevi AS, Sasthy GN (2013) Cation- $\pi$  Interaction: Its Role and Relevance in Chemistry, Biology, and Material Science. *Chem. Rev.* 113: 2100–2138.
28. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* 32: 665–667.
29. Vondráček J, Mason PE, Heyda J, Collins KD, Jungwirth P (2009) The molecular origin of like arginine-arginine pairing in water. *J Phys Chem B* 113: 9041–9045.
30. Zerrad L, Merli A, Schröder GF, Varga A, Graczer E, et al. (2011) A spring-loaded release mechanism regulates domain movement and catalysis in phosphoglycerate kinase. *J Biol Chem* 286: 14040–14048.
31. Varga A, Flachner B, Konarev P, Graczer E, Szabo J, et al. (2006) Substrate-induced double sided H-bond network as a means of domain closure in 3-phosphoglycerate kinase. *FEBS Letters* 580: 2698–2706.
32. Merino F, Rivas-Pardo JA, Caniguir A, García I, Guixé V (2012) Catalytic and regulatory roles of divalent metal cations on the phosphoryl-transfer mechanism of ADP-dependent sugar kinases from hyperthermophilic archaea. *Biochimie* 94: 516–524.
33. Cleland WW (1970). Steady State Kinetics. In: Boyer, PD, editor. *The Enzymes* vol II, third edition. Academic Press, New York. 1–61.
34. Kellermann G, Vicentia F, Tamura E, Rocha M, Tolentino H, et al. (1997) The small-angle X-ray scattering beamline of the Brazilian synchrotron light laboratory. *J Appl Cryst* 30: 880–883.
35. Svergun DI (1992) Determination of regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Cryst* 25: 495–503.
36. Svergun DI, Petoukhov MV, Koch MHJ (2001) Determination of domain structure of protein from X-ray solution scattering. *Biophys J* 80: 2946–2953.
37. Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* 66: 125–132.
38. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62: 72–82.
39. Winn MD (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67: 235–242.
40. McCoy J, Grosse-Kunstleve RW, Adams PD, Winn MD, Strohriani L, et al. (2007) Phaser crystallographic software. *J App Cryst* 40: 658–674.
41. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60: 2126–2132.
42. Ahushudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, et al. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 67: 355–367.
43. Davis IW, et al (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35: W375–W383.
44. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14: 33–38.
45. Pei J, Kim BH, Grishin NV (2006) PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res* 36: 2295–2300.
46. Roberts E, Eargle J, Wright D, Luthy-Schulten Z (2006) MultiSeq: unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics* 7: 382.
47. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.
48. Svergun DI, Barberato C, Koch MHJ (1995) CRYSOLE - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J Appl Cryst* 28: 768–773.

### Supporting Information:

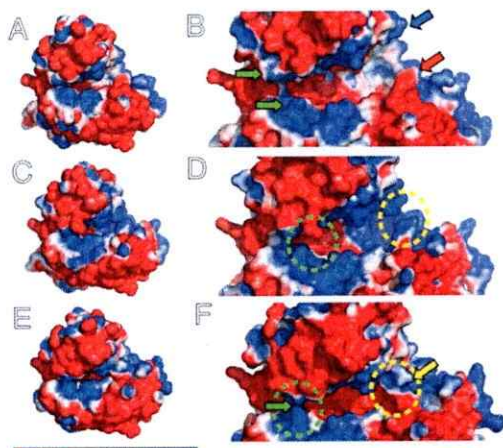
Rvas-Pardo JA, Herrera-Morande A, Castro-Fernandez V, Fernandez FJ, Vega MC and Guixé V. (2013) **Crystal Structure, SAXS and Kinetic Mechanism of Hyperthermophilic ADP-Dependent Glucokinase from *Thermococcus litoralis* Reveal a Conserved Mechanism for Catalysis.** PLoS ONE 8(6): e66687.



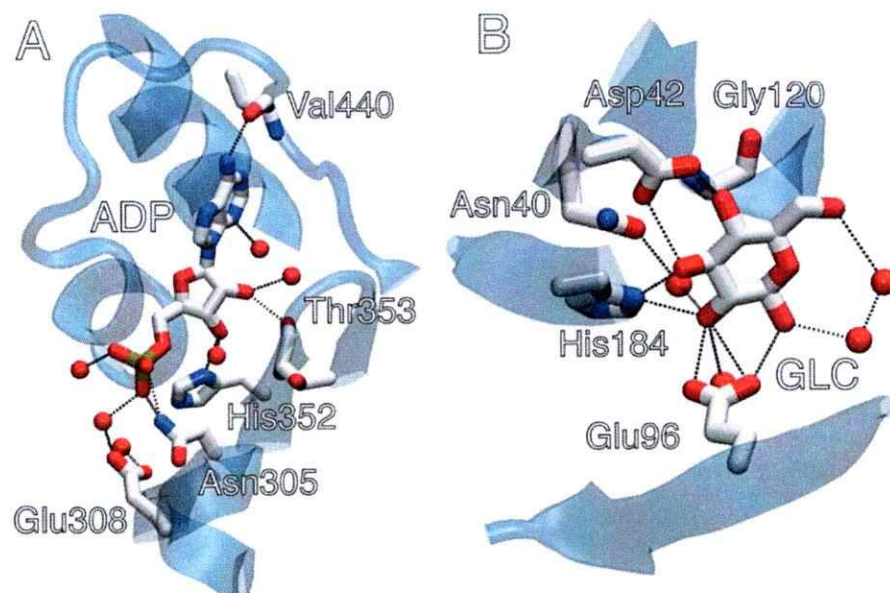
**Figure S1. Structural alignment of TIGK structures.** Apo TIGK (yellow), TIGK·Mg·ADP described by Ito et al. 2001 (grey) and TIGK·Mg·ADPβS·D-glucose (blue). Ito S, Fushinobu S, Yoshioka I, Koga S, Matsuzawa H, Wakagi T (2001) Structural basis for the ADP-specificity of a novel glucokinase from a hyperthermophilic archaeon. Structure 9: 205-214.

#### Electrostatic Analysis of the enzyme TIGK

Electrostatic analysis using the APBS tools available in PyMOL (DeLano, WL 2002) of the crystal structures shows a large rearrangement of the charge distribution at the inner space of the active site, especially after ADP binding (Figure S1A and C). The electrostatic map in the apo-enzyme shows a relatively neutral character that is altered after nucleotide binding, eliciting a change to a positive character. In the structure with ADPβS·D-glucose the active site is totally occluded by the small domain and the solvent accessible surface area leaves only a small space between both domains, thereby preventing access or release of further molecules (Figure S1A versus E, and B versus F).

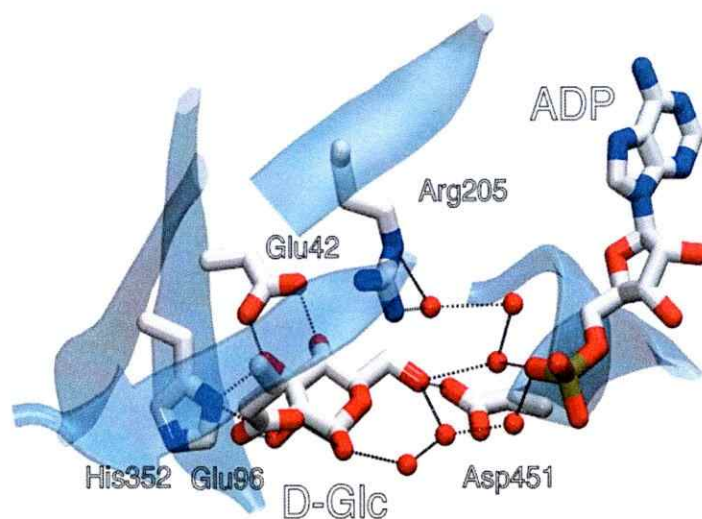


**Figure S2. Representation of Molecular Electrostatic Potentials distribution of TIGK in the open to closed transition.** (A) (C) and (E) Electrostatic potential representation of the solvent accessible surface in the apo-enzyme, TIGK·Mg·ADP and TIGK·Mg·ADPβS·D-glucose, respectively. (B) (D) and (F) Surface representation of the enzyme using the same electrostatic map, under different conditions: apo-enzyme, TIGK·Mg·ADP and TIGK·Mg·ADPβS·D-glc, respectively. Bar scale ranges from +1 Kcal/mol to -1Kcal/mol, using the blue color for positive charge, white for neutral, and red for negative charge. The arrows indicate the position of the Arg202 (blue arrow), Tyr354 (red arrow) and Lys74 and Lys246 (green arrows). DeLano, WL (2002) The PyMOL Molecular Graphics System, DeLano Scientific LLC, San Carlos, CA.



**Figure S3. TIGK substrate contacts.** (A) View of the ADP binding pocket in the TIGK ADPβS·D-glucose ternary complex. Val440 established H bonds with the NH6 of the ADP, while residues Thr353 and His352 established this kind of interaction with positions 2' and 3' of the ribose moiety, respectively. Asn305 makes contact with the oxygen from the α-phosphate through two H-bonds. Five water molecules establish H bonds with the nucleotide, one in the N3 from the adenine moiety, two with the ribose in the position 2' and 3' and two more with the oxygen from the α-phosphate. In the second sphere, Glu308 of the conserved NXXE motif accepts H bonds from water molecules (shown as red spheres). (B) View of the glucose binding pocket in the TIGK ADPβS·D-glucose complex. Asp42 establish two H-bonds with the 3' and 4' of the glucose, Glu96 can accept two H bonds from positions 1' and 2' of D-glucose, His184 is hydrogen bonded to positions 2' and 3' of glucose. Finally, the amine group from Gly120 is involved in an H bond with position 4' of the D-glucose. Furthermore, four water molecules interact with the sugar molecule in positions 1', 2' and 6'. Hydrogen bonds are represented by dashed lines. Side chains and adenosine and glucose atoms are in blue for nitrogen atoms, red for oxygen atoms, white for carbon atoms and orange for phosphorous atoms.





**Figure S4. Nucleotide and sugar binding sites in the ternary complex TIGK·Mg·ADP·D-glucose.** Residue Asp451, the likely catalytic base of the transfer reaction, is only at 2.52 Å from the 6'OH of the sugar. Arg205 can contribute with its positive charge to transition state stabilization, favoring the  $\beta$ -phosphate transfer from the nucleotide to the sugar. Water molecules involved in the reaction are shown as red spheres.

**Table S1. Product inhibition parameters for the TIGK reaction at 40 °C.**

Inhibitor	Variable substrate	Inhibition	$K_{iq}$ ( $\mu\text{M}$ )	$K_{ip}$ ( $\mu\text{M}$ )
Mg·AMP	Mg·ADP	C	28.0	-
Mg·AMP	D-Glucose	NC	72.2	-
Glucose-6-P	Mg·ADP	NC	-	-
Glucose-6-P	D-Glucose	NC	-	782.3

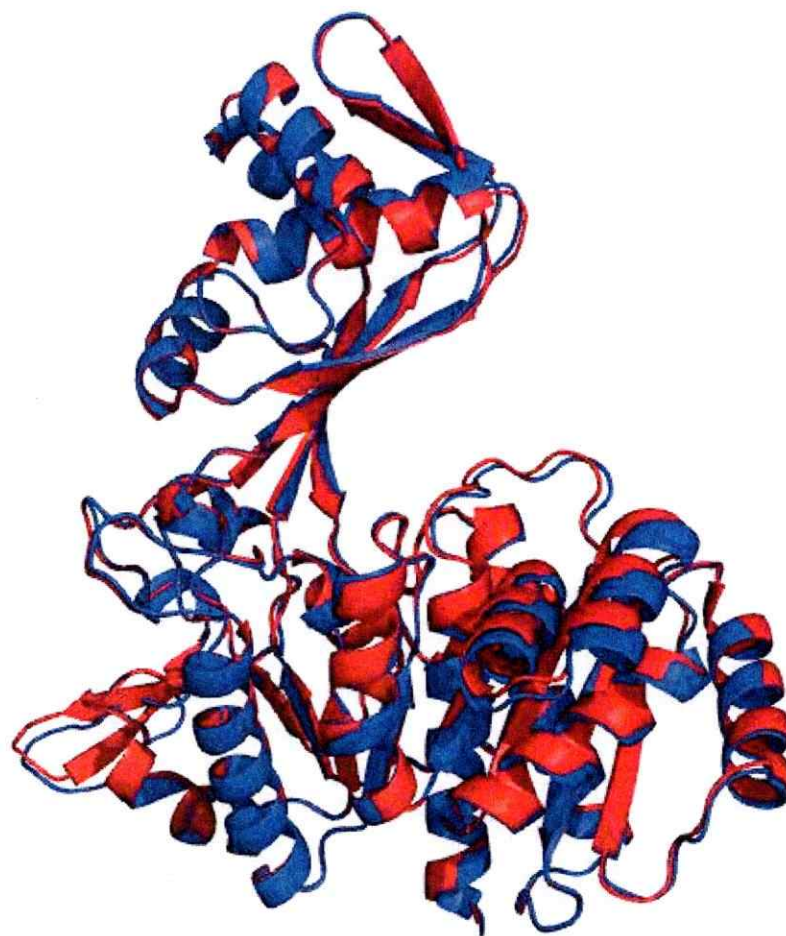
C, competitive inhibition; NC, non-competitive inhibition.  $K_{iq}$  and  $K_{ip}$  represents the inhibition constants for Mg·AMP and D-glucose-6-P, respectively.

#### 5.4.4 Cambios conformacionales evidenciados por determinación de la estructura de la fosfofructoquinasa dependiente de ADP de *Pyrococcus horikoshii*

Hasta la fecha la única PFK dependiente de ADP cristalizada es la de *Pyrococcus horikoshii* de la cual existen 2 estructuras depositadas en la base de datos PDB, códigos 1U2X y 3DRW. La primera estructura se encuentra sin sustratos y la segunda se encuentra con AMP en su sitio activo [85]. Ambas estructuras presentaron dos unidades asimétricas en la celda unitaria (cadenas A y B) las que se encuentran en la misma conformación (abierta) al comparar la estructura apo y con AMP (**Figura 20**), por lo que no hay evidencia que los ligandos provoquen un cambio conformacional en esta enzima.

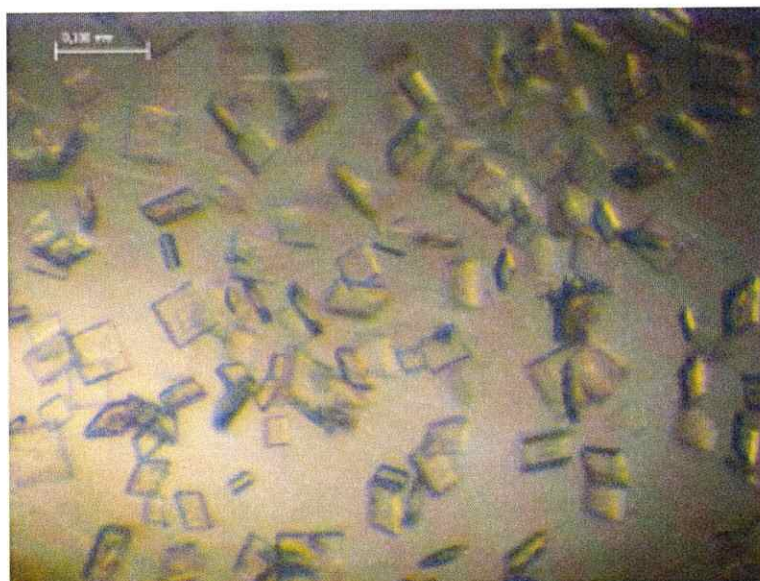
Con el fin de obtener una estructura cristalográfica en presencia de ambos ligando se realizaron pruebas de cristalización en presencia de AMP y fructosa-6P, encontrando una condición donde se obtuvieron cristales (**Figura 21**) y patrones de difracción adecuados para resolver una estructura (**Figura 22**) (**Tabla 11**). La resolución de las fases se realizó mediante reemplazo molecular utilizando la cadena A de la estructura 3DRW, la búsqueda de las fases se realizó con el dominio menor y mayor por separado. La estructura refinada no mostró densidad electrónica apreciable en el sitio activo para asignar el AMP o la fructosa-6P y solo se pudieron asignar algunos  $\text{SO}_4^{2-}$  provenientes de la solución de cristalización. Interesantemente, aparecieron dos unidades asimétricas en conformaciones distintas, la cadena A presentó la misma conformación encontrada en las estructuras previamente determinadas (1U2X y 3DRW), pero la cadena B presentó una conformación más cerrada respecto a todas las otras estructuras (**Figura 23**). Al analizar los radios de giro ( $R_g$ ) teóricos con CRY SOL[95], se encontró que la cadena A

(24,86 Å) respecto a la cadena B (24,31 Å) presenta un cierre 0,55 Å . Este cierre es más evidente al analizar la distancia entre los centros de masa de los dominios menor y mayor de ambas cadenas; la distancia entre los dominios de la cadena A es 32,4 Å, mientras que para la cadena B es 29,4 Å, evidenciando un acercamiento de 3 Å entre los dominios en la cadena B respecto a la cadena A determinada en este trabajo y respecto de las otras estructuras depositadas en el banco PDB. Esta estructura muestra que las enzimas PFK también presentarían una movilidad de los dominios, al igual que las GKs dependientes de ADP, aunque en el caso de las PFKs el movimiento de los dominios podría no estar correlacionado con la unión de sustratos a la enzima.



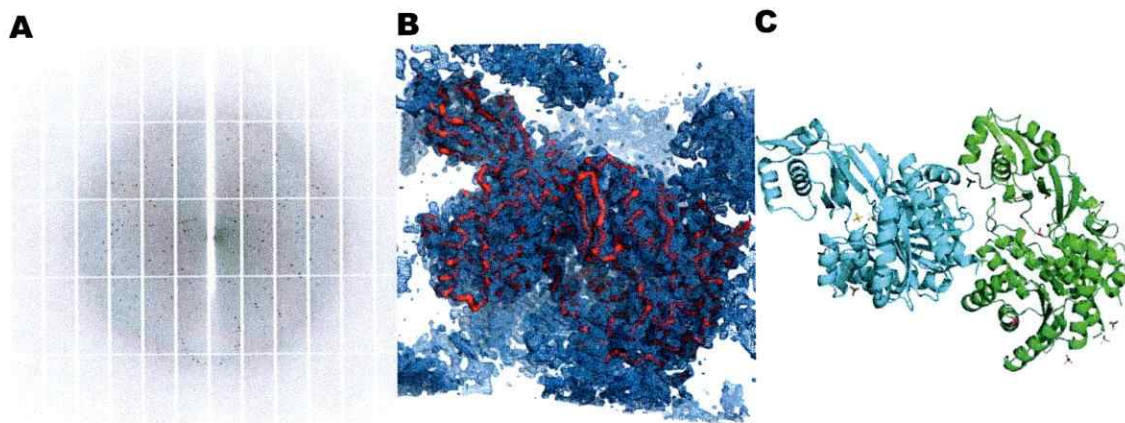
**Figura 20. Superposición de estructuras de *Ph*PFK depositadas en el banco PDB.**  
1U2X\_A (Rojo) con 3DRW\_A (Azul), se puede apreciar que 3DRW (cristalizada con AMP) presenta la misma conformación que la enzima en ausencia de ligando.





**Figura 21. Cristales de *PhPFK*.**

Cristales obtenidos por el metodo de gota sentada mezclando 1  $\mu$ L de condicion de cristalizacion (0.5 M Li<sub>2</sub>SO<sub>4</sub>, 15% w/v PEG 8000) con 1  $\mu$ L de proteina (*PhPFK* 8 mg/mL, 13 mM AMP, 20 mM MgCl<sub>2</sub>, 20 mM Fructosa-6P, 25 mM Hepes pH 7.8)

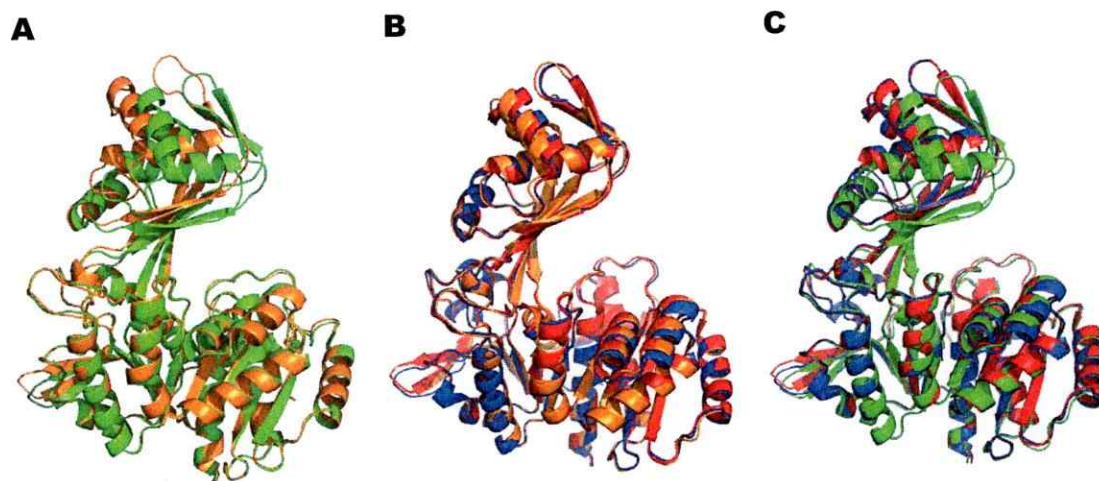


**Figura 22. Determinación de la estructura de *PhPFK*.**

A. Patrón de difracción, B, Mapa de densidad electrónica obtenido por remplazo molecular. C, Modelo obtenido luego del refinamiento.

**Tabla 11. Información cristalográfica y estadística del refinamiento para PhPFK**

Línea	i02
Long. de onda (Å)	0.97950
Rango de resol (Å)	81.32-2.533
Grupo especial	I 4 <sub>1</sub> 2 2
Celda unitaria	a=b=137.98(Å) c=294.29(Å) $\alpha = \beta = \gamma = 90(^{\circ})$
Reflexiones totales/ reflexiones únicas	94875 / 47518
Multiplicidad	8.8 ( 8.2 - 8.4)
Compleitud (%)	99.97 ( 99.4 - 100.0)
Promedio I/sigma(I)	15.09 ( 41.3 - 2.8)
R-work/ R-free	0.1980 / 0.2263
RMS(enlaces)	0.005
RMS(ángulos)	0.89
Residuos de proteína	902
Promedio de factor B (Å <sup>2</sup> )	56.50
macromoléculas	56.50
SO <sub>4</sub> <sup>2-</sup>	69.50
Solvente	46.00
Numero de átomos no hidrogeno	7547
macromoléculas	7295
SO <sub>4</sub> <sup>2-</sup>	90
Agua	162
Favorecido de Ramachandran (%)	95
Fuera de Ramachandran (%)	0.11



**Figura 23. Superposición entre la estructura de *PhPFK* determinada y las estructuras depositadas en el PDB.**

A. Superposición de cadenas A (naranja) y B (verde) de *PhPFK* determinada. B. Superposición de cadena A (naranja) con estructuras 1U2X (rojo) y 3DRW (azul), no se observa ningún cambio conformacional. C. Superposición de cadena B (verde) con estructuras 1U2X(rojo) y 3DRW (azul), se observa claramente la conformación más cerrada de la cadena B.

## 5.5 *Discusión*

Se sabe muy poco sobre los mecanismos por los cuales las estructuras de proteínas evolucionan aunque sabemos con seguridad que el cambio en las estructuras es más lento que en las secuencias de aminoácidos que las componen. De hecho, es interesante hacer notar que proteínas que se encuentran en diferentes organismos de los tres dominios de la vida, poseen estructuras que presentan un alto grado de conservación, lo que indica que el plegamiento de esas proteínas ya se encontraba presente en el organismo (o los organismos) de los cuales desciende la vida descrita en la tierra. Esta antigüedad de las estructuras de la gran mayoría de las proteínas ubicuas de la vida es el gran obstáculo que se encuentra al intentar estudiar la evolución de estructuras. Por otra parte, los cambios en las secuencias de proteínas y funciones se pueden estudiar en tiempos tan cortos como años o décadas, como es el caso de la aparición de la resistencia a antibióticos en las bacterias[21].

Los resultados de porcentajes de estructura secundaria obtenidos por difracción circular en la familia de quinasa dependientes de ADP, muestran que su estructura se ha mantenido altamente conservada entre glucoquinasas y fosfofructoquinasas, lo que se evidencia también al comparar las estructuras de GKS y PFKs actuales. Esta familia de proteínas presente en arqueas ha estado presente en estos organismos desde la era precámbrica y de acuerdo a los estimados publicados en *The Tree Time Of Life* [96] y basándonos en los eventos de especiación entre *Methanococcales* y *Thermococcales* (3594 millones de años), así como entre *Methanococcales* y *Methanosarcinales* (3468 millones de años) podemos decir que la proporción de los elementos de estructuras

secundarias en esta familia se han mantenido prácticamente inalteradas por sobre 3 billones de años. Considerando además la conservación de estructura terciaria de las proteínas actuales de la familia es bastante seguro proponer que también la estructura terciaria de ha mantenido casi inalterada. Al analizar las estructuras de GKs y PFKs actuales (*TlGK* y *PhPFK* PDB: 4B8R y 1U2X) solo se evidencian pequeños cambios como el largo de la sabana anti paralela de la cresta del dominio menor formada por las hojas  $\beta$  9 y 10, las cuales se encuentran más desarrolladas en las PFKs ya que al parecer en las GKs de *Thermococcales* ocurrió una delección de por lo menos 4 residuos. Caso contrario son las hojas  $\beta$  17 y 18, las cuales no están presentes en las PFKs de *Thermococcales* y *Methanococcales*, pero si en todas las demás proteínas de la familia, lo cual indicaría que ocurrió una delección de 4 a 5 residuos en la región en que se encuentran estas hojas  $\beta$ , lo que habría ocurrido anterior a la aparición del ancestro ancPFK-MT, cuya secuencia también habría presentado dicha delección. Cabe hacer notar que respecto a estos elementos de estructura secundaria que se perdieron durante la evolución en algunos grupos y persisten en otros, no se encuentran formas transitorias en las proteínas actuales, las que podrían estar presentes en la forma de la desaparición de 1 residuo, luego de 2 residuos y así sucesivamente. La desaparición completa de los 4 o 5 residuos repentinamente es lo que hace sentido con un mecanismo de evolución de equilibrio interrumpido, donde los cambios en las estructuras son abruptos y se dan en un periodo corto de tiempo, donde en la mayor parte del tiempo la estructura se encontraría sin cambios significativos.

Recientemente Ingles-Prieto y colaboradores (2013) publicaron un artículo donde determinaron las estructuras de tioredoxinas ancestrales de los tres dominios de la vida y de algunos grupos de eucariontes y de bacterias, mediante difracción de rayos X, el que tiene por título *Conservation of protein structure over four billion years* [97]. Los autores no encontraron ninguna diferencia en la estructura terciaria entre 7 ancestros resucitados y las proteínas actuales, salvo el largo de la primera hélice  $\alpha$  (denominada  $\alpha 1$ ), la cual presenta inserciones de 4 residuos en la rama de los eucariontes. Estos resultados están en total acuerdo con los obtenidos en este trabajo de tesis y resaltan la manera al parecer interrumpida de la evolución de las estructuras de proteínas.

El origen termófilo de la vida está apoyado por estudios filogenéticos realizados con árboles de secuencias de la subunidad pequeña del ARN ribosomal, en los cuales organismos termófilos, tanto de arqueas como de bacterias, se encuentran en la raíz. Así también el origen termófilo se apoya con dataciones geológicas en las cuales se proponen evaporaciones masivas de océanos y una alta temperatura en el origen mismo de la tierra. Los resultados de las denaturaciones térmicas de los ancestros mostraron que las proteínas provenientes de organismos hipertermófilos, mesófilos y psicrófilos provienen de ancestros termófilos, lo cual concuerda con el origen termófilo de la vida en la tierra y con su existencia en periodos previos a los 3.000 millones de años. Este origen de la vida en ambientes termófilos también ha sido estudiado por otros autores mediante la metodología de resurrección de proteínas ancestrales, en los cuales podemos mencionar estudios con nucleósido difosfato quinasas [98], factores de elongación Tu de la traducción en ribosomas [92],  $\beta$ -lactamasas [93] y tioredoxinas [99]. En todos estos

estudios los autores han encontrado que las proteínas actuales descienden de ancestros termófilos, apoyando los resultados de esta tesis y el origen de la vida en un ambiente de alta temperatura.

La relevancia de los cambios conformacionales en la glucoquinasa de *T. litoralis* se correlacionan con el mecanismo cinético determinado para la enzima. Además, y con seguridad las interacciones que sustentan la relación de estos cambios conformacionales con la unión de los sustratos, se encuentra totalmente conservada en las GKs de *Thermococcales*, aunque si bien algunas interacciones se conservan en la familia, se pueden encontrar ciertas variaciones en las PFKs de *Thermococcales*. La determinación de *Ph*PFK en una conformación más cerrada pero en ausencia de ligando presume una diferencia en estas enzimas en cuanto al mecanismo de unión de sustratos. Si bien, en las GKs el mecanismo más probable de unión es mediante ajuste inducido, en las PFKs podría actuar un mecanismo de selección conformacional, al menos respecto a la entrada del primer sustrato. La metodología que podría dar una respuesta a esta incertidumbre entre ambos mecanismos es el estudio de cinéticas de estado pre-estacionario mediante sistemas de paro de flujos, pero la ausencia de una metodología que permita diferenciar directamente el azúcar fosforilado del sustrato inicial o diferenciar ADP de AMP restringe estos estudios. Otras metodologías biofísicas como espectrometría de masas con intercambio de hidrogeno/deuterio podrían ser útiles para dar evidencia sobre el tipo de mecanismo de unión de sustrato en este tipo de enzimas, ya que permitiría determinar que estados (conformaciones) del paisaje energético son visitados, y si estos están restringidos por la presencia o ausencia de sustratos[100]. Si

ciertos estados son visitados solo en presencia de los sustratos estaríamos en presencia de *ajuste inducido* y por el contrario si los sustratos no restringen el número de estados visitados sino que solo aumentan la población de algunos estados, estaríamos frente a un mecanismo de *selección conformacional*.



## 5.6 Conclusiones

- Las proporciones de elementos de estructura secundaria se han mantenido a través de la evolución de la familia de acuerdo a la estimación por difracción circular.
- Los ancestros de la familia fueron proteínas termófilas, pero la estabilidad térmica ha ido en aumento generando las proteínas hipertermófilas actuales.
- Los cambios conformacionales de *TIGK* evidenciados en solución y en estado cristalino se correlacionan con el mecanismo cinético de la enzima. La presencia de grupos de residuos conservados, principalmente en el grupo de las GKs de *Thermococcales*, estabilizarían la conformación cerrada, en tanto que otros grupos darían cuenta de la apertura de los dominios.
- La solución de la estructura de *PhPFK* evidencia que las enzimas PFK dependientes de ADP presentarían cambios en la conformación de los dominios mayor y menor en ausencia de ligandos.
- La gran conservación estructural a través de la evolución con cambios de pequeños elementos de estructura secundaria apoya la evolución de estructuras mediante un mecanismo interrumpido o puntuado.

## **6 CAPÍTULO CUATRO**

**Historia evolutiva del uso de piridoxal e hidroximetil-pirimidina en la familia de quinasas de coenzimas dependientes de ATP**

### ***6.1 Aparición de la fosforilación de piridoxal durante la evolución de las quinasas de hidroximetil pirimidina a través de un ancestro promiscuo.***

En la familia de quinasas de coenzimas dependientes de ATP se han descrito enzimas bifuncionales que catalizan la fosforilación de hidroximetil-pirimidina (HMP) y piridoxal (PL). Estas enzimas se encuentran dentro del grupo de quinasas de hidroximetil-pirimidina (HMPK) y lejanas a las quinasas de piridoxal (PLKs) (Newman, et. al. 2006). Para investigar como la actividad bifuncional apareció en el grupo de HMPKs hemos reconstruido la filogenia de estas enzimas e inferido la secuencia de tres ancestros de la familia. Además, hemos expresado y caracterizado el último ancestro en común entre enzimas bifuncionales y HMPKs específicas y hemos modelado y realizado ensayos de docking con los respectivos sustratos para las tres proteínas ancestrales.

A continuación se presenta el manuscrito correspondiente a estos resultados el que se publicó en la revista *FEBS Letters* (2014) 588, 3068-3073. El trabajo fue ideado por Victor Castro-Fernandez, Cesar A. Ramírez -Sarmiento y Victoria Guixé, el trabajo *in silico* y experimental fue encabezado por Victor Castro-Fernández y Felipe Bravo-Moraga.



## Emergence of pyridoxal phosphorylation through a promiscuous ancestor during the evolution of hydroxymethyl pyrimidine kinases



Victor Castro-Fernandez<sup>1</sup>, Felipe Bravo-Moraga<sup>1</sup>, Cesar A. Ramirez-Sarmiento, Victoria Guixe\*

Departamento de Biología, Facultad de Ciencias, Universidad de Chile, Santiago, Chile

### ARTICLE INFO

#### Article history:

Received 7 April 2014

Revised 5 June 2014

Accepted 6 June 2014

Available online 19 June 2014

Edited by Miguel De la Rosa

#### Keywords:

HMP kinase

Pyridoxal kinase

Bifunctional enzyme

Enzyme evolution

Resurrection ancestral enzyme

### ABSTRACT

In the family of ATP-dependent vitamin kinases, several bifunctional enzymes that phosphorylate hydroxymethyl pyrimidine (HMP) and pyridoxal (PL) have been described besides enzymes specific towards HMP. To determine how bifunctionality emerged, we reconstructed the sequence of three ancestors of HMP kinases, experimentally resurrected, and assayed the enzymatic activity of their last common ancestor. The latter has ~8-fold higher specificity for HMP due to a glutamine residue (Gln44) that is a key determinant of the specificity towards HMP, although it is capable of phosphorylating both substrates. These results show how a specific enzyme with catalytic promiscuity gave rise to current bifunctional enzymes.

© 2014 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

### 1. Introduction

Although textbooks and research articles traditionally highlighted the remarkable specificity of enzyme action, the promiscuity of enzymes has received considerable attention in the past few years due to its role in the evolution of new functions. Several observations support the hypothesis that evolutionary progenitors and intermediates display broad specificity or high promiscuity. Therefore, bifunctional or promiscuous proteins could reflect ancestral forms from which two different specificities split and originated through a gene duplication event during evolution. Catalytic promiscuity can be defined as a secondary catalytic activity, since either affinity or intracellular substrate concentrations are not physiologically relevant [1], while bifunctionality refers to catalytic activities that have physiological relevance.

The family of ATP-dependent vitamin kinases from the ribokinase superfamily features a single domain with a conserved Rossmann-like fold [2], thus structurally diverging from other

enzymes of this superfamily that have, in addition, a small domain composed of a  $\beta$ -sheet that may also include some  $\alpha$ -helical insertions [3]. This family includes hydroxymethylpyrimidine kinases (HMPK, EC 2.7.1.49), pyridoxal kinases (PLK, EC 2.7.1.35) and hydroxyethylthiazole kinases (THZK, EC 2.7.1.50), enzymes that participate in the biosynthesis *de novo* of active (phosphorylated) forms of vitamin B1 (thiamine pyrophosphate) and B6 (PLP) in bacteria, essential cofactors for several enzymes involved in the metabolism of amino acids [4] and carbohydrates [5]. Although at first sight it seems that these enzymes can be easily split in different groups based on their substrate specificities, the fact is that several members of this protein family are able to phosphorylate more than one substrate. Some enzymes have been classified based either in enzyme kinetics or in structural comparisons as being specific for HMP, such as *Escherichia coli* and *Salmonella typhimurium* HMPKs derived from the *thiD* gene [6,7], or for pyridoxal (PL), like the protein codified by the *pxdY* gene of *E. coli*, where the PL moiety is so tightly bound to the enzyme that it is not released unless the protein is subjected to denaturation [8]. However, several bifunctional HMPK/PLK enzymes that are able to phosphorylate both PL and HMP have been described, such as the protein encoded by the *pxdK* gene of *E. coli* [9,10], *Trypanosoma brucei* and *Plasmodium falciparum* [11], and the protein encoded by the *thiD* gene of *Bacillus subtilis* [12]. Also, the crystal structures of bifunctional kinases from *B. subtilis* and *Staphylococcus aureus* have been recently solved [13,14]. Both enzymes have been described as possessing significant activity towards both PL and the thiamine

**Abbreviations:** HMP, 4-amino-5-hydroxymethyl-2-methylpyrimidine; HMPK, 4-amino-5-hydroxymethyl-2-methylpyrimidine kinase; PL, pyridoxal; PLK, pyridoxal kinase; PLP, pyridoxal-5'-phosphate; THZK, 4-methyl-5- $\beta$ -hydroxy-ethylthiazole kinase

\* Corresponding author. Address: Departamento de Biología, Facultad de Ciencias, Universidad de Chile, Las Palmeras 3425, Ñuñoa, Santiago 7800003, Chile. Fax: +56 2 227 12983.

E-mail address: [vguixe@uchile.cl](mailto:vguixe@uchile.cl) (V. Guixe).

<sup>1</sup> These authors contributed equally to this work.

<http://dx.doi.org/10.1016/j.febslet.2014.06.033>

0014-5793/© 2014 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.



precursor HMP, thus conferring to these enzymes a dual function in both the pyridoxal and thiamine biosynthesis pathways. Moreover, the *thiD* gene of *B. subtilis* has been found not to be essential for bacterial survival [15] and located outside the operon where thiamine biosynthetic genes are usually clustered [13,16]. Although this evidence sheds light about the emergence of PLK activity inside the group of HMPK enzymes as a late event of convergence in the evolution of these kinases, there is still no proof that the trait of their last common ancestor corresponded to the specificity towards HMP.

In order to test the evolutionary history of the ATP-dependent vitamin kinases and the role of bifunctionality in the appearance of the HMPK and PLK activities, we performed phylogenetic analysis, ancestral enzyme reconstruction, molecular modeling and docking, and we also tested experimentally the kinetic features of the resurrected last common ancestor of the HMPK enzymes. Our results show that this ancestor is able to phosphorylate both PL and HMP, but has a 8-fold higher preference for HMP. Also, the expected catalytic rate at near-physiological concentrations is higher for HMP than PL. These results provide strong evidence of how PLK activity emerged during the evolution of this protein family.

## 2. Materials and methods

### 2.1. Structural alignment and phylogenetic tree of PLKs and HMPKs

Crystal structures of PLKs and HMPKs (PDB ID: 2I5B, 4C5L, 1JXI, 1UBO, 1TD2, 2YXT, 1LHP and 2DDM) were structurally aligned using STAMP [17], and the resulting alignment was used to reconstruct a phylogenetic tree with MrBayes 3.1.2 [18]. For the analysis we performed 2 runs with 4 chains per run and heating temperature set to 0.2 for  $1 \times 10^6$  generations using the mixed analysis of fixed amino acid substitution models with fixed site-specific rates. Samples were collected every 100 steps and the initial 40% of samples were removed before summarizing trees and parameters, ensuring that the average standard deviation of split frequencies was less than 0.01 upon convergence. The consensus tree was processed in Dendroscope 3.2.10 [19].

### 2.2. Multiple sequence alignment

Sequences were collected from the non-redundant protein database (nr) using a PSI-BLAST algorithm with 3 iterations through Protein BLAST Server ([blast.ncbi.nlm.nih.gov](http://blast.ncbi.nlm.nih.gov)), using HMPKs and PLKs with known structure as templates (PDB ID: 2I5B, 1JXH, 1TD2, 2F7K and 2DDM). An initial multiple sequence alignment (MSA) was constructed using MAFFT [20], then redundant sequences were removed using QR Sequence tool of Multiseq in VMD [17] with a PID cutoff of 88%. A second MSA was constructed based on three-dimensional and secondary structure constraints using Promals3D [21]. Misaligned C-terminal, N-terminal and loop positions were manually corrected based on the structural alignment obtained previously. The final MSA used in this work is available in the Supplementary information.

### 2.3. HMPKs phylogenetic tree

Sequences of HMPK were selected from MSA and their phylogeny was inferred with MrBayes version 3.1.2 [18], where human PLK was used as outgroup. For the analysis we used WAG as the fixed model, based on its posterior probability of 1.0 in the mixed model analysis, and gamma-shaped rate variation across sites with a proportion of invariable sites. Runs were performed as previously described, but increasing the number of generations to  $1 \times 10^7$  and

only removing the initial 30% of samples before summarizing trees and parameters. The average standard deviation of the split frequencies was less than 0.01 upon convergence.

### 2.4. Ancestral sequence reconstruction

We used a modified method described by Hall [22]. Briefly, ancestral sequences were inferred through a hierarchical Bayes approach implemented in MrBayes 3.1.2 with model and parameters used for phylogeny inference [23]. Each target node was constrained and probabilities for each amino acid were calculated in each position of alignment, where the amino acid with highest probability in each position were selected. Sequence gaps were inferred as described by Hall [22].

### 2.5. Gene synthesis, protein expression and purification

The gene of the last common ancestor was codon-optimized for expression in *E. coli* and synthesized by GENSCRIPT (Piscataway, NJ, USA), then cloned into a modified pET-28b vector and verified by DNA sequencing. *E. coli* BL21(DE3) were transformed and grown in LB broth containing 35 µg/mL kanamycin at 37 °C until OD600 reached ~0.8. Expression of the recombinant protein was induced with 1 mM of isopropyl-β-D-thiogalactopyranoside overnight. Cells were harvested by centrifugation, resuspended in binding buffer (50 mM Tris-HCl pH 7.6, 500 mM NaCl, 20 mM imidazole and 5 mM MgCl<sub>2</sub>) and disrupted by sonication. After centrifugation (18514 g for 45 min), the soluble fraction was loaded onto a Ni<sup>2+</sup>-NTA affinity column (HisTrap HP, GE Healthcare, UK). Protein was eluted with a linear gradient between 20 and 500 mM imidazole and fractions with enzyme activity were pooled and stored at 4 °C with 1 mM ATP and 5% glycerol. Enzyme purity was analyzed by SDS-PAGE stained with Coomassie blue.

### 2.6. Enzyme activity assays

HMPK activity was measured following the appearance of ADP with a coupled assay containing pyruvate kinase/lactate dehydrogenase. Briefly, enzyme preparation was mixed with reaction buffer containing 25 mM Tris-HCl pH 7.8, 0.8 U/mL of pyruvate kinase, 2.4 U/mL lactate dehydrogenase, 0.3 mM phosphoenolpyruvate (PEP), 125 mM KCl, 0.2 mM NADH, 10 mM ATP, 15 mM HMP and 15 mM MgCl<sub>2</sub>.

PLK activity was measured spectrophotometrically by following PLP formation at 388 nm. The reaction mixture consisted of 10 mM ATP, 15 mM PL, 15 mM MgCl<sub>2</sub> and 25 mM PIPES pH 6.5. An extinction coefficient of 6.22 mM<sup>-1</sup> cm<sup>-1</sup> for NADH and 2.886 mM<sup>-1</sup> cm<sup>-1</sup> for PLP were used, and the enzymatic unit (U) was defined as µmol min<sup>-1</sup>. Both activities were measured at 37 °C.

### 2.7. Homology modeling and docking

Fifty models were constructed for each ancestral protein with MODELLER 8 [24]. The best 10 models were chosen based on DOPE potential, and its quality evaluated with PROSA2003 [25], Procheck [26] and VERIFY3D [27].

Docking assays were performed with AutodockVina 1.0 [28], the protonation state of the ionizable residues was calculated using the web server H++ [29] and partial charges were derived with Gasteiger method using AutoDockTools [30]. Hydrogens of HMP and PL substrates were added and optimized with Gaussian [31]. Docking results with the lowest interaction energy and the phosphoryl acceptor hydroxyl oriented towards the GXGD(C) motif were selected, since the aspartic acid (or cysteine) is considered the catalytic base for the phosphate group transfer in all ribokinase superfamily members [32].



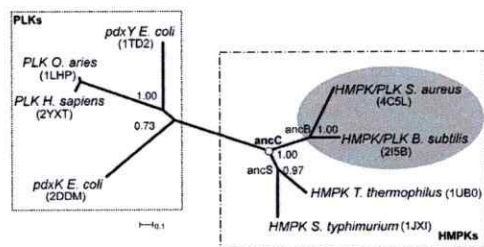
### 3. Results and discussion

#### 3.1. PLK and HMPK activity, evidence for bifunctional enzymes

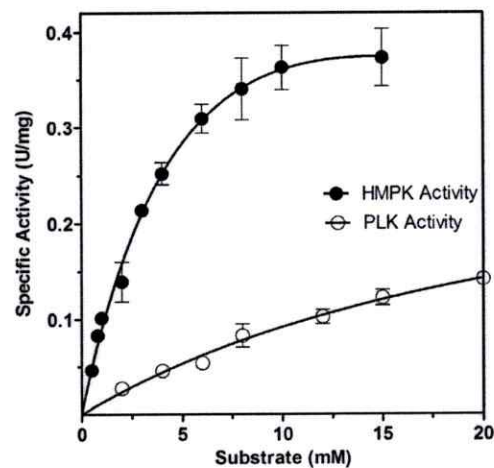
The relationship between PLKs or HMPKs and bifunctional enzymes was analyzed by constructing a phylogenetic tree based on structural alignment (Fig. 1). The resulting dendrogram shows that enzymes HMPK or PLK appear as separate groups, while bifunctional enzymes cluster inside the HMPK group, as described previously [13,14]. Therefore, the HMPK group can be divided into specific HMPKs (*Thermus thermophilus* and *S. typhimurium*) and bifunctional HMPK/PLKs (*B. subtilis* and *S. aureus*) whereas the PLK group includes enzymes from prokaryotes (*pdxY* and *pdxK*) and eukaryotes (human and sheep). It is interesting to analyze the evolutionary history of the HMPK group, since the ability to phosphorylate PL within the group can be an event of convergent evolution, in which PLK activity reappears in the HMPK group as proposed by Newman et al. [13] or, conversely, the bifunctionality is the ancestral trait [33] and HMP specificity would be a recent event in the evolutionary history of these enzymes. These two hypotheses can be scrutinized by reconstructing the last common ancestor of the HMPK group (Fig. 1). The experimental resurrection of this ancestor and its kinetic characterization, regarding its ability to phosphorylate PL and HMP, will reveal if this ancestor was either specific for HMP (convergent evolution) or bifunctional (ancestral conservation of PLK activity) (see Fig. 2).

#### 3.2. HMPK phylogeny and ancestral sequence reconstruction

In order to reconstruct the last common ancestor of HMPK enzymes, we established the phylogeny of bacterial HMPKs by bayesian method (Fig. S1), using the human PLK as outgroup to root the tree. As expected, these enzymes cluster in two groups; one that contains the sequences of HMPK/PLK from *B. subtilis* and *S. aureus* (putative bifunctional group) with a posterior probability (pp) of 1.0, and other that includes specific HMPK enzymes from *S. typhimurium* and *T. thermophilus* (HMPK specific group) with a support of 0.93. The posterior probability for the node of the last common ancestor for all HMPKs was 0.92. We inferred the ancestral sequences of the last common ancestor of HMPKs enzymes (common ancestor, ancC), of the specific HMPK enzymes (ancS) and of the putative bifunctional group of enzymes (ancB) using the hierarchical bayesian approach [23]. Sequences of ancestors ancS and ancB presented a very high support of posterior probability of reconstruction; ancS (Fig. S2A) shows 94% of residues over pp = 0.90 (254 of 268 aa) (Fig. S2A), while ancB (Fig. S2B) displays 92% of residues over pp = 0.90 (245 of 266 aa) (Fig. S2B). AncC was more ambiguously reconstructed (mean pp = 0.84 overall)



**Fig. 1.** Dendrogram of PLK and HMPK activities. Enzymes with determined structure were structurally aligned and phylogeny was reconstructed. Enzymes with dual activity are shadowed in gray, while squares indicate PLK and HMPK groups. The last common ancestor of HMPK group (ancC), ancestor of HMPK specific (ancS) and ancestor of bifunctional enzymes (ancB) are labeled.



**Fig. 2.** Dependence of the enzymatic activity of the last common ancestor (ancC) as a function of the concentration of HMP or PL.

(Fig. S2C), with 55% of its residues having pp  $\geq 0.90$  (Fig. S2C). Nevertheless, the reconstruction of ligand-contacting sites was considerably more robust (mean pp = 0.94) (Table S1). This mean posterior probability is a good support for the validity of the ancestral sequence reconstruction, as reported in other publications [34,35].

#### 3.3. Experimental resurrection of the last common ancestor of HMPK enzymes

We expressed and purified the last common ancestor of HMPK enzymes (ancC). The ancestral enzyme showed activity with both substrates (PL and HMP) with an optimum pH of 8.0 and 6.5 for the HMPK and PLK activity, respectively (Fig. S3). The  $K_m$  for ATP using HMP as co-substrate was 1.00 mM; this substrate also shows a slight inhibition with a  $K_i$  of 44 mM. A  $K_m$  of 7 mM was obtained when HMP was used as substrate, a value very similar to those reported for current enzymes, like the one present in *B. subtilis* ( $K_m$  of 2 mM) [12]. At HMP concentrations above 10 mM, substrate inhibition was observed with a  $K_i$  of 31 mM. When employing PL as substrate a  $K_m$  of 28 mM, a  $k_{cat}$  of  $0.2 \text{ s}^{-1}$  and a  $K_i$  of 85 mM for substrate inhibition (Table 1) were obtained. This  $K_m$  value for PL is much higher than the ones reported for current enzymes (Table S2); both the *pdxK*-derived enzyme from *E. coli* as well as the *thiD*-derived enzyme from *B. subtilis* presents  $K_m$  values for PL in the micromolar range. Comparison of the  $k_{cat}/K_m$  values for HMP and PL show that ancC displays a HMP preference of 8 times over PL, which results mainly due to its high  $K_m$  for PL. This suggest that the catalytic activity of ancC with PL corresponds to catalytic promiscuity since, due to the high chemical reactivity of this metabolite with proteins [36], its intracellular concentration in

**Table 1**  
Kinetic constants for ancC.

Parameters	HMP	PL	Mg-ATP
$K_m$ (mM)	$7 \pm 2$	$28 \pm 9$	$1 \pm 0.2$
$k_{cat}$ ( $\text{s}^{-1}$ )	$0.40 \pm 0.06$	$0.20 \pm 0.04$	$0.080 \pm 0.005$
$K_i$ (mM)	$31 \pm 17$	$85 \pm 54$	$44 \pm 19$
$k_{cat}/K_m$ ( $\text{M}^{-1} \text{s}^{-1}$ )	57	7	80

*E. coli* is 27  $\mu\text{M}$  total and only 10  $\mu\text{M}$  free, not-protein bound [37], thus suggesting that it would not be physiologically relevant. In fact, in this concentration range, the catalytic rate is significantly higher when HMP is used as substrate compared to PL. This idea is reinforced by the fact that if  $k_{\text{cat}}/K_m$  values are calculated from initial velocity slopes the difference between HMP versus PL as substrates is 11-fold (Fig. S4). In the case of the bifunctional enzyme from *B. subtilis*, there is a  $\sim 4$ -fold preference of PL over HMP that is accompanied by a  $K_m$  for PL in the micromolar range, thus exhibiting a higher catalytic rate for PL for the expected intracellular concentrations (Fig. S5) [12]. Moreover, in vivo covalent modification by analogs of the antibacterial compound rugulactone of the enzyme encoded by the *thiD* gene of *S. aureus*, which has a 65-fold preference for PL over HMP and possess a millimolar  $K_m$  for HMP [14], leads to inhibition of thiamine biosynthesis [38], thus providing indirect evidence that the millimolar  $K_m$  values observed for these enzymes are of biological relevance. Altogether, these results support that the last common ancestor of HMPK preferred HMP over PL under near-physiological conditions due to significantly different catalytic efficiencies, although it was capable of phosphorylating both substrates. Therefore, the PLK activity must have been tuned during the evolutionary history of the family in order to reappear later in the HMPK group as a convergent evolutionary event, a process that may have been favored in the absence of selective pressure as evidenced by the location of genes encoding current bifunctional enzymes outside the thiamine biosynthesis operon [12,14], and in agreement with the proposal by Newman et al. [13]. The fact that the sequence identity between modern

bifunctional enzymes and HMPKs (on average, 38%) is higher than when compared with PLKs (on average, 17%), along with the noticeable differences in the residue composition of the binding sites of both types of kinases (Fig. 4), reinforces the idea that the ability of phosphorylating pyridoxal emerged in HMPKs as a convergent, independent process.

### 3.4. Molecular modeling of ancestral HMPKs

In order to find the structural determinants that account for the specificity of ancC by HMP, we built homology models for this ancestor and for the ancestral ancS and ancB sequences (Table S3), and then performed docking assays with PL and HMP. Docking results showed the orientation for both ligands were similar to those present in crystallized enzymes. For ancC, the closest residues to PL and HMP moieties are: Met81, Ser13, Val43, Gln45, His50, His211, Val108 and Cys215 (Fig. 3). For ancB these residues are Met80, Thr12, Val42, Gln44, Val49, His209, Val107 and Cys213, and for ancS with HMP the residues are the following: Met81, Thr12, Val42, Gln44, Val49, His210, Val108 and Cys214. A comparison of the active sites of current bifunctional HMPK/PLK and specific HMPKs enzymes, shows that the main differences between them correspond to residues in positions 44 and 52 where the bifunctional enzyme from *B. subtilis* has Met and His respectively, while in the specific HMPK enzyme from *S. typhimurium* Gln44 and Val49 are found in equivalent positions [13]. In the ancC ancestor we found glutamine in position 45 and histidine in position 50, so that Gln44 can explain the HMP specificity and His50 could

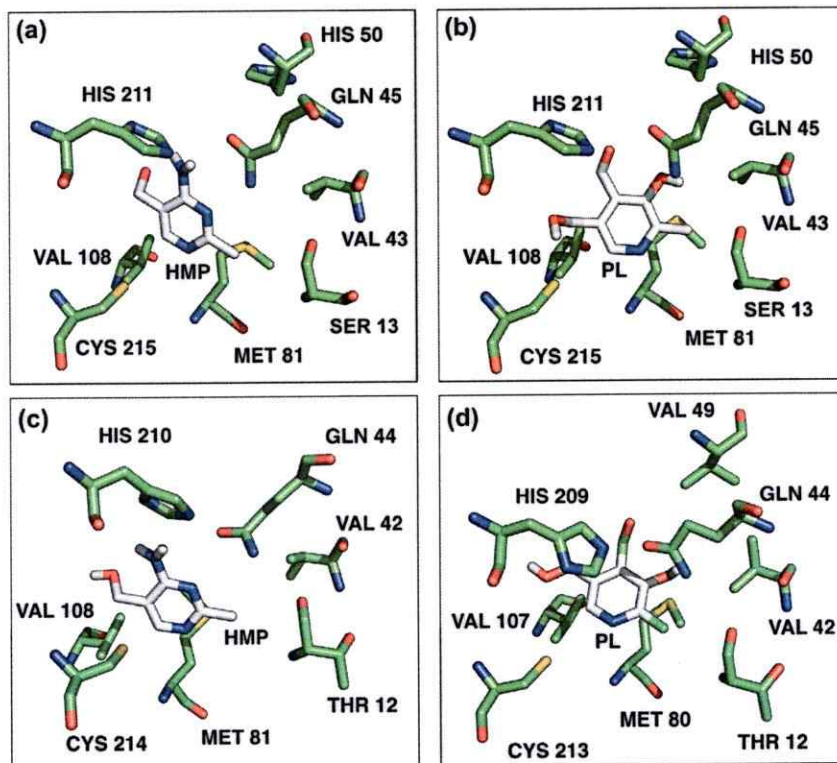
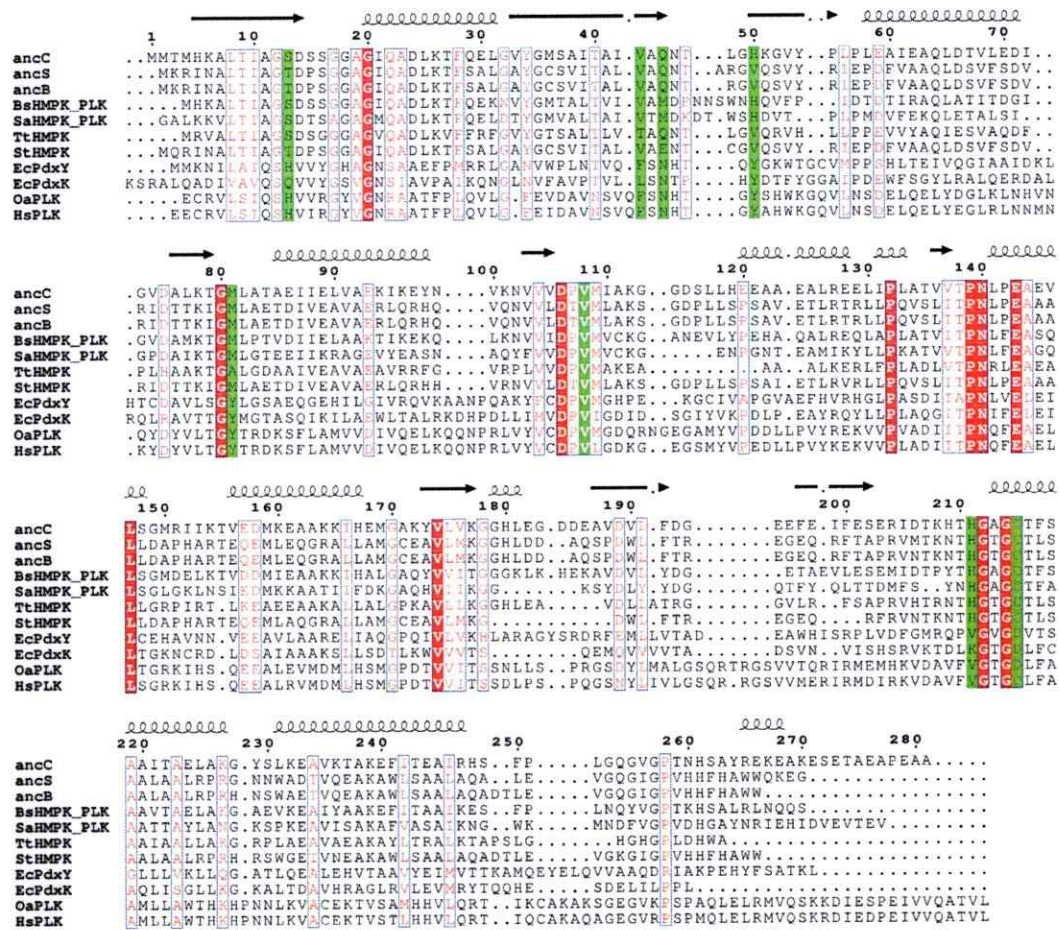


Fig. 3. Docking assays with HMP and PL on the binding site of the ancestral enzymes. (a) AncC with HMP. (b) ancC with PL. (c) ancS with HMP. (d) ancB with PL.





**Fig. 4.** Structure-based alignment of models of ancestral enzymes and structures of modern protein. BsHMPK\_PLK (*B. subtilis*, PDB: 215B), SaHMPK\_PLK (*S. aureus*, PDB: 4C5L), TtHMPK (*T. thermophilus*, PDB: 1UB0), StHMPK (*S. thyphimurium*, PDB: 1JXI), EcPdxY (*E. coli*, PDB: 1TD2), EcPdxK (*E. coli*, PDB: 2DDM), OaPLK (*O. aries*, PDB: 1LHP), HsPLK (*H. sapiens*, PDB: 2YXI). Alignment was performed with Stamp. The similarity coloring scheme is based on blossum62 matrix. Residues of the active site of ancC are highlighted in green and the secondary structure elements are annotated based on the homology model of ancC.

account for the promiscuous activity with PL, highlighting the role of Gln in HMP specificity. Moreover, ancS presents glutamine (Gln44) and valine (Val49) in the corresponding positions, which also stressed the importance of glutamine for this specificity. Interestingly, ancB presents glutamine (Gln44) and valine (Val49) in the corresponding positions, which does not match with the active sites of current bifunctional enzymes since *S. aureus* and *B. subtilis* enzymes presents Met and His in those positions (Fig. 4). A search for residue conservation at the active site inside the group of bifunctional enzymes shows that some of them (enzymes from *Megamonas funiformis*, *Caldalkalibacillus thermanum*, *Exiguobacterium sp.*, *Acetonea longum*) retain a glutamine residue like the ancestral enzyme. These enzymes are found in different branches of the bifunctional group in the phylogenetic tree (Fig. S1), therefore suggesting some of these members may also use HMP as the preferred substrate.

#### 4. Conclusions

The experimental resurrection of the last common ancestor of the HMPK group showed that probably this protein was not able to use PL under physiological conditions. Therefore, the PLK activity present in the current bifunctional enzymes must have appeared in a convergent event independently of the PLK activity of *pdxY* and *pdxK* genes, as was proposed by Newman et al. [13]. The ability of ancC to phosphorylate pyridoxal, which is 8-times less preferred than the phosphorylation of HMP, was considered by us as a promiscuous activity, since its high  $K_m$  value for PL would not be physiologically relevant. The promiscuous activity of enzymes has been proposed as the starting point for new activities [1,39]. In our case, this trait would have allowed the appearance of an activity already present in this family, in a convergent and independent manner.



## Acknowledgment

This work was supported by Fondo Nacional de Desarrollo Científico y Tecnológico (Fondecyt, Chile) Grant 1110137.

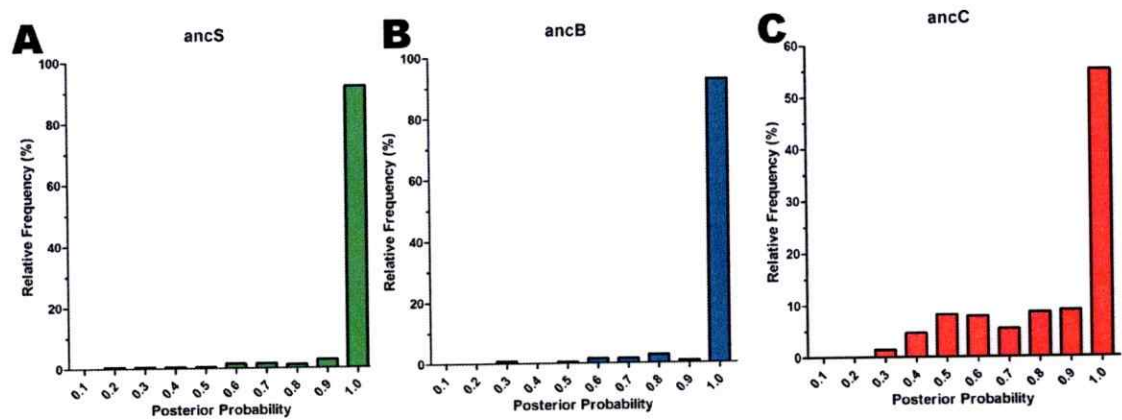
## Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version, at <http://dx.doi.org/10.1016/j.febslet.2014.06.033>.

## References

- Allewell, N.M. (2012) Thematic minireview series on enzyme evolution in the post-genomic era. *J. Biol. Chem.* 287, 1–2.
- 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, 183–193.
- Rivas-Pardo, J.A., Herrera-Morande, A., Castro-Fernandez, V., Fernandez, F.J., Vega, M.C. and Guixé, V. (2013) Crystal structure, SAXS and kinetic mechanism of hyperthermophilic ADP-dependent glucokinase from *Thermococcus litoralis* reveal a conserved mechanism for catalysis. *PLoS One* 8, e66687.
- Eliot, A.C. and Kirsch, J.F. (2004) Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. *Annu. Rev. Biochem.* 73, 383–415.
- Pohl, M., Sprenger, G.A. and Müller, M. (2004) A new perspective on thiamine catalysis. *Curr. Opin. Biotechnol.* 15, 335–342.
- 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, 225–235.
- Reddick, J.J., Kinslanda, C., Nicewongera, R., Christianb, T., Downsb, D.M., Winklerc, M.E. and Begley, T.P. (1998) Overexpression, purification and characterization of two pyrimidine kinases involved in the biosynthesis of thiamin: 4-amino-5-hydroxymethyl-2-methylpyrimidine kinase and 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate kinase. *Tetrahedron* 54, 15983–15991.
- 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, 8074–8082.
- Mizote, T. and Nakayama, H. (1989) Purification and properties of hydroxymethylpyrimidine kinase from *Escherichia coli*. *Biochim. Biophys. Acta* 991, 109–113.
- Yang, Y., Tsui, H.C., Man, T.K. and Winkler, M.E. (1998) Identification and function of the pdxY gene, which encodes a novel pyridoxal kinase involved in the salvage pathway of pyridoxal 5'-phosphate biosynthesis in *Escherichia coli* K-12. *J. Bacteriol.* 180, 1814–1821.
- Wrenger, C., Eschbach, M.-L., Müller, I.B., Laun, N.P., Begley, T.P. and Walter, R.D. (2006) Vitamin B1 de novo synthesis in the human malaria parasite *Plasmodium falciparum* depends on external provision of 4-amino-5-hydroxymethyl-2-methylpyrimidine. *Biol. Chem.* 387, 41–51.
- Park, J.-H., Burns, K., Kinsland, C. and Begley, T.P. (2004) Characterization of two kinases involved in thiamine pyrophosphate and pyridoxal phosphate biosynthesis in *Bacillus subtilis*: 4-amino-5-hydroxymethyl-2-methylpyrimidine kinase and pyridoxal kinase. *J. Bacteriol.* 186, 1571–1573.
- Newman, J.A., Das, S.K., Sedelnikova, S.E. and Rice, D.W. (2006) The crystal structure of an ADP complex of *Bacillus subtilis* pyridoxal kinase provides evidence for the parallel emergence of enzyme activity during evolution. *J. Mol. Biol.* 363, 520–530.
- Nodwell, M.B., Koch, M.F., Alte, F., Schneider, S. and Sieber, S.A. (2014) A sub-family of bacterial ribokinases utilizes a hemithioacetal for pyridoxal phosphate salvage. *J. Am. Chem. Soc.* 136, 4992–4999.
- Kobayashi, K., Ehrlich, S.D., Albertini, A., et al. (2003) Essential *Bacillus subtilis* genes. *Proc. Natl. Acad. Sci. USA* 100, 4678–4683.
- Petersen, L.A. and Downs, D.M. (1997) Identification and characterization of an operon in *Salmonella typhimurium* involved in thiamine biosynthesis. *J. Bacteriol.* 179, 4894–4900.
- Roberts, E., Eargle, J., Wright, D. and Luthey-Schulten, Z. (2006) MultiSeq: unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics* 7, 382.
- Ronquist, F. and Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574.
- Huson, D.H. and Scornavacca, C. (2012) Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.* 61, 1061–1067.
- Katoh, K., Misawa, K., Kuma, K. and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066.
- Pei, J., Kim, B.-H. and Grishin, N.V. (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* 36, 2295–2300.
- Hall, B.G. (2006) Simple and accurate estimation of ancestral protein sequences. *Proc. Natl. Acad. Sci. USA* 103, 5431–5436.
- Huelsenbeck, J.P. and Bollback, J.P. (2001) Empirical and hierarchical Bayesian estimation of ancestral states. *Syst. Biol.* 50, 351–366.
- Sali, A. and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815.
- Sippl, M.J. (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins Struct. Funct. Genet.* 17, 355–362.
- 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. Crystallogr.* 26, 283–291.
- Bowie, J.J., Lüthy, R. and Eisenberg, D. (1991) A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 253, 164–170.
- Trott, O. and Olson, A.J. (2009) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461.
- 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, W368–W371.
- Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S. and Olson, A.J. (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 30, 2785–2791.
- M.J. Frisch et al., **Gaussian09 Revision A.1.**, Gaussian 09, Revis. A02, 2009.
- Gandhi, A.K., Ghatge, M.S., Musayev, F.N., Sease, A., Aboagye, S.O., di Salvo, M.L., Schirch, V. and Safo, M.K. (2009) Kinetic and structural studies of the role of the active site residue Asp235 of human pyridoxal kinase. *Biochem. Biophys. Res. Commun.* 381, 12–15.
- Khersonsky, O. and Tawfik, D.S. (2010) Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* 79, 471–505.
- Eick, G.N., Colucci, J.K., Harms, M.J., Ortlund, E.A. and Thornton, J.W. (2012) Evolution of minimal specificity and promiscuity in steroid hormone receptors. *PLoS Genet.* 8, e1003072.
- Hobbs, J.K., Shepherd, C., Saul, D.J., Demetras, N.J., Haaning, S., Monk, C.R., Daniel, R.M. and Arcus, V.L. (2012) On the origin and evolution of thermophily: reconstruction of functional precambrian enzymes from ancestors of *Bacillus*. *Mol. Biol. Evol.* 29, 825–835.
- Christensen, H.N. (1958) Three Schiff base types formed by amino acids, peptides and proteins with pyridoxal and pyridoxal-5-phosphate. *J. Am. Chem. Soc.* 80, 99–105.
- Fu, T.F., di Salvo, M. and Schirch, V. (2001) Distribution of B6 vitamers in *Escherichia coli* as determined by enzymatic assay. *Anal. Biochem.* 298, 314–321.
- Nodwell, M.B., Menz, H., Kirsch, S.F. and Sieber, S.A. (2012) Rugulactone and its analogues exert antibacterial effects through multiple mechanisms including inhibition of thiamine biosynthesis. *ChemBioChem* 13, 1439–1446.
- Khersonsky, O., Roodveldt, C. and Tawfik, D.S. (2006) Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr. Opin. Chem. Biol.* 10, 498–508.





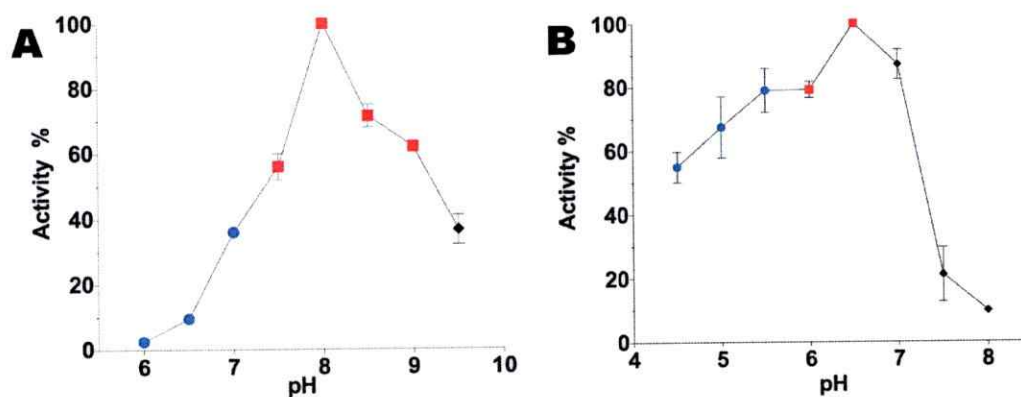
**Figure S2. Support for the reconstruction of ancestral sequences.** Posterior probability histograms: **A**, ancS. **B**, ancB. **C**, ancC.

**Table S1. Posterior probability of active site residues**

ancC	PP	ancB	PP	ancS	PP
HIS50	0.58*	VAL49	1.00	VAL49	1.00
VAL108	1.00	VAL107	1.00	VAL108	1.00
MET81	1.00	MET80	1.00	MET81	1.00
GLY12	1.00	GLY11	1.00	GLY11	1.00
SER12	0.99	THR12	1.00	THR12	1.00
ASP14	1.00	ASP13	1.00	ASP13	1.00
VAL43	0.77	VAL42	1.00	VAL42	1.00
GLN45	0.91	GLN44	0.79	GLN44	1.00
ALA19	1.00	ALA18	1.00	ALA18	1.00
GLY20	1.00	GLY19	1.00	GLY21	1.00
CYS215	1.00	CYS213	1.00	CYS214	1.00
HIS211	1.00	HIS209	1.00	HIS210	1.00
LYS112	1.00	LYS111	1.00	LYS112	1.00
mean	0.94		0.98		1.00

\*second most probable residue was Val with PP= 0.23





**Figure S3. ancC activity at different pH values.** **A**, HMPK activity. **B**, PLK activity. pH 4.5, 5.0 and 5.5 for acetate/acetic acid; pH 6.0, 6.5 and 7.0 for PIPES/NaOH; pH 7.5, 8.0, 8.5 and 8.9 for Tris-HCl, pH 9.5 Glycine/NaOH.

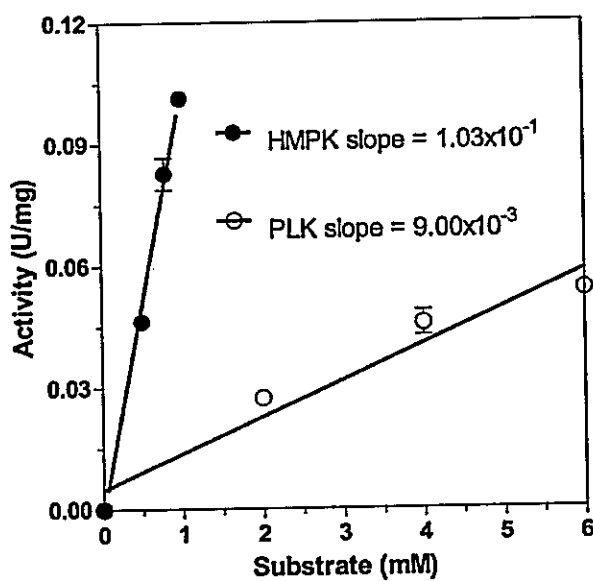
**Table S2. Reported kinetic parameters for PLKs and HMPKs**

	Km PL ( $\mu\text{M}$ )	kcat PL ( $\text{s}^{-1}$ )	Km HMP ( $\mu\text{M}$ )	kcat HMP ( $\text{s}^{-1}$ )	kcat/Km PL ( $\text{M}^{-1}\text{s}^{-1}$ )	kcat/Km HMP ( $\text{M}^{-1}\text{s}^{-1}$ )	Ref.
<b>PLK Human</b>	24	0.48	NR	NR	$20 \times 10^4$	-	[1]
<b>PLK <i>O. aries</i></b>	40	NR	NR	NR	-	-	[2]
<b>PdxK <i>E. coli</i></b>	50	4.16	NR	NR	$8.3 \times 10^4$	-	[3]
<b>ThiDS. aureus</b>	111	0.16	1988	0.045	$1.5 \times 10^3$	$2.3 \times 10^1$	[4]
<b>ThiDB. subtilis</b>	46	0.032	2030	0.36	$6.9 \times 10^2$	$1.8 \times 10^2$	[5]
<b>ThiDE. coli</b>	NR	NR	150	0.86	-	$5.7 \times 10^3$	[6]
<b>Thi3 <i>Z. mays</i></b>	NR	NR	6.6	0.019	-	$2.9 \times 10^3$	[7]

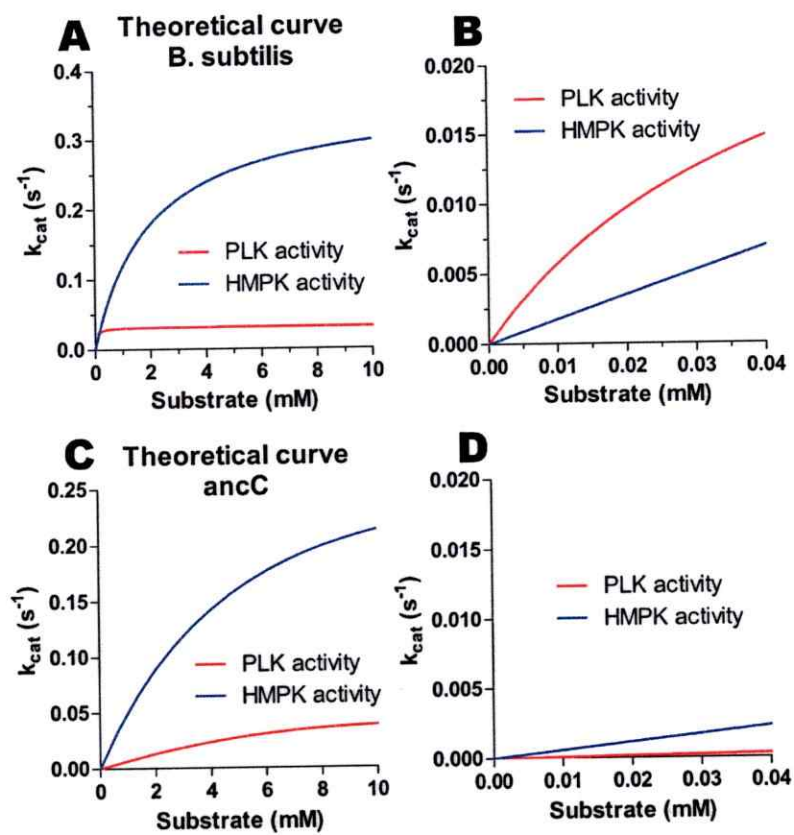
NR: not reported.

**Table S3. Homology Model Validation**

Ancestor	Dope	Procheck				Verify3D	Prosa
		Favorite	Additional	Generous	Restricted	Score>0.2	Z-score
ancC	-29812	227	19	0	0	96,61	-7,04
ancS	-29702	214	17	3	1	89,63	-7,33
ancB	-29779	213	18	1	1	92,88	-7,83



**Figure S4. Determination of  $k_{cat}/K_m$  from the initial slope of the saturation curve for the ancC.** Transformation of the slope units ( $\text{U mg}^{-1} \text{mM}^{-1}$ ) to  $k_{cat}/K_m$  unit was  $51.3 \text{ M}^{-1}\text{s}^{-1}$  for HMPK and  $4.5 \text{ M}^{-1}\text{s}^{-1}$  for PLK, resulting in an 11-fold difference.



**Figure S5. Theoretical curves in physiological range of pyridoxal for HMPK/PLK enzymes from *B. subtilis* and ancestral enzymes ancC.** A, simulated curves for Michaelis-Menten model for HMPK and PLK activity of *B. subtilis* enzyme, range 0-10 mM. B, Theoretical curves of *B. subtilis* enzymes between 0-40 μM range. C, simulated curves for substrate inhibition model for ancestral enzyme ancC, range 0-10 mM. D, Theoretical curves of ancC enzymes between 0-40 μM range. Simulations were performed with  $k_{cat}$  and  $K_m$  and  $K_i$  reported in Table S2 and Table 1.

### Supplementary Information References

1. Gandhi AK, Ghatge MS, Musayev FN, Sease A, Aboagye SO, di Salvo ML, Schirch V & Safo MK (2009) Kinetic and structural studies of the role of the active site residue Asp235 of human pyridoxal kinase. *Biochem. Biophys. Res. Commun.* 381, 12–15.
2. Kerry JA, Rohde M & Kwok F (1986) Brain pyridoxal kinase. *Eur. J. Biochem.* 585, 581–585.
3. Safo MK, Musayev FN, di Salvo ML, Hunt S, Claude J-B & Schirch V (2006) Crystal structure of pyridoxal kinase from the *Escherichia coli* pdxK gene: implications for the classification of pyridoxal kinases. *J. Bacteriol.* 188, 4542–52.
4. Nodwell MB, Koch MF, Alte F, Schneider S & Sieber SA (2014) A sub-family of bacterial ribokinases utilizes a hemithioacetal for pyridoxal phosphate salvage. *J. Am. Chem. Soc.* 136, 4992–4999.
5. Park J-H, Burns K, Kinsland C & Begley TP (2004) Characterization of two kinases involved in thiamine pyrophosphate and pyridoxal phosphate biosynthesis in *Bacillus subtilis*: 4-amino-5-hydroxymethyl-2-methylpyrimidine kinase and pyridoxal kinase. *J. Bacteriol.* 186, 1571–1573.
6. Reddicka JJ, Kinsland C, Nicewongera R, Christianb T, Downsb DM, Winklerc ME & Begley TP (1998) Overexpression, purification and characterization of two pyrimidine kinases involved in the biosynthesis of thiamin: 4-amino-5-hydroxymethyl-2-methylpyrimidine kinase and 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate kinase. *Tetrahedron* 54, 15983–15991.
7. Rapala-kozik M, Olczak M, Ostrowska K, Starosta A & Kozik A (2007) Molecular characterization of the thi3 gene involved in thiamine biosynthesis in *Zea mays*: cDNA sequence and enzymatic and structural properties of the recombinant bifunctional protein with 4-amino-5-hydroxymethyl-2-methylpyrimidine (phosphate) kinase. 159, 149–159.

## 7 DISCUSIÓN GENERAL

La evolución de proteínas aún tiene muchas preguntas sin resolver, aunque estudios pioneros nos han dado ciertas directrices y sin duda la reconstrucción de proteínas ancestrales está comenzando a ser una metodología de gran ayuda en la búsqueda de respuestas. Pál C. y colaboradores [101] ponen de manifiesto algunos de estos problemas; ¿Cómo procede la adaptación a nivel molecular? Estudios teóricos y experimentales como de evolución dirigida muestran que la adaptación a nivel molecular a menudo involucra unas pocas mutaciones de gran efecto sobre la aptitud y que posteriormente este efecto es ajustado con muchas mutaciones de pequeño efecto. ¿La neutralidad es un rasgo evolutivo? Las mutaciones neutrales no son solo el resultado de las restricciones que tienen las proteínas sobre la estructura y la función, sino que además ofrecerían robustez a los cambios genéticos y perturbaciones ambientales, una proteína con una alta fracción de sitios sometidos a mutaciones neutrales se encontraría más relajada respecto a la selección purificadora [102]. ¿Cuál es la relación entre eficiencia metabólica y evolución de proteínas? Si dos aminoácidos pueden realizar la misma función en una proteína, se selecciona el aminoácido cuya síntesis requiera menos energía, así lo refleja la composición de aminoácidos de los proteomas de *E. coli* y *Bacillus subtilis* [103]. Los resultados de resurrección de proteínas ancestrales de esta tesis apoyan claramente la necesidad de pasar por estados generalistas para la aparición de nuevas funciones y sustentan la hipótesis de que la adaptación a nivel molecular ocurre con pocas mutaciones de gran efecto en la aptitud de una proteína y que luego ocurren un gran número de mutaciones de pequeño efecto que ajustan esa aptitud. Tanto



en el caso de las quinasas dependientes de ADP, como en la quinasas de vitaminas, los cambios que habrían permitido la transición entre enzimas bifuncionales y específicas son de solo uno residuo en el sitio activo, permitiendo explicar la evolución de la familia de quinasas dependientes de ADP con un simple modelo de rasgos discretos, utilizando solo un parámetro. Estudios futuros permitirán evaluar el número de mutaciones relevantes de pequeño efecto que permiten obtener las constantes cinéticas de las proteínas actuales.

Respecto a la evolución de los parámetros cinéticos de las enzimas estudiadas, la eficiencia catalítica de las enzimas ha ido en aumento a través de la evolución, lo que está de acuerdo con que la presión selectiva actúa sobre este parámetro [24], siendo las enzimas de las vías principales del metabolismo de un organismo sometidas fuertemente a esta selección, mientras que en metabolismos secundarios esta presión sería más débil [25]. De acuerdo a esto, se podrían explicar las diferencias entre los parámetros cinéticos de las quinasas dependientes de ADP provenientes de organismos *Thermococcales* y *Methanococcales*, ya que en los *Thermococcales* la glicolisis sería una vida metabólica muy relevante en la obtención de energía, por lo que las enzimas de estos organismos presentan mayores  $k_{cat}$  y pequeñas  $K_m$ , mientras que en los *Methanococcales*, la glicolisis podría ser un metabolismo secundario, ya estos organismos obtienen la energía a través de la metanogénesis. De acuerdo a la caracterización de la enzima bifuncional de *M. maripaludis*, la glicolisis en estos organismos podría estar siendo utilizada en la dirección de síntesis de glucosa y en el almacenamiento de glicógeno en condiciones

ambientales específicas, permitiendo que las enzimas de esta vía puedan tener  $k_{cat}$  menores y  $K_m$  más altas sin alterar el metabolismo general de estos organismos.

De acuerdo a los modelos clásicos que nos permiten explicar la aparición de nuevas funciones en proteínas, ninguna de las dos familias de enzimas estudiadas se ajusta a un modelo de neofuncionalización, ya que este modelo propone la aparición de una nueva función en una de las copias de un gen posterior a una duplicación génica. Más bien, la aparición de nuevas actividades enzimáticas en ambas familias estudiadas concuerda con los modelos tipo subfuncionalización. De los dos tipos de modelos de subfuncionalización [9], el modelo de DDC contempla la aparición de una proteína bifuncional, donde luego de la duplicación génica ambas copias se especializan por sus respectivas funciones solo por deriva génica, obteniéndose finalmente dos copias mono funcionales. Por otra parte, el modelo de EAC se diferencia del anterior en que la especialización por cada una de las funciones son fijadas en la población mediante selección positiva de una mutante que permite la especialización de alguna de las copias duplicadas. La evolución de la especificidad por fructosa-6P encontrada en las PFKs de *Thermococcales* concuerda con un modelo de EAC, ya que a partir de una proteína bifuncional, que fue adquirida por THG por un ancestro *Thermococcal*, se habría perdido la función glucoquinasa mediante la mutación E72A (ancPFK-MT). Esto habría permitido que este gen se especializara por F6P y otorgando a estos organismos la ventaja de utilizar más eficientemente la glicolisis, rasgo que habría sido seleccionado y permitido la fijación de este gen en la población de los *Thermococcales*. Por otra parte, en la familia de quinasas de vitaminas la aparición de la actividad PLK en el grupo de

las HMPKs sería un estado inicial de un modelo de subfuncionalización, donde la promiscuidad catalítica del ancestro ancC por PL y la posterior mutación Q44M habrían proporcionado el surgimiento de la actividad bifuncional en las enzimas HMPK. Es importante notar que estos modelos “clásicos” no consideran las actividades promiscuas de las proteínas como puntos de partida en la generación de nuevas funciones. Es así que recientemente se ha propuesto un modelo denominado innovación, amplificación y divergencia (IAD) [14], el cual considera la selección por el nivel de expresión de un gen (*dosage selection*), selección positiva por especialización de función y considera las actividades promiscuas como el material de partida para la generación de nuevas funciones en proteínas, permitiendo explicar cómo la bifuncionalidad aparece con pocas mutaciones, tomando como punto de partida las actividades promiscuas. Así como la expresión de un gen (dosis) permite la fijación en una población de una duplicación génica, también considera que la THG puede ser clave en la amplificación de genes, sobre todo en organismos procariotas. El modelo IAD también da cuenta de la aparición de especificidad de las PFKs de *Themococcales*, considerando que la especialización por F6P otorgaría una ventaja adaptativa y concuerda con la promiscuidad por glucosa persistente luego de la especialización por F6P, lo que permite deducir que tuvo que existir un ancestro que fue específico por glucosa pero que tenía una actividad promiscua por F6P. Por otra parte, el caso estudiado en la familia de quinasas de vitaminas concuerda con un estadio inicial de este modelo, en el cual la promiscuidad por PL del ancestro ancC permite la aparición de la actividad PLK en el grupo de las HMPKs.

La conservación estructural por sobre los 3.000 millones de años en las quinasas dependientes de ADP pese a la gran divergencia de secuencias en la familia, la que alcanza hasta el 20 % de identidad dentro de las arqueas y es aún más bajo si consideramos los miembros de la familia presentes en organismos eucariontes, resalta la robustez del plegamiento a las mutaciones y sugiere que los plegamientos que actualmente conocemos aparecieron tempranamente en la evolución de la vida. Aunque poco es el conocimiento de cómo evolucionan las estructuras de las proteínas, ciertas características se han evidenciado mediante estudios de biología estructural y con proyectos como los de genómica estructural, que tienen como uno de sus objetivos caracterizar todos los tipos de plegamientos que existen en las proteínas [104]. Estos estudios muestran que las estructuras evolucionan a través de la inserción/delección de elementos de estructura secundaria, permutaciones circulares y arreglos de estructuras betas (*Hairpin flip/swap - invasión/withdawal*). Respecto a esto podemos analizar la evolución de la superfamilia Riboquinasa (**Figura 1**) en la cual se puede observar como marcador evolutivo la presencia del dominio menor, el cual se encuentra ausente en la familia de quinasas de vitaminas, la que de acuerdo a la hipótesis del aumento de complejidad de proteínas monoméricas [105] representaría la forma más ancestral de la superfamilia. Posteriormente, mediante la adquisición de hebras betas por parte de esta estructura se habría derivado en la estructura de la familia Riboquinasa, la cual presenta un dominio menor representado por una sábana de hojas betas que es estabilizada por la formación de dímeros. Luego, a partir de la familia Riboquinasa descenderían las quinasas dependientes de ADP las cuales además de la sábana beta, presentan inserciones de hélices alfa que permiten estabilizar la sábana beta y posibilita que estas

enzimas se presenten como monómeros en solución. Esta misma transición entre la estructura tipo familia Riboquinasa, cuyos miembros son dependientes de ATP, hacia las estructura tipo quinasas dependientes de ADP, contempla una permutación de la porción C-terminal la cual no habría interferido en el plegamiento de estas enzimas, conservando la misma posición de los elementos de estructura secundaria en el espacio tridimensional, pero alterando el orden de estos elementos en la secuencia primaria de estas proteínas. Por ende, la topología entre las enzimas dependientes de ATP y las dependientes de ADP es diferente en sus extremos carboxilo terminal [106]. Todos estos cambios a través de la evolución de la superfamilia concuerdan con un mecanismo de evolución tipo puntuado o interrumpido como lo analizado en la pérdida de pequeños elementos de estructura secundaria dentro de la familia de quinasas dependientes de ADP en el capítulo tres de esta tesis.

La aproximación bioquímica tradicional para estudiar las proteínas y entender como una secuencia determina una estructura y una función, contempla la caracterización de proteínas homologas actuales con diferentes funciones para intentar encontrar los residuos claves que sustentan las diferentes funciones entre la gran mayoría de diferencia que no están involucrados directamente en la función. Esta aproximación “horizontal” en el estudio de proteínas ignora completamente la historia evolutiva que ocurrió para obtener ambas funciones, ya que las funciones de las proteínas actuales son el resultado de la acumulación de mutaciones a través del tiempo de manera “vertical” en linajes de proteínas ancestrales, mientras que la comparación horizontal de las proteínas modernas solo considera las puntas del árbol evolutivo [34]. Como lo

demostrado en este trabajo, la resurrección de proteínas ancestrales ofrece la posibilidad de realizar el estudio vertical de las funciones en una familia de proteínas, permitiendo identificar de manera directa cuales fueron los cambios que ocurrieron durante la historia evolutiva de la familia que explican las funciones actuales. También, ofrece la posibilidad de explorar condiciones de los ambientes de los organismos ancestrales como es el caso de la temperatura, que fue analizada en este trabajo, aunque otros estudios también han evaluado el pH, dando soporte a un ambiente precámbrico ácido [99]. Actualmente, los métodos de reconstrucción contemplan solo modelos de evolución de secuencias de aminoácidos, nucleótidos o codones, futuros trabajos deberían buscar desarrollar modelos evolutivos que consideren rasgos estructurales no limitándose únicamente a la información obtenida de un alineamiento múltiple de secuencias.

## 8 CONCLUSIONES GENERALES

- La resurrección de proteínas ancestrales es una poderosa metodología que nos permitió realizar un estudio vertical a través de la evolución de las enzimas de la superfamilia Riboquinasa, identificando las mutaciones esenciales para la aparición de ciertas funciones y permitiendo identificar funciones en enzimas actuales.
- Las actividades enzimáticas de las familias de enzimas estudiadas en este trabajo evolucionan atravesando por estados bifuncionales, en los cuales las actividades promiscuas son las responsables que estos estados aparezcan con un número pequeño de mutaciones, como lo evidenciado en la familia de quinasas de vitaminas, mientras que después de la especialización de estos estados bifuncionales, la promiscuidad pueden ser un vestigio evolutivo de actividades ancestrales como lo mostrado en la familia de quinasas dependientes de ADP.
- Tanto la aparición de los estados bifuncionales como la especialización de éstos muestran que es necesario solo el cambio de un residuo en el sitio activo de las familias estudiadas, a pesar del gran número de diferencias en las secuencias entre ancestros y descendientes. Esto evidencia que la adaptación molecular surge con pocas mutaciones de gran efecto y que posteriormente hay un gran número de mutaciones de pequeño efecto.

- La evolución de la actividad PLK en el grupo de las HMPKs y de la actividad PFK específica de la familia de quinasas dependientes de ADP, se explica mediante el modelo de innovación, amplificación y divergencia (IAD), verificando que este mecanismo puede ser una manera general de evolución de funciones en proteínas.
- Los parámetros cinéticos entre ancestros y descendientes mostraron que la eficiencia catalítica aumenta a través de la evolución, pero que este aumento además de ser limitado por factores fisicoquímicos, sería dependiente de la importancia de la función enzimática en el metabolismo del organismo.
- La gran conservación estructural con cambios discretos en elementos pequeños de estructura secundaria en la familia de quinasas dependientes de ADP y el análisis de los cambios estructurales de la superfamilia Riboquinasa apoyan la evolución de estructuras de proteínas mediante un mecanismo puntuado o interrumpido de evolución.



## REFERENCIAS

- 1 Ureta T (2011) *Origen y evolución de proteínas y enzimas*, 1ª Edición Editorial Universitaria, Santiago.
- 2 Darwin C (1859) *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* John Murray, London.
- 3 Pincheira-Donoso D (2012) *Selección y evolución adaptativa* Ediciones UC, Santiago.
- 4 Gould SJ & Eldredge N (1993) Punctuated equilibrium comes of age. *Nature* **366**, 223–7.
- 5 Zuckerkandl E (1987) On the molecular evolutionary clock. *J. Mol. Evol.* **26**, 34–46.
- 6 Kimura M (1968) Evolutionary rate at the molecular level. *Nature* **217**, 624–626.
- 7 Wagner A (2008) Neutralism and selectionism: a network-based reconciliation. *Nat. Rev. Genet.* **9**, 965–974.
- 8 Orr HA (2005) The genetic theory of adaptation: a brief history. *Nat. Rev. Genet.* **6**, 119–27.
- 9 Conant GC & Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.* **9**, 938–50.
- 10 Ohno S (1979) *Evolution by gene duplication* Springer-Verlag, New York.
- 11 Lynch M (2007) The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 8597–8604.
- 12 Noda-García L, Camacho-Zarco AR, Medina-Ruíz S, Gaytán P, Carrillo-Tripp M, Fülöp V & Barona-Gómez F (2013) Evolution of substrate specificity in a recipient's enzyme following horizontal gene transfer. *Mol. Biol. Evol.* **30**, 2024–34.
- 13 Treangen TJ & Rocha EPC (2011) Horizontal Transfer , Not Duplication , Drives the Expansion of Protein Families in Prokaryotes. *PLoS Genet.* **7**, e1001284.
- 14 Bergthorsson U, Andersson DI & Roth JR (2007) Ohno's dilemma: evolution of new genes under continuous selection. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 17004–9.

- 15 Voet D & Voet J (2006) Bioquímica. In *Biochemistry* 3rd editio, p. 1756. Panamericana, Montevideo, Uruguay.
- 16 Todd AE, Orengo CA & Thornton JM (2001) Evolution of function in protein superfamilies, from a structural perspective. *J. Mol. Biol.* **307**, 1113–43.
- 17 Bartlett GJ, Borkakoti N & Thornton JM (2003) Catalysing New Reactions during Evolution: Economy of Residues and Mechanism. *J. Mol. Biol.* **331**, 829–860.
- 18 Gould steaphen J (1980) *The Panda's Thumb* W. W. Norton & Company, New York.
- 19 Ji Q, Norell MA, Gao KQ, Ji SA & Ren D (2001) The distribution of integumentary structures in a feathered dinosaur. *Nature* **410**, 1084–1088.
- 20 Khersonsky O & Tawfik DS (2010) Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective. *Annu. Rev. Biochem.* **79**, 471–505.
- 21 Khersonsky O, Roodveldt C & Tawfik DS (2006) Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr. Opin. Chem. Biol.* **10**, 498–508.
- 22 Fersht AR (1974) Catalysis, binding and enzyme-substrate complementarity. *Proc. R. Soc. Lond. B. Biol. Sci.* **187**, 397–407.
- 23 Crowley PH (1975) Natural selection and the Michaelis constant. *J. Theor. Biol.* **50**, 461–475.
- 24 Cornish-Bowden A (1976) The effect of natural selection on enzymic catalysis. *J. Mol. Biol.* **101**, 1–9.
- 25 Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS & Milo R (2011) The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **50**, 4402–4410.
- 26 Hammes G (2002) Multiple conformational changes in enzyme catalysis. *Biochemistry* **41**, 8221–8228.
- 27 Yang Z, Kumar S & Nei M (1995) A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* **141**, 1641–1650.
- 28 Huelsenbeck JP & Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755.
- 29 Yokoyama S, Yang H & Starmer WT (2008) Molecular basis of spectral tuning in the red- and green-sensitive (M/LWS) pigments in vertebrates. *Genetics* **179**, 2037–43.

- 30 Field SF & Matz M V (2010) Retracing evolution of red fluorescence in GFP-like proteins from *Faviina* corals. *Mol. Biol. Evol.* **27**, 225–33.
- 31 Ugalde J a, Chang BSW & Matz M V (2004) Evolution of coral pigments recreated. *Science* **305**, 1433.
- 32 Bridgham JT, Eick GN, Larroux C, Deshpande K, Harms MJ, Gauthier ME a., Ortlund E a., Degnan BM & Thornton JW (2010) Protein Evolution by Molecular Tinkering: Diversification of the Nuclear Receptor Superfamily from a Ligand-Dependent Ancestor. *PLoS Biol.* **8**, e1000497.
- 33 Thomson JM, Gaucher E a, Burgan MF, De Kee DW, Li T, Aris JP & Benner S a (2005) Resurrecting ancestral alcohol dehydrogenases from yeast. *Nat. Genet.* **37**, 630–5.
- 34 Harms MJ & Thornton JW (2010) Analyzing protein structure and function using ancestral gene reconstruction. *Curr. Opin. Struct. Biol.* **20**, 360–366.
- 35 Bork P, Sander C & 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**, 31–40.
- 36 Ronimus RS & Morgan HW (2004) Cloning and biochemical characterization of a novel mouse ADP-dependent glucokinase. *Biochem. Biophys. Res. Commun.* **315**, 652–658.
- 37 Richter S, Richter JP, Mehta SY, Gribble AM, Sutherland-Smith AJ, Stowell KM, Print CG, Ronimus RS & Wilson WR (2012) Expression and role in glycolysis of human ADP-dependent glucokinase. *Mol. Cell. Biochem.* **364**, 131–145.
- 38 Merino F & Guixé V (2011) On the specialization history of the ADP-dependent sugar kinase family. In *Gene duplication* Intech.
- 39 Datta R, Das I, Sen B, Chakraborty A, Adak S, Mandal C & Datta AK (2005) Mutational analysis of the active-site residues crucial for catalytic activity of adenosine kinase from *Leishmania donovani*. *Biochem. J.* **387**, 591–600.
- 40 Ito S, Fushinobu S, Jeong J, Yoshioka I, Koga S, Shoun H & 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**, 871–883.
- 41 Maj MC, Singh B & Gupta RS (2000) Structure-activity studies on mammalian adenosine kinase. *Biochem. Biophys. Res. Commun.* **275**, 386–393.

- 42 Sigrell JA, Cameron AD & Mowbray SL (1999) Induced fit on sugar binding activates ribokinase. *J. Mol. Biol.* **290**, 1009–1018.
- 43 Sakuraba H, Yoshioka I, Koga S, Takahashi M, Kitahama Y, Satomura T, Kawakami R & Ohshima T (2002) ADP-dependent glucokinase/phosphofructokinase, a novel bifunctional enzyme from the hyperthermophilic archaeon *Methanococcus jannaschii*. *J. Biol. Chem.* **277**, 12495–12498.
- 44 Newman J a, Das SK, Sedelnikova SE & Rice DW (2006) The crystal structure of an ADP complex of *Bacillus subtilis* pyridoxal kinase provides evidence for the parallel emergence of enzyme activity during evolution. *J. Mol. Biol.* **363**, 520–30.
- 45 Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD & Bairoch A (2005) Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook* (Walker JM, ed), pp. 571–607. Humana Press.
- 46 Ronquist F & Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.
- 47 Hall BG (2005) Comparison of the accuracies of several phylogenetic methods using protein and DNA sequences. *Mol. Biol. Evol.* **22**, 792–802.
- 48 Pei J, Kim B-H & Grishin N V (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36**, 2295–300.
- 49 Wernersson R (2003) RevTrans: multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Res.* **31**, 3537–3539.
- 50 Posada D (2008) jModelTest: Phylogenetic Averaging. *Mol. Biol. Evol.* **25**, 1253–1256.
- 51 Hall BG (2006) Simple and accurate estimation of ancestral protein sequences. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 5431–5436.
- 52 Hall BG (2008) Reconstructing Ancestral Sequences. In *Phylogenetic Trees Made Easy* Third Edit, pp. 165–176. Sinauer Associate Inc.
- 53 Abascal F, Zardoya R & Posada D (2005) ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104–2105.
- 54 Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–91.

- 55 Hanson-Smith V, Kolaczkowski B & Thornton JW (2010) Robustness of ancestral sequence reconstruction to phylogenetic uncertainty. *Mol. Biol. Evol.* **27**, 1988–1999.
- 56 Maddison WP & Maddison DR (2011) Mesquite: a modular system for evolutionary analysis. v2.75. , Available from <http://mesquiteproject.org>.
- 57 Pagel M & Meade A (2006) Bayesian Analysis of Correlated Evolution of Discrete Characters by Reversible-Jump Markov Chain Monte Carlo. *Am. Nat.* **167**.
- 58 Rambaut A, Suchard M, Xie D & Drummond A (2014) Tracer v1.6. , Available from <http://beast.bio.ed.ac.uk/Tracer>.
- 59 Kengen SW, De Bok FA, Van Loo ND, Dijkema C, Stams AJ & De Vos WM (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**, 17537–17541.
- 60 Bilski P, Li MY, Ehrenshaft M, Daub ME & Chignell CF (2000) Vitamin B6 (Pyridoxine) and Its Derivatives Are Efficient Singlet Oxygen Quenchers and Potential Fungal Antioxidants. *Photochem. Photobiol.* **71**, 129.
- 61 Eargle J, Wright D & Luthey-Schulten Z (2006) Multiple Alignment of protein structures and sequences for VMD. *Bioinformatics* **22**, 504–506.
- 62 Humphrey W, Dalke A & Schulten K (1996) VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **14**, 33–38.
- 63 Sali A & Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815.
- 64 Sippl MJ (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins Struct. Funct. Genet.* **17**, 355–362.
- 65 Laskowski RA, MacArthur MW, Moss DS & Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
- 66 Bowie JU, Luthy R & Eisenberg D (1990) Fold into a Known Three-Dimensional Structure. *Science.* **253**, 164–170.
- 67 Trott O & Olson AJ (2009) AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **31**, 455–461.

- 68 Gordon JC, Myers JB, Folta T, Shoja V, Heath LS & Onufriev A (2005) H<sup>++</sup>: a server for estimating pK<sub>a</sub>s and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* **33**, W368–371.
- 69 Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS & Olson AJ (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**, 2785–2791.
- 70 Guixé V & Merino F (2009) The ADP-dependent sugar kinase family: kinetic and evolutionary aspects. *IUBMB Life* **61**, 753–61.
- 71 Merino F, Rivas-Pardo JA, Caniuguir A, García I & Guixé V (2012) Catalytic and regulatory roles of divalent metal cations on the phosphoryl-transfer mechanism of ADP-dependent sugar kinases from hyperthermophilic archaea. *Biochimie* **94**, 516–524.
- 72 Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L & Schulten K (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802.
- 73 MacKerell A. D., Bashford D, Bellott, Dunbrack R. L., Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D & Karplus M (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **102**, 3586–3616.
- 74 Winter G (2010) Xia2: An expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190.
- 75 McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC & Read RJ (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674.
- 76 Potterton E, Briggs P, Turkenburg M & Dodson E (2003) A graphical user interface to the CCP4 program suite. *Acta Crystallogr. - Sect. D Biol. Crystallogr.* **59**, 1131–1137.
- 77 Emsley P & Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132.
- 78 Adams PD, Afonine P V., Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC & Zwart PH (2010)

- PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221.
- 79 Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS & Richardson DC (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 12–21.
- 80 Tuininga JE, Verhees CH, van der Oost J, Kengen SW, Stams a J & de Vos WM (1999) Molecular and biochemical characterization of the ADP-dependent phosphofructokinase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Biol. Chem.* **274**, 21023–21028.
- 81 Ito S, Fushinobu S, Yoshioka I, Koga S, Matsuzawa H & Wakagi T (2001) Structural basis for the ADP-specificity of a novel glucokinase from a hyperthermophilic archaeon. *Struct. London Engl.* **1993** **9**, 205–214.
- 82 Hansen T & Schönheit P (2004) ADP-dependent 6-phosphofructokinase, an extremely thermophilic, non-allosteric enzyme from the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus* strain 7324. *Extrem. life under Extrem. Cond.* **8**, 29–35.
- 83 Koga S, Yoshioka I, Sakuraba H, Takahashi M, Sakasegawa S, Shimizu S & 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**, 1079–1085.
- 84 Verhees CH, Tuininga JE, Kengen SWM, Stams AJM, Van Der Oost J & De Vos WM (2001) ADP-dependent phosphofructokinases in mesophilic and thermophilic methanogenic archaea. *J. Bacteriol.* **183**, 7145–7153.
- 85 Currie M a, Merino F, Skarina T, Wong AHY, Singer A, Brown G, Savchenko A, Caniuguir A, Guixé V, Yakunin AF & Jia Z (2009) ADP-dependent 6-phosphofructokinase from *Pyrococcus horikoshii* OT3: structure determination and biochemical characterization of PH1645. *J. Biol. Chem.* **284**, 22664–71.
- 86 Merino F & 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.
- 87 Pagel M, Meade A & Barker D (2004) Bayesian estimation of ancestral character states on phylogenies. *Syst. Biol.* **53**, 673–684.

- 88 Grishin N V (2001) Fold change in evolution of protein structures. *J. Struct. Biol.* **134**, 167–85.
- 89 Giegé R (2013) A historical perspective on protein crystallization from 1840 to the present day. *FEBS J.* **280**, 6456–97.
- 90 Kelly SM, Jess TJ & Price NC (2005) How to study proteins by circular dichroism. *Biochim. Biophys. Acta* **1751**, 119–39.
- 91 Sreerama N & Woody RW (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* **287**, 252–260.
- 92 Gaucher E a, Govindarajan S & Ganesh OK (2008) Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* **451**, 704–7.
- 93 Risso V a, Gavira J a, Mejia-Carmona DF, Gaucher E a & Sanchez-Ruiz JM (2013) Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian  $\beta$ -lactamases. *J. Am. Chem. Soc.* **135**, 2899–902.
- 94 Tsuge H, Sakuraba H, Kobe T, Kujime A, Katunuma N & 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**, 2456–2463.
- 95 Svergun D, Barberato C & Koch MH (1995) CRY SOL - A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773.
- 96 Hedges SB, Dudley J & Kumar S (2006) TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* **22**, 2971–2.
- 97 Ingles-Prieto A, Ibarra-Molero B, Delgado-Delgado A, Perez-Jimenez R, Fernandez JM, Gaucher E a, Sanchez-Ruiz JM & Gavira J a (2013) Conservation of protein structure over four billion years. *Structure* **21**, 1690–7.
- 98 Akanuma S, Nakajima Y, Yokobori S, Kimura M, Nemoto N, Mase T, Miyazono K, Tanokura M & Yamagishi A (2013) Experimental evidence for the thermophilicity of ancestral life. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 11067–72.
- 99 Perez-Jimenez R, Inglés-Prieto A, Zhao Z-M, Sanchez-Romero I, Alegre-Cebollada J, Kosuri P, Garcia-Manyes S, Kappock TJ, Tanokura M, Holmgren A, Sanchez-Ruiz JM, Gaucher EA & Fernandez JM (2011) Single-molecule paleoenzymology probes the chemistry of resurrected enzymes. *Nat. Struct. Mol. Biol.* **18**, 592–596.



- 100 Harms MJ, Eick GN, Goswami D, Colucci JK, Griffin PR, Ortlund E a & Thornton JW (2013) Biophysical mechanisms for large-effect mutations in the evolution of steroid hormone receptors. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 11475–80.
- 101 Pál C, Papp B & Lercher MJ (2006) An integrated view of protein evolution. *Nat. Rev. Genet.* **7**, 337–48.
- 102 Wagner A (2005) Robustness, evolvability, and neutrality. *FEBS Lett.* **579**, 1772–8.
- 103 Rocha EPC & Danchin A (2004) An analysis of determinants of amino acids substitution rates in bacterial proteins. *Mol. Biol. Evol.* **21**, 108–16.
- 104 Chandonia J-M & Brenner SE (2006) The impact of structural genomics: expectations and outcomes. *Science.* **311**, 347–51.
- 105 Zhang Y, Dougherty M, Downs DM & Ealick SE (2004) Crystal structure of an aminoimidazole riboside kinase from *Salmonella enterica*: implications for the evolution of the ribokinase superfamily. *Structure* **12**, 1809–21.
- 106 Merino F (2011) Tesis de Magister: Substrate specificity evolution of de ADP-dependent sugar kinase family: structure-activity studies on ADP-dependent kinases from hyperthermophilic archaea. , Facultad de Ciencias, Universidad de Chile.