



**UNIVERSIDAD DE CHILE**  
**FACULTAD DE CIENCIAS**  
**ESCUELA DE POSGRADO**

**“HISTORIA EVOLUTIVA DE LA ACTIVIDAD 4-AMINO-5-HIDROXIMETIL-2-METILPIRIMIDINA FOSFATO QUINASA DE BACTERIAS”**

Tesis entregada a La Universidad De Chile en cumplimiento parcial de los requisitos para optar al grado de Magíster en Ciencias Biológicas

**NICOLÁS ANDRÉS FUENTES UGARTE**

Director de Tesis:  
Dr. Víctor Hugo Castro Fernández

Co-directora de Tesis:  
Dra. Victoria Cristina Guixé Leguía

Enero de 2023  
Santiago – Chile

ESCUELA DE POSGRADO – FACULTAD DE CIENCIAS – UNIVERSIDAD DE CHILE



**UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS  
ESCUELA DE POSGRADO**

**INFORME DE APROBACION**

**TESIS DE MAGÍSTER**

**Se informa a la Escuela de Postgrado de la Facultad de Ciencias que la Tesis de Magister presentada por el candidato Nicolás Andrés Fuentes Ugarte.**

**“HISTORIA EVOLUTIVA DE LA ACTIVIDAD 4-AMINO-5-HIDROXIMETIL-2-METILPIRIMIDINA FOSFATO QUINASA DE BACTERIAS”**

Ha sido aprobada por la comisión de Evaluación de la tesis como requisito para optar al grado de Magíster en Ciencias Biológicas, en el examen de Defensa Privada de Tesis rendido el día .....

Director de Tesis:

Dr. Víctor Hugo Castro Fernández:

\_\_\_\_\_

Co-directora de Tesis

Dra. Victoria Cristina Guixé Leguía:

\_\_\_\_\_

Comisión de Evaluación de la Tesis

Dr. Francisco Pablo Chávez Espinosa:

\_\_\_\_\_

Dr. Marco Antonio Méndez Torres:

\_\_\_\_\_

*“Nunca consideres el estudio como una obligación,  
sino como una oportunidad para penetrar en el bello y maravilloso mundo del saber”*

*Albert Einstein*

A mis padres, porque ellos siempre estuvieron a mi lado brindándome su apoyo y sus  
consejos para hacer de mí una mejor persona.

## Biografía



Nací un domingo del mes de abril del año 1995 en Santiago, pero prontamente me remonté a concepción un par de años por motivos laborales de mi padre, gracias a esto en mis primeros años de vida crecí en presencia de inviernos lluviosos, los cuales agradan hasta el día de hoy. A la edad de cuatro años volví a Santiago, ciudad donde he vivido hasta el día de hoy y he desarrollado toda mi trayectoria académica.

Durante mi etapa escolar temprana empecé a desarrollar mi interés en el área de las ciencias naturales y matemáticas que perduran hasta la actualidad. En esos tiempos para optar a una mejor calidad de educación, en el año 2007 ingrese al Internado Nacional Barros Arana (INBA), donde realice mi educación secundaria, mi estadía en dicho recinto no fue la más tranquila dado diversos acontecimientos como: las constantes movilizaciones estudiantiles de la “revolución pingüina”, el terremoto del año 2010 que dañó gravemente al liceo y la extensa movilización estudiantil del año 2011. Pero a pesar de todas esas circunstancias, valoro haber ingresado al INBA, dado que conocí a grandes docentes que contribuyeron no solo en mi formación académica, sino que también en mi formación como persona, lo cual agradezco hasta el día de hoy.

El año 2013 tras rendir la PSU, logre ingresar a la carrera de Ingeniería en Biotecnología Molecular de la Universidad de Chile, donde durante el transcurso de la carrera conocí a grandes amigos y a la vez desarrolle el interés en las proteínas, lo cual llevó a que en

2017 ingresara al laboratorio de Bioquímica y Biología Molecular a desarrollar una pasantía junto a la Dra. Victoria Guixé, pasantía que fue fructífera, y que a futuro me llevo a desarrollar mi seminario de título en la caracterización de enzimas halófilas bajo la guía de la Dra. Guixé y el Dr. Víctor Castro, gracias a quienes desarrolle una pasión por la cinética enzimática y la biología estructural.

En afán de continuar aprendiendo sobre cinética enzimática, biología estructural y evolución de proteínas, en el año 2019 ingresé al programa de Magister en Ciencias Biológicas de la Facultad de Ciencias de la Universidad de Chile para fortalecer mi formación como científico e investigador, donde aprendí nuevas metodologías que se aplican en las áreas de biología estructural y bioinformática. Lamentablemente, por la pandemia que vivimos en los últimos años, este proyecto de tesis no se llevó a cabo como se había planificado inicialmente, pero aun así aprendí nuevas técnicas y experticias que se aprovecharon en el desarrollo de esta tesis.

Pese a todo lo malo de la pandemia, puedo destacar algunas cosas buenas dado que en nuestro laboratorio fuimos los primeros en operar de manera remota la línea MANACÁ del sincrotrón SIRIUS del Laboratorio Brasileño de Luz Sincrotrón (LNLS), São Paulo, Brasil, instancia de la cual participe como coinvestigador y se obtuvieron excelentes resultados que se muestran en esta tesis.

En la actualidad me encuentro *ad-portas* de finalizar mi magister y cursando mi primer año en el programa de Doctorado en Ciencias mención Biología Molecular, Celular y Neurociencias de la Facultad de Ciencias de la Universidad de Chile para renovar mis bases como científico y tener una base más sólida en las ciencias biológicas, para ser un mejor investigador a futuro.

## Agradecimientos

Primero que nada, agradezco a mis padres Luis y Alicia quienes siempre me ha dado su amor incondicional y apoyo en las decisiones que he tomado en mi vida. Por su esfuerzo en criarme y contribuir de gran manera a construir la persona que soy hoy en día. A mi hermano menor Vicente, por ser una gran compañía durante su vida y ahora en su juventud ser una muy grata compañía y permitirme seguir su crecimiento como persona y futuro profesional.

A Sebastián, un gran amigo que hice durante mi primer año de universidad que han estado en las buenas y en las malas hasta el día de hoy y quien ha sido una grata compañía durante nuestro desarrollo formativo en la facultad de ciencias.

A Felipe asistente de investigación del lab cuando llegue al laboratorio y actual estudiante de doctorado, quien me recibió con los brazos abiertos y me enseñó la rigurosidad del trabajo de laboratorio, realizando los experimentos “a lo caballero” lo cual sigo aplicando hasta la actualidad.

A la Dra. Victoria Guixé y al Dr. Víctor Castro, quienes dirigieron esta tesis, por su paciencia y a la libertad de trabajo que me otorgaron durante el desarrollo de esta tesis, también les agradezco su constante preocupación y guía al estar siempre disponibles para discutir los resultados de los experimentos y solventar dudas que surgieron durante el desarrollo de este seminario. En especial a Víctor, director de esta tesis a quien agradezco por sus enseñanzas y a transmitirme sus experiencias durante el desarrollo de la tesis, al realizar “ingeniería de laboratorio” con arreglos a la chilena de algunos equipos en desuso y a esas largas conversaciones sobre los resultados de mesón, lo cual fue muy grato y satisfactorio.

A la gente del laboratorio de Bioquímica y Biología Molecular que estuvieron presentes durante el desarrollo de esta tesis, y a aquellos que ya no están como: Ale, Amparo, Gisse, Juan Carlos, Leslie, Luis, Matu, Pablo C., Pablo V., Ricardo Coñuecar y aquellos quienes aún siguen en el laboratorio: Exe, Gabo, Isa, Myriam, Nachito, Nico H, Ricardo Cabrera, Sixto, Señora Elizabeth, quienes con aquellas conversaciones de pasillo ayudaron a enriquecer este trabajo.

## Financiamiento

Esta tesis fue desarrollada en el Laboratorio de Bioquímica y Biología Molecular de la Facultad de Ciencias de la Universidad de Chile, con el apoyo de los proyectos:

FONDECYT 1221667.

FONDECYT 1191321.

Proyecto de colaboración internacional para el apoyo a la cristalografía REDI170497.

Proyecto FONDEQUIP EQM120208 equipo de cristalización de proteínas automatizado de alto rendimiento ARI Crystal Gryphon.

Proyecto FONDEQUIP EQM200266 línea de rayos X a bajo ángulo SAXSpoint 2.0 de Anton Paar.

Este proyecto de tesis uso recursos del Laboratorio nacional Argonne, Chicago USA, propuesta 65859. Se agradece al equipo de la línea 23-B del “Advanced Photon Source (APS)” por su asistencia durante los experimentos de difracción.

Este proyecto de tesis uso recursos del Laboratorio Brasileño de Luz Sincrotrón (LNLS), São Paulo, Brasil y del centro nacional para investigación en energía y materiales (CNPEM), propuestas 20210185 y 20221561. Se agradece al equipo de la línea MANACA del sincrotrón SIRIUS por su asistencia durante los experimentos de difracción.



# Índice de contenidos

|  |       |
|--|-------|
| Biografía .....  | iv    |
| Agradecimientos .....  | vi    |
| Financiamiento .....   | viii  |
| Índice de contenidos .....   | ix    |
| Índice de tablas .....   | xiv   |
| Índice de figuras .....  | xvi   |
| Lista de abreviaturas .....  | xix   |
| Resumen .....  | xxi   |
| Abstract .....   | xxiii |
| 1. Introducción .....  | 1     |
| 1.1. Superfamilia Riboquinasa .....  | 1     |
| 1.2. Familia quinasas de vitaminas dependientes de ATP de bacterias .....                                    | 3     |
| 1.3. Enzimas piridoxal quinasas tipo hidroximetil pirimidina quinasa bifuncionales<br>(PdxK-PLK/HMPPK) ..... | 7     |
| 1.4. Enzimas hidroximetil pirimidina fosfato quinasas (ThiD-HMPPK) .....                                     | 9     |
| 1.5. Hipótesis .....   | 14    |
| 1.6. Objetivos .....   | 14    |
| 1.6.1. Objetivo general .....  | 14    |
| 1.6.2. Objetivos específicos .....   | 14    |
| 2. Materiales y Métodos .....  | 15    |

|   |    |
|---|----|
| 2.1. Materiales.....  | 15 |
| 2.1.1. Vectores .....   | 15 |
| 2.1.2. Medios de cultivos .....   | 15 |
| 2.1.3. Soluciones generales .....   | 15 |
| 2.1.4. Reactivos químicos generales .....   | 16 |
| 2.2. Métodos.....   | 17 |
| 2.2.1. Cuantificación de proteínas .....  | 17 |
| 2.2.2. Electroforesis de proteínas en condiciones desnaturalantes .....   | 17 |
| 2.2.3. Purificación de AncEnHMPPK.....  | 17 |
| 2.2.4. Determinación del estado oligomérico de AncEnHMPPK mediante HPLC<br>.....  | 19 |
| 2.2.5. Determinación del estado oligomérico de AncEnHMPPK mediante SAXS<br>.....  | 19 |
| 2.2.6. Separación de poblaciones homogéneas y cristalografía de proteínas para<br>la realización de ensayos de cristalización.....      | 20 |
| 2.2.7. Difracción de rayos X de cristales de AncEnHMPPK y análisis de datos   | 21 |
| 2.2.8. Reconstrucción de los complejos Mg-ATP de las estructuras<br>cristalográficas en presencia de HMP y HMP-P .....                  | 22 |
| 2.2.9. Recolección de secuencias de quinasas de vitaminas dependientes de<br>ATP de bacteria .....                                      | 23 |
| 2.2.10. Generación del perfil evolutivo e Inferencia filogenética de las quinasas de<br>vitaminas dependientes de ATP de bacteria ..... | 24 |

|  |    |
|--|----|
| 2.2.11. Reconstrucción de secuencias ancestrales .....   | 25 |
| 3. Resultados.....   | 26 |
| 3.1. Expresión y purificación de la proteína ancestral AncEnHMPPK.....   | 26 |
| 3.2. Determinación del estado oligomérico de AncEnHMPPK .....  | 26 |
| 3.3. Ensayos de co-cristalización de AncEnHMPPK con HMP, HMP-P y análogos no hidrolizables de ATP .....                    | 31 |
| 3.4. Difracción de rayos X y análisis preliminar de los datos.....   | 33 |
| 3.5. Reemplazo molecular y refinamiento.....   | 34 |
| 3.6. Análisis estructurales de AncEnHMPPK con HMP, HMP-P y análogos no hidrolizables de ATP .....                          | 39 |
| 3.7. Análisis del bolsillo de unión a sustratos de las estructuras de AncEnHMPPK .....                                     | 41 |
| 3.8. Reconstrucción filogenética por máxima verosimilitud .....  | 48 |
| 3.9. Reconstrucción de secuencias ancestrales de la familia de quinasas de vitaminas dependientes de ATP de bacteria ..... | 54 |
| 3.10. Historia evolutiva de la actividad HMP-P quinasa de bacterias .....  | 55 |
| 4. Discusión .....   | 59 |
| 4.1. Estado oligomérico quinasas de vitaminas dependientes de ATP de bacterias .....                                       | 59 |
| 4.2. Estructuras cristalinas de AncEnHMPPK en presencia de sustratos .....   | 61 |
| 4.3. Reordenamiento del sitio activo de quinasas de vitaminas dependientes de ATP y segunda fosforilación de HMP.....      | 63 |

|        |  |    |
|--------|--|----|
| 4.4.   | Historia evolutiva de la actividad HMP-P quinasa de bacterias.....                   | 67 |
| 5.     | Conclusiones .....   | 70 |
| 6.     | Bibliografía.....  | 71 |
| 7.     | Anexo .....  | 80 |
| 7.1.   | Información adicional.....   | 80 |
| 7.1.1. | Estructuras cristalográficas disponibles de ThiD-HMPPK y PdxK-PLK/HMPPK .....        | 80 |
| 7.2.   | Determinación del estado oligomérico de AncEnHMPPK .....                             | 80 |
| 7.2.1. | Cromatografía de exclusión molecular mediante HPLC.....                              | 80 |
| 7.2.2. | Dispersión de rayos X a bajo ángulo .....  | 81 |
| 7.2.3. | Análisis de interfaz de dimerización mediante servidor PISA.....                     | 82 |
| 7.3.   | Cristalografía de proteínas.....   | 83 |
| 7.3.1. | Separación de poblaciones homogéneas para realizar ensayos de cristalización .....   | 83 |
| 7.3.2. | Ensayos de cristalización para la proteína ancestral AncEnHMPPK.....                 | 84 |
| 7.3.3. | Patrón de difracción de rayos X representativo para los cristales de AncEnHMPPK..... | 85 |
| 7.3.4. | Análisis estructurales de quinasas de vitaminas dependientes de ATP de bacteria..... | 86 |
| 7.3.5. | Análisis del bolsillo de unión a sustratos de las estructuras de AncEnHMPPK.....     | 88 |

|            |   |    |
|------------|---|----|
| 7.4.       | MSA, inferencia filogenética y reconstrucción de secuencias ancestrales .....       | 90 |
| 7.4.1.     | Colección de secuencias para construcción de MSA.....                               | 90 |
| 7.4.2.     | Selección de modelo evolutivo para inferencia filogenética .....                    | 93 |
| 7.4.3.     | Comandos utilizados para inferencia filogenética y reconstrucción ancestral .....   | 93 |
| 7.4.3.1.   | Comando utilizado en PhyML para inferencia filogenética.....                        | 93 |
| 7.4.3.2.   | Comando utilizado en Lazarus para la reconstrucción de secuencias ancestrales ..... | 93 |
| 7.4.3.3.   | Secuencias proteínas ancestrales .....  | 93 |
| 7.4.3.3.1. | Secuencia AncEnHMPPK.....   | 93 |
| 7.4.3.3.2. | Secuencia AncHMPPK.....   | 94 |
| 7.4.3.3.3. | Secuencia AncC.....   | 94 |
| 7.4.3.3.4. | Secuencia AncPLK/HMPPK .....  | 94 |
| 7.4.3.3.5. | Secuencia AncUnk .....  | 94 |
| 7.4.3.4.   | Robustez reconstrucción secuencias ancestrales.....                                 | 95 |
| 7.4.4.     | Filogenia de las quinasas de vitaminas dependientes de ATP de bacterias.....        | 96 |

## Índice de tablas

|   |    |
|---|----|
| Tabla 1: Reactivos químicos utilizados. ....  | 16 |
| Tabla 2: Condiciones de cristalización AncEnHMPPK. ....   | 21 |
| Tabla 3: Volúmenes de elución para AncEnHMPPK. ....   | 28 |
| Tabla 4: Estimación bayesiana de masa molar de AncEnHMPPK mediante SAXS. ....   | 30 |
| Tabla 5: Condiciones de cristalización de AncEnHMPPK en presencia de ligandos. ...  | 31 |
| Tabla 6: Condiciones de cristalización para AncEnHMPPK. ....  | 33 |
| Tabla 7: Resumen de condiciones de cristalización de las estructuras refinadas de AncEnHMPPK. ....  | 36 |
| Tabla 8: Información cristalográfica y estadísticas del refinamiento de las estructuras de AncEnHMPPK co-cristalizado con distintos ligandos. ....                | 37 |
| Tabla 9: Información cristalográfica y estadísticas del refinamiento de las estructuras de AncEnHMPPK co-cristalizado con distintos ligandos (continuación). .... | 38 |
| Tabla 10: Comparación por RMSD de las conformaciones de las estructuras cristalográficas de los complejos ternarios de AncEnHMPPK. ....                           | 41 |
| Tabla 11: Residuos a 5 Å de los ligandos HMP/HMP-P y análogo no hidrolizable del nucleótido. ....   | 42 |
| Tabla 12: Posiciones invariantes obtenidas del alineamiento múltiple de secuencias de las quinasas de vitaminas dependientes de ATP de Bacteria. ....             | 49 |
| Tabla 13: Comparación de las secuencias ancestrales reconstruidas. ....   | 55 |
| Tabla 14: Probabilidades posteriores de residuos del sitio activo de las secuencias ancestrales reconstruidas involucrados en la unión de HMP-P. ....             | 57 |
| Tabla 15: Masa molar y estados de oligomerización reportados para enzimas ThiD-HMPPK y PdxK-PLK/HMPPK. ....   | 59 |

|   |    |
|---|----|
| Tabla suplementaria 1: Estructuras cristalográficas disponibles para enzimas PLK/HMPPK actuales.....  | 80 |
| Tabla suplementaria 2: Estructuras cristalográficas disponibles para enzimas ThiD-HMPPK actuales.....   | 80 |
| Tabla suplementaria 3: Volúmenes de elución para estándar Biorad.....   | 80 |
| Tabla suplementaria 4: Residuos de la interfaz de dimerización de AncEnHMPPK. ....  | 82 |
| Tabla suplementaria 5: Estructuras cristalográficas con interfaz de dimerización análoga a la de AncEnHMPPK.....                              | 83 |
| Tabla suplementaria 6: Distancias atómicas de residuos de la capa de coordinación de ligandos de los complejos HMP/Mg-ATP y HMP-P/MgATP. .... | 89 |
| Tabla suplementaria 7: Distancias atómicas de los grupos dadores y aceptores de fosforilo en AncEnHMPPK.....                                  | 89 |
| Tabla suplementaria 8: Listado de secuencias colectadas para construcción de MSA e inferencia filogenética.....                               | 90 |
| Tabla suplementaria 9: Valores de akaike para los 5 mejores modelos evolutivos evaluados en software Protest.....                             | 93 |

## Índice de figuras

|   |    |
|---|----|
| Figura 1: Estructura cuaternaria de miembros de la superfamilia riboquinasa. ....                                     | 2  |
| Figura 2: Síntesis de vitamina B1 y B6. ....  | 4  |
| Figura 3: Reacciones enzimáticas catalizadas por las quinasas de vitaminas dependientes de ATP. ....                  | 5  |
| Figura 4: Filogenia esquemática de quinasas de vitaminas de ATP. ....   | 6  |
| Figura 5: Estructura tridimensional de la ThiD-HMPPK de <i>Salmonella typhimurium</i> . ...                           | 11 |
| Figura 6: Purificación de AncEnHMPPK. ....  | 26 |
| Figura 7: Cromatogramas Superdex 75 Increase 10/300. ....   | 28 |
| Figura 8: Analisis por SAXS para AncEnHMPPK. ....   | 30 |
| Figura 9: Modelo <i>ab initio</i> de AncEnHMPPK en solución a una concentración de 6,6 mg/mL. ....                    | 30 |
| Figura 10: Cristalización AncEnHMPPK. ....  | 32 |
| Figura 11: Modelo representativo de los 4 monómeros en la unidad asimétrica de la estructura de AncEnHMPPK. ....      | 36 |
| Figura 12: Estructura cristalográfica representativa de AncEnHMPPK. ....  | 39 |
| Figura 13: Superposición estructural de las estructuras cristalográficas de la proteína ancestral AncEnHMPPK. ....    | 40 |
| Figura 14: Bolsillo de unión de nucleótido de AncEnHMPPK en presencia de HMP y HMP-P. ....                            | 43 |
| Figura 15: Conformaciones de los análogos de ADP y ATP en el bolsillo de unión a sustrato de AncEnHMPPK. ....         | 45 |
| Figura 16: Bolsillo de unión a sustrato de AncEnHMPPK unido a HMP-P y ATP <sub>γS</sub> en conformación atípica. .... | 47 |



|   |    |
|---|----|
| Figura 17: Reconstrucción de los complejos Mg-ATP para los bolsillos de unión de AncEnHMPPK en complejo ternario con HMP/Mg-AMPCP y HMP-P/ATP <sub>γS</sub> ..... | 48 |
| Figura 18: Alineamiento representativo de quinasas de vitaminas dependientes de ATP de bacterias.....   | 51 |
| Figura 19: Relaciones filogenéticas de las quinasas de vitaminas dependientes de ATP de bacterias.....  | 52 |
| Figura 20: Histograma de distribución de probabilidades posteriores para la reconstrucción de secuencias ancestrales.....   | 55 |
| Figura 21: Ruta evolutiva de las quinasas de vitaminas dependientes de ATP de bacterias.....  | 58 |
| Figura 22: Cambios conformacionales inducidos por unión a ligandos en AncEnHMPPK. ....  | 64 |
| Figura 23: Bolsillo de unión a sustrato de SaPLK/HMPPK.....   | 67 |
| Figura suplementaria 1: Patrón de dispersión de SAXS representativo para AncEnHMPPK. ....   | 81 |
| Figura suplementaria 2: Interfaz de dimerización de AncEnHMPPK.....   | 82 |
| Figura suplementaria 3: Cromatografía de exclusión molecular preparativa de AncEnHMPPK para la separación de poblaciones homogéneas. ....                         | 83 |
| Figura suplementaria 4: Análogos no hidrolizables de ADP y ATP empleados para los ensayos de cristalización por gota sentada.....                                 | 84 |
| Figura suplementaria 5: Patrón de difracción representativo de AncEnHMPPK.....  | 85 |
| Figura suplementaria 6: Estructuras cristalográficas de AncEnHMPPK.....   | 86 |
| Figura suplementaria 7: Comparación de topología quinasas de vitaminas dependientes de ATP. ....  | 87 |

|  |    |
|--|----|
| Figura suplementaria 8: Conformaciones relevante de los grupos fosfato de los análogos de ATP en el bolsillo de unión a sustrato de AncEnHMPPK. .... | 88 |
| Figura suplementaria 9: Probabilidad posterior respecto posición de secuencia para las proteínas ancestrales inferidas. ....                         | 95 |
| Figura suplementaria 10: Alineamiento representativo extendido quinasas de vitaminas dependientes de ATP de bacterias. ....                          | 96 |
| Figura suplementaria 11: Filogenia extendida de las quinasas de vitaminas dependientes de ATP de bacterias. ....                                     | 97 |

## Lista de abreviaturas

|                   |  |
|-------------------|--|
| Aa                | Aminoácido   |
| ADP               | Adenosín difosfato   |
| ADP <sub>βS</sub> | Adenosín-5'-(β-tio) difosfato  |
| AMP-PNP           | Adenilil imido difosfato   |
| AncEnHMPPK        | Ancestro de 4-amino-5-fosfohidroximetil-2-metilpirimidina quinazas del orden Enterobacteriales |
| ATP               | Adenosín trifosfato  |
| ATP <sub>γS</sub> | Adenosín-5'-(γ-tio) trifosfato   |
| C $\alpha$        | Carbono alfa   |
| FPLC              | Cromatografía líquida de proteínas a alta velocidad  |
| HEPES             | Ácido 4-(2-hidroxietil)-1-piperazineetanosulfónico   |
| HMP               | 4-amino-5-hidroximetil-2-metilpirimidina   |
| HMP-P             | 4-amino-5-hidroximetil-2-metilpirimidina fosfato   |
| HMP-PP            | 4-amino-5-hidroximetil-2-metilpirimidina pirofosfato   |
| HMPK              | 4-amino-5-hidroximetil-2-metilpirimidina quinasa   |
| HMPPK             | 4-amino-5-hidroximetil-2-metilpirimidina fosfato quinasa                                       |
| HPLC              | Cromatografía líquida de alta eficacia   |
| IPTG              | Isopropil β-D-1-tiogalactopiranosido   |
| ML                | Máxima Verosimilitud   |
| OD                | Densidad óptica  |
| PAGE              | Gel de electroforesis de poliacrilamida  |
| <i>pdxK</i>       | Gen codificante de la piridoxal quinasa  |

|             |                                      |
|-------------|--------------------------------------|
| <i>pdxY</i> | Gen codificante de piridoxal quinasa |
| PL          | Piridoxal                            |
| PLK         | Piridoxal quinasa                    |
| PLK/HMPPK   | Piridoxal quinasa tipo HMPPK         |
| PLP         | Piridoxal fosfato                    |
| PM          | Piridoxina                           |
| PMSF        | Fluoruro de fenil metil sulfonilo    |
| PN          | Piridoxamina                         |
| PP          | Probabilidad posterior               |
| RMSD        | <i>Root Mean Square Deviation</i>    |
| SAXS        | Dispersión de rayos X a bajo ángulo  |
| SDS         | Dodecil sulfato de sodio             |
| TAE         | Tampón Tris-Ácido acético-EDTA       |
| TG-SDS      | Solución tampón Tris-Glicina-SDS     |
| <i>thiD</i> | Gen codificante de la HMPPK          |
| TRIS        | Tris(hidroximetil)aminometano        |

## Resumen

Las quinasas de vitaminas son fosfotransferasas dependientes de ATP que participan en la biosíntesis de vitamina B1 y B6 fosforilando en general un grupo alcohol primario (R-OH) del sustrato. Dentro de esta familia se han descrito dos enzimas homólogas que catalizan la fosforilación de 4-amino-5-hidroxi-2-metilpirimidina (HMP) a 4-amino-5-hidroximetil-2-metilpirimidina fosfato (HMP-P) en la síntesis de vitamina B1. Una de ellas corresponde a una enzima piridoxal quinasa bifuncional codificada por el gen *pdxK-tipo-HMPPK* (proteína PdxK-PLK/HMPPK), que también puede catalizar la fosforilación de piridoxal a piridoxal fosfato en la síntesis de vitamina B1. La otra, corresponde a la enzima 4-amino-5-hidroxi-2-metilpirimidina fosfato quinasa codificada por el gen *thiD* (proteína ThiD-HMPPK), enzima que también puede catalizar la fosforilación de HMP-P a 4-amino-5-hidroximetil-2-metilpirimidina pirofosfato (HMP-PP). Esta segunda reacción de fosforilación es única en la familia de quinasas de vitaminas y en la superfamilia Riboquinasa, dado que se realiza sobre un grupo metil fosfato (R-CH<sub>2</sub>-PO<sub>4</sub>) y no sobre un alcohol primario, siendo pocos los tipos de mecanismos catalíticos descritos para la fosforilación de un grupo fosfato.

Con el objetivo de determinar los factores estructurales involucrados en la fosforilación de un grupo metil fosfato en las enzimas ThiD-HMPPK y la conservación de la capacidad de fosforilar el grupo metil fosfato durante la evolución de esta familia de enzimas, se utilizó un enfoque combinado empleando cristalografía de proteínas y la reconstrucción de secuencias ancestrales. Para ello se cristalizó una enzima ThiD-HMPPK ancestral del orden *Enterobacteriales* (AncEnHMPPK), para la cual se conocían varias condiciones de cristalización en condición APO. Cristalizamos AncEnHMPPK en presencia de distintas combinaciones de HMP, HMP-P y análogos no hidrolizables de ATP o ADP. De los experimentos de cristalización y difracción de rayos X, se obtuvieron ocho estructuras cristalinas en complejos ternarios: **1.** HMP/análogo-ATP; **2.** HMP-P/análogo-ADP y **3.** HMP-P/análogo-ATP. Las estructuras presentan la topología

conservada de la superfamilia Riboquinasa, y sorprendentemente, no se encontraron conformaciones diferentes en elementos de estructura secundaria próximos al sitio activo al comparar complejos ternarios de HMP y HMP-P. Interesantemente, las cadenas laterales de los residuos His179 y Thr211 se acercan y se reorientan hacia los grupos fosfato de la molécula de ATP cuando se analiza el complejo ternario con HMP-P, lo cual sugeriría que esta dupla de residuos podría ser crucial y específica para la actividad HMP-P quinasa y no para la actividad HMP quinasa.

Debido a que la actividad enzimática de fosforilación del grupo metil fosfato no está presentes en las enzimas homólogas PdxK-PLK/HMPPK, inferimos las relaciones filogenéticas entre las enzimas ThiD-HMPPK y PdxK-PLK/HMPPK. Del análisis evolutivo se desprende que la actividad HMPPK sería el rasgo ancestral de esta rama de la familia y que la pérdida de la actividad HMPPK en la ruta evolutiva hacia las PdxK-PLK/HMPPK sería concomitante con la aparición de la actividad PLK. Por otro lado, el análisis de la conservación de los residuos del sitio activo en las secuencias ancestrales muestra que entre el ancestro ThiD-HMPPK y el último ancestro en común con las enzimas PdxK-PLK/HMPPK (AncC) solo habría ocurrido la sustitución T211A (numeración de AncEnHMPPK). Posteriormente, el descendiente de AncC que corresponde a AncPLK/HMPPK, presenta las sustituciones Q44M, A110C y H179S, junto a una inserción de dos residuos en la horquilla  $\beta$ 3- $\beta$ 4. Basándonos en este análisis evolutivo del sitio activo, las sustituciones Q44M y A110C estarían involucradas en la aparición de actividad PLK, mientras que las sustituciones T211A y H179S serían responsables de la pérdida de la actividad HMPPK en las enzimas PdxK-PLK/HMPPK.

Finalmente, en este trabajo se proponen determinantes estructurales de la actividad HMPPK en las enzimas ThiD-HMPPK, mientras que el análisis evolutivo constituye un ejemplo único de evolución molecular donde una actividad enzimática emerge concomitante con la desaparición de otra actividad en una familia de enzimas.

## Abstract

Vitamin kinases are ATP-dependent phosphotransferases involved in vitamin B1 and B6 biosynthesis phosphorylating different substrates on a primary alcohol group (R-OH). In this family, two homologous enzymes that catalyze the phosphorylation of 4-amino-5-hydroxy-2-methylpyrimidine (HMP) to 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate (HMP-P) have been described. One of them is a bifunctional pyridoxal kinase enzyme encoded by the *pdxK-like-HMPPK* gene (PdxK-PLK/HMPPK), which also catalyze the phosphorylation of pyridoxal (PL) to pyridoxal phosphate (PLP, vitamin B6). The other one, corresponds to a 4-amino-5-hydroxy-2-methylpyrimidine phosphate kinase encoded by the *thiD* gene (ThiD-HMPPK), also able to catalyze the phosphorylation of HMP-P to 4-amino-5-hydroxymethyl-2-methylpyrimidine pyrophosphate (HMP-PP). This second phosphorylation reaction is unique in the family of vitamin kinases and in the Ribokinase superfamily, since it is performed on a methyl phosphate group (R-CH<sub>2</sub>-PO<sub>4</sub>) and not on a primary alcohol, being very few the types of catalytic mechanisms describing the phosphorylation of a phosphate group.

To determine the structural factors involved in the phosphorylation of a methyl phosphate group in the ThiD-HMPPK enzymes and the conservation of the ability to phosphorylate a methyl phosphate group during the evolution of this family a combined approach of protein crystallography and ancestral sequence reconstruction was employed. For this purpose, we used an ancestral ThiD-HMPPK enzyme from *Enterobacteriales* order (AncEnHMPPK) for which several crystallization conditions for the APO enzyme were known. The AncEnHMPPK was crystallized in the presence of different combinations of HMP, HMP-P and non-hydrolysable ATP or ADP analogues. From the crystallization assays and X-ray diffraction experiments, eight crystal structures of the following ternary complexes were obtained: **1.** HMP/ATP-analogue, **2.** HMP-P/ADP-analogue and **3.** HMP-

P/ATP-analogue. The structures present the conserved topology for the Ribokinase superfamily, and surprisingly no different conformations in secondary structure elements close to the active site were observed when HMP and HMP-P ternary complexes were compared. Interestingly, in the ternary complex with HMP-P the side chains of residues His179 and Thr211 approach and reorient towards the phosphate groups of the ATP molecule, suggesting that this pair of residues might be crucial and specific for the HMP-P kinase activity and not for HMP kinase activity.

Since the enzymatic activity of a methyl phosphate group phosphorylation is not present in the homologous PdxK-PLK/HMPPK enzymes, we inferred the phylogenetic relationships between ThiD-HMPPK and PdxK-PLK/HMPPK enzymes. The evolutionary analysis shows that the HMPPK activity would be the ancestral trait in this family branch while the loss of the HMPPK activity in the evolutionary pathway towards PdxK-PLK/HMPPK would be concomitant with the appearance of PLK activity. On the other hand, analysis of the active site residues conservation, shows that between the ancestor of ThiD-HMPPK and the last common ancestor of the PdxK-PLK/HMPPK enzymes, (AncC) only one substitution (T211A) occurs (AncEnHMPPK numbering). Subsequently, the AncC descendant corresponding to AncPLK/HMPPK, has the Q44M, A110C and H179S substitutions, along with an insertion of two residues in the  $\beta$ 3- $\beta$ 4 hairpin. Based on this active site evolutionary analysis, the Q44M and A110C substitutions would be involved in the appearance of PLK activity, while the T211A and H179S substitutions would be responsible for the loss of HMPPK activity in the PdxK-PLK/HMPPK enzymes. Finally, structural determinants of HMPPK activity in ThiD-HMPPK enzymes were proposed in this work and the evolutionary analysis constitutes a unique example of molecular evolution where one enzyme activity emerges concomitant with the disappearance of another activity in an enzyme family.

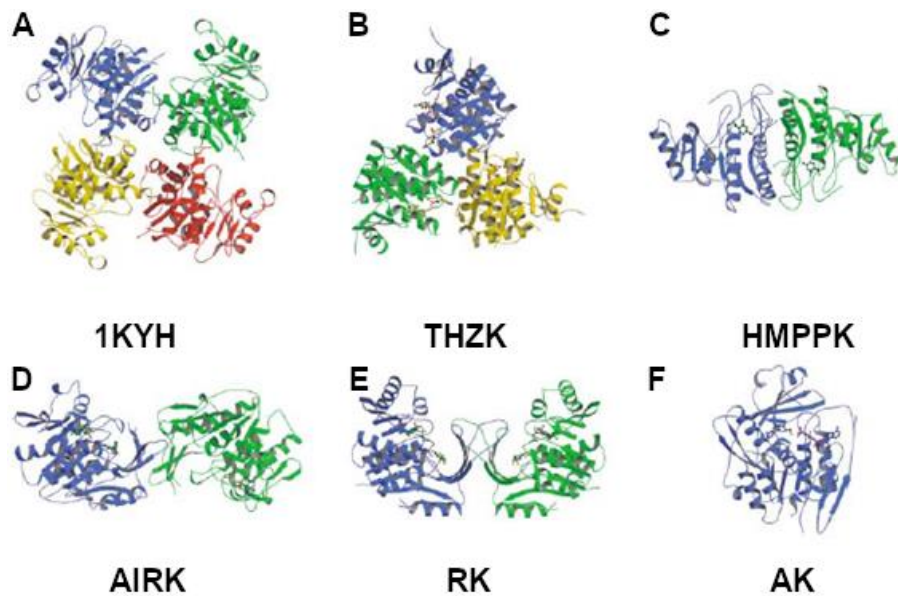


# 1. Introducción

## 1.1. Superfamilia Riboquinasa

La superfamilia Riboquinasa (EC 2.7.1.-) comprende a enzimas fosfotransferasas presentes en los tres dominios de la vida y que tienen por aceptor del fosforilo un grupo alcohol primario (R-OH). Los miembros de esta superfamilia realizan la transferencia de grupos fosforilo entre dos moléculas, por lo tanto, presentan dos sustratos: el primero corresponde al dador del fosforilo, el cual suele ser la molécula de adenosín trifosfato (ATP), mientras que el segundo sustrato corresponde al aceptor del grupo fosforilo, dentro del cual hay una gran variedad de moléculas, encontrando desde distintos tipos de azúcares, compuestos aromáticos y nucleósidos (Ramírez-Sarmiento y col., 2020).

Las enzimas de la superfamilia riboquinasa han sido estudiadas tanto a nivel estructural como de secuencia, reportándose la presencia de motivos conservados responsables de la transferencia de fosforilo (GXGD) y de la estabilización del complejo metal-nucleótido (NXXE) (Merino y col., 2012). Estructuralmente, el plegamiento de la superfamilia riboquinasa es altamente conservado, el que presenta un plegamiento tipo Rossmann compuesto de una sábana- $\beta$  central, formada de al menos ocho hebras- $\beta$  en una disposición paralela, las que se encuentran rodeadas de  $\alpha$ -hélices. Además, los miembros de esta superfamilia con una mayor extensión de secuencia presentan un dominio menor, el cual ha sido relacionado tanto con la actividad enzimática actuando como una tapa del sitio activo durante la catálisis (Sigrell y col., 1998) y/o con el estado de oligomerización de estas proteínas. En este sentido, se han reportado enzimas con distinta estructura cuaternaria, presentando diferentes estados de oligomerización (**Figura 1**), encontrando desde monómeros a tetrámeros (Zhang y col., 2004).



**Figura 1: Estructura cuaternaria de miembros de la superfamilia riboquinasa.**

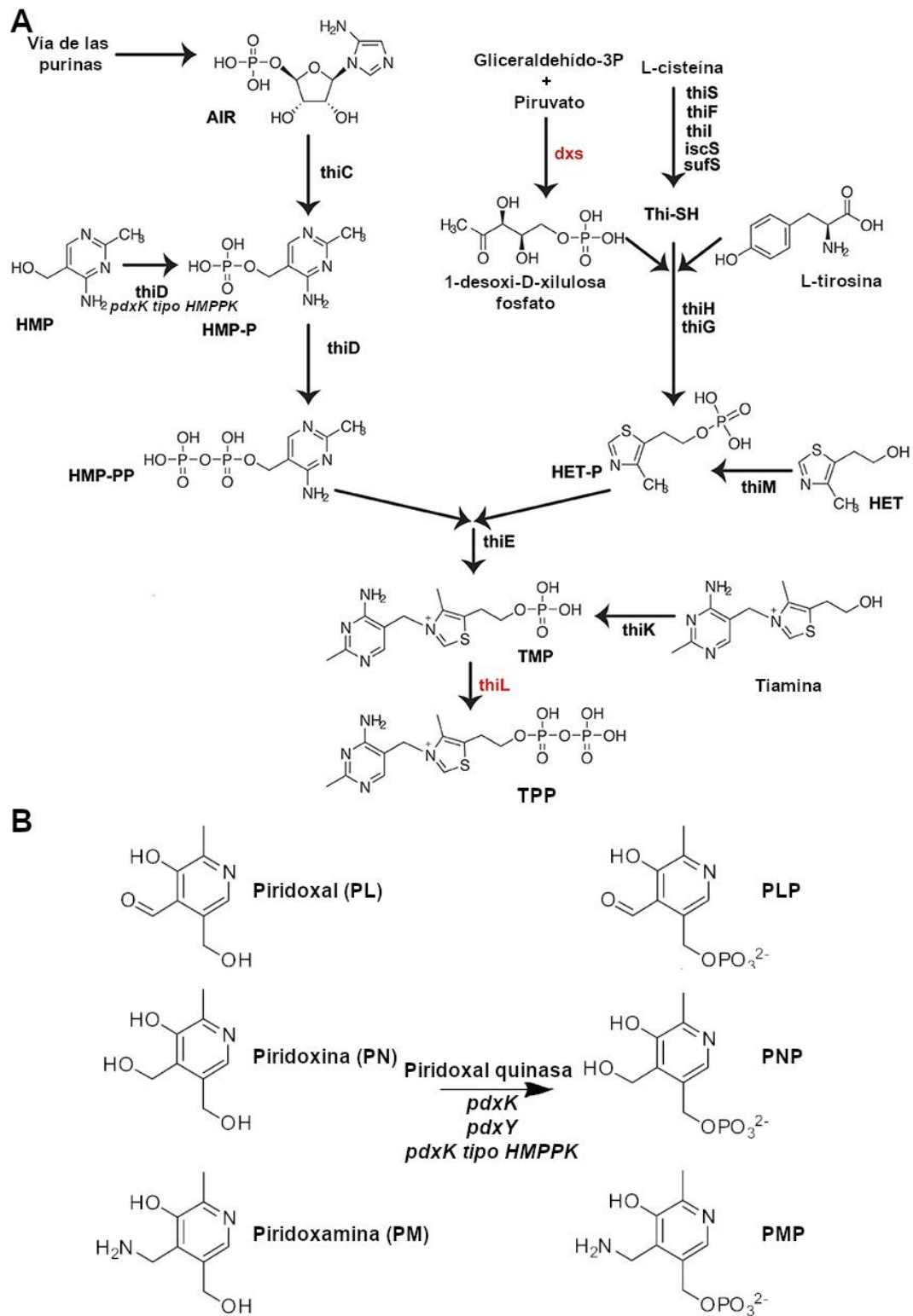
Se presentan como “cartoon” las estructuras tridimensionales de: **(A)** 1KYH: Quinasa putativa de *Bacillus subtilis*, **(B)** Tiazol quinasa. **(C)** Hidroximetil pirimidina fosfato quinasa. **(D)** Aminoimidazol ribósido quinasa. **(E)** Riboquinasa. **(F)** Adenosina quinasa. Modificado de Zhang y col., 2004

Dentro de esta superfamilia, sus miembros se pueden clasificar principalmente en tres grupos (Ramírez-Sarmiento y col., 2020): **1.** Las quinasas de azúcares dependientes de ATP (o familia riboquinasa), las cuales presentan todas la características descritas anteriormente; **2.** Las quinasas de azúcares dependientes de ADP, enzimas presentes principalmente en arqueas y en eucaria que tienen la particularidad de usar ADP como dador de fosforilo en vez de ATP y que presentan un reordenamiento en el meandro- $\beta$  y **3.** Las quinasas de vitaminas dependientes de ATP, enzimas encargadas de fosforilar precursores de vitamina B1 y B6, las que solo presentan el plegamiento tipo Rossmann, careciendo del dominio menor. Las quinasas de vitaminas dependientes de ATP serán el principal sujeto de estudio de esta tesis.

## 1.2. Familia quinasas de vitaminas dependientes de ATP de bacterias

La familia de quinasas de vitaminas dependientes de ATP son enzimas fosfotransferasas que participan en la biosíntesis de tiamina (vitamina B1) y piridoxal fosfato (vitamina B6) en bacterias y que son cofactores esenciales de enzimas participantes de la biosíntesis de aminoácidos (Grubman y col., 2010) y carbohidratos (Dick y col., 2010). Por otra parte, se ha reportado que la vitamina B6 es crucial para la virulencia y sobrevivencia de bacterias patogénicas como *Streptococcus pneumoniae*, *Helicobacter pylori*, *Mycobacterium tuberculosis*, y *Vibrio cholerae* (Nodwell y col., 2014).

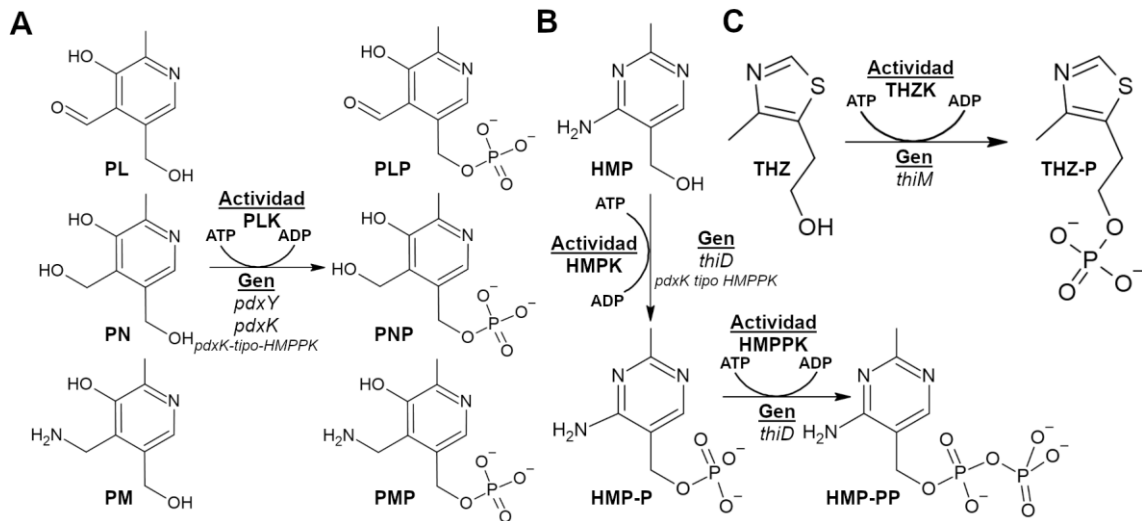
En bacterias, esta familia presenta las siguientes actividades enzimáticas: 4-metil-5- $\beta$ -hidroxietil tiazol quinasa (THZK, EC 2.7.1.50), 4-amino-5-hidroximetil-2-metilpirimidina quinasa (HMPK, EC 2.7.1.49), 4-amino-5-hidroximetil-2-metilpirimidina fosfato quinasa (HMPPK, EC 2.7.1.7) y piridoxal quinasa (PLK, EC 2.7.1.35) (Castro-Fernandez y col., 2014). Tanto las actividades THZK como HMPK y HMPPK participan en la vía de síntesis de tiamina, fosforilando 4-metil-5- $\beta$ -hidroxietil tiazol (THZ) a 4-metil-5- $\beta$ -hidroxietil tiazol fosfato (THZ-P) y 4-amino-5-hidroximetil-2-metilpirimidina (HMP) a 4-amino-5-hidroximetil-2-metilpirimidina fosfato (HMP-P) y posteriormente a 4-amino-5-hidroximetil-2-metilpirimidina pirofosfato (HMP-PP) respectivamente. Luego estos precursores, por acción de la actividad tiamina sintasa generan tiamina fosfato (TMP, forma activada de la tiamina, Figura 2, panel A). Por otro lado, la actividad piridoxal quinasa fosforila precursores de vitamina B6 (**Figura 2**, panel B) como piridoxal (PL), piridoxamina (PM) y piridoxina (PN), siendo el piridoxal fosfato (PLP) la forma activa de la vitamina B6 (Nodwell y col., 2014).



**Figura 2: Síntesis de vitamina B1 y B6.**

(A) Biosíntesis de vitamina B1 (tiamina monofosfato), modificado de Lünse y col., 2014. (B) Vía de rescate de piridoxal fosfato, modificado de Nodwell y col., 2014.

A continuación, se muestran en un esquema las reacciones que son catalizadas por estas actividades enzimáticas en bacterias (**Figura 3**), donde se observa que la fosforilación de HMP a HMP-P puede ser realizada por dos enzimas independientes (gen *thiD* y *pdxK*-tipo-HMPPK), existiendo una redundancia funcional.



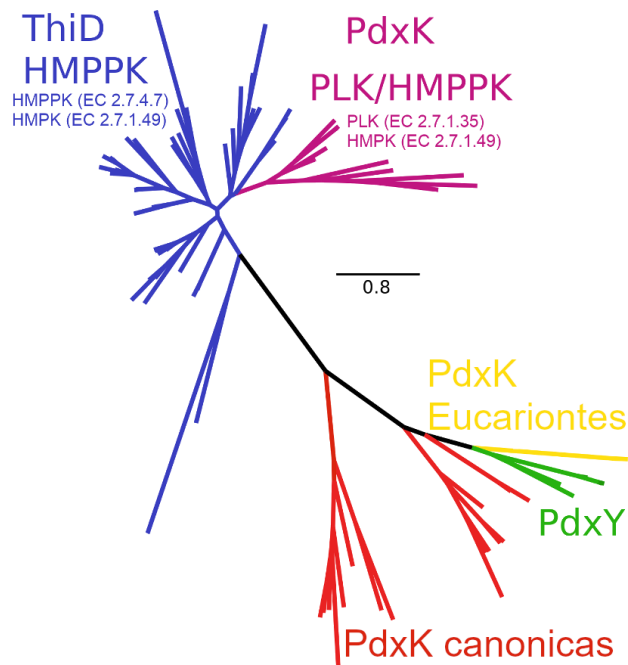
**Figura 3: Reacciones enzimáticas catalizadas por las quinasas de vitaminas dependientes de ATP.**

**(A)** Síntesis de PLP. **(B)** Síntesis de HMP-PP. **(C)** Síntesis de THZ-P.

La redundancia funcional presente en las enzimas de esta familia ha sido un desafío para la clasificación y nomenclatura de estas proteínas, lo que ha causado una mala anotación de éstas en distintas bases de datos. En bacterias, la actividad THZK está codificada por el gen *thiM*, mientras que tanto la actividad HMPK como la HMPPK están presentes en la misma enzima codificada por el gen *thiD*, enzima que lleva a cabo ambas fosforilaciones sobre el HMP de manera consecutiva a expensas de dos moléculas de ATP. Por otra parte, en el caso de la actividad PLK, se han identificado tres genes homólogos cuyos productos presentan actividad piridoxal quinasa, correspondientes a

los genes *pdxY*, *pdxK*-canónico y *pdxK*-tipo-HMPPK (proteína que nos referiremos como PdxK PLK/HMPPK de ahora en adelante).

La relación evolutiva de los productos génicos con actividades PLK y HMPK/HMPPK se ha estudiado mediante reconstrucción filogenética por máxima verosimilitud (**Figura 4**), la cual muestra que las enzimas codificadas por los genes *pdxY* y *pdxK*-canónico serían homólogos, presentando una mayor similitud de secuencia entre ellas, y una menor identidad de secuencia con el producto génico de las *pdxK*-tipo-HMPPK, las cuales serían homólogos cercanos del producto génico del gen *thiD*. Por lo tanto, los genes *pdxK*-tipo-HMPPK serían homólogos muy lejanos de los genes *pdxK*-canónicos (PLK clásicas) habiendo tenido un origen evolutivo distinto, pero con una convergencia en la función de fosforilación de piridoxal (Cea y col., 2020).



**Figura 4: Filogenia esquemática de quinasas de vitaminas de ATP.**

Árbol filogenético por máxima verosimilitud no enraizado de quinasas de vitaminas de ATP, las ramas se encuentran coloreadas de la siguiente manera: ThiD-HMPPK en azul, PdxK-PLK/HMPPK en púrpura, PdxY PLK en amarillo, PdxK PLK canónicas en verde y PdxK eucariontes en rojo. Las actividades enzimáticas son indicadas respecto sus números EC. Modificado de (Cea y col., 2020).

Para esta familia de quinasas se ha descrito un motivo adicional relacionado con la unión del metal divalente (DPV) (Ramírez-Sarmiento y col., 2020) junto a una modificación del residuo catalítico presente en la superfamilia Riboquinasa, encontrándose el motivo como GXGC en vez de GXGD en las enzimas ThiD-HMPPK (Castro-Fernandez y col., 2014) lo que implicaría que el residuo catalítico (base catalítica) correspondería a una cisteína en vez de un aspartato. Dentro de las quinasas de vitaminas dependientes de ATP, se ha visto que las THZK presentan una menor similitud de secuencia con el resto de la familia proponiéndose que estas enzimas serían las más similares a la proteína ancestral que derivó en las quinasas de vitaminas dependientes de ATP (Newman y col., 2006b). Por otro lado, las PLK presentan un mayor grado de similitud con las HMPPK, pero en este proyecto de tesis nos centraremos en la rama formada por las ThiD-HMPPK y las PdxK-PLK/HMPPK (**Figura 4**) grupo que presenta una mayor similitud de secuencia entre ellas.

### **1.3. Enzimas piridoxal quinasas tipo hidroximetil pirimidina quinasa bifuncionales (PdxK-PLK/HMPPK)**

Las enzimas con actividad piridoxal quinasa (PLK) catalizan la fosforilación de piridoxal (PL) a piridoxal fosfato (PLP) y como se mencionó en el párrafo anterior, la actividad PLK en bacterias presenta redundancia génica, encontrándose presente en tres enzimas de esta familia, codificadas por los genes *pdxY*, *pdxK-canónico* y *pdxK-tipo-HMPPK*. Estas enzimas presentan homología a nivel estructural y de secuencia, por lo que en algún momento durante la evolución habrían tenido algún ancestro en común.

Como se mencionó antes, en esta tesis nos centraremos en las PdxK-PLK/HMPPK codificadas por el gen *pdxK-tipo-HMPPK*, las que habrían divergido a partir del gen *thiD* y solo estarían presentes en bacterias Gram positivas según lo descrito por Cea y col.,

2020. Las enzimas PdxK PLK/HMPPK han sido estudiadas previamente, habiéndose reportado caracterizaciones cinéticas y estructuras cristalográficas de las enzimas de *Bacillus subtilis* (BsPLK/HMPPK) y *Staphylococcus aureus* (SaPLK/HMPPK) en presencia de distintos ligandos (**Tabla suplementaria 1**).

La información cristalográfica reportada para las enzimas de *Bacillus subtilis* y *Staphylococcus aureus* muestra que estas enzimas presentan el característico plegamiento tipo Rossmann y la ausencia del dominio menor, de acuerdo con lo reportado para la familia de las quinasas de vitaminas dependientes de ATP de bacteria (**Figura 5**). Además, en el trabajo de Nodwell y col., 2014, se reportó que SaPLK/HMPPK fosforila piridoxal a través de un mecanismo que involucra la formación de un hemitioacetal entre una cisteína del sitio activo y el piridoxal, estando dicha cisteína conservada en el grupo de las PdxK-PLK/HMPPK. Además, el bolsillo de unión a ligandos de las PdxK-PLK/HMPPK estaría resguardado por una horquilla ( $\beta 3$ - $\beta 4$ ) que cumpliría una función de tapa del sitio activo. Esta horquilla en las PdxK-PLK/HMPPK presentan una inserción de dos residuos, la cual está presente en todos sus miembros, siendo un rasgo que las distinguen de las ThiD-HMPPK que presentarían una versión más corta de esta horquilla. Asimismo, en el trabajo de Nodwell y col. se ha reportado que SaPLK/HMPPK puede catalizar la fosforilación tanto de PL como HMP realizando una única fosforilación en ambos sustratos generando piridoxal fosfato (PLP) y 4-amino-5-hidroximetil-2-metilpirimidina fosfato (HMP-P) respectivamente, al igual que otras enzimas PdxK-PLK/HMPPK (redundancia funcional con ThiD, **Figura 3**, panel B).



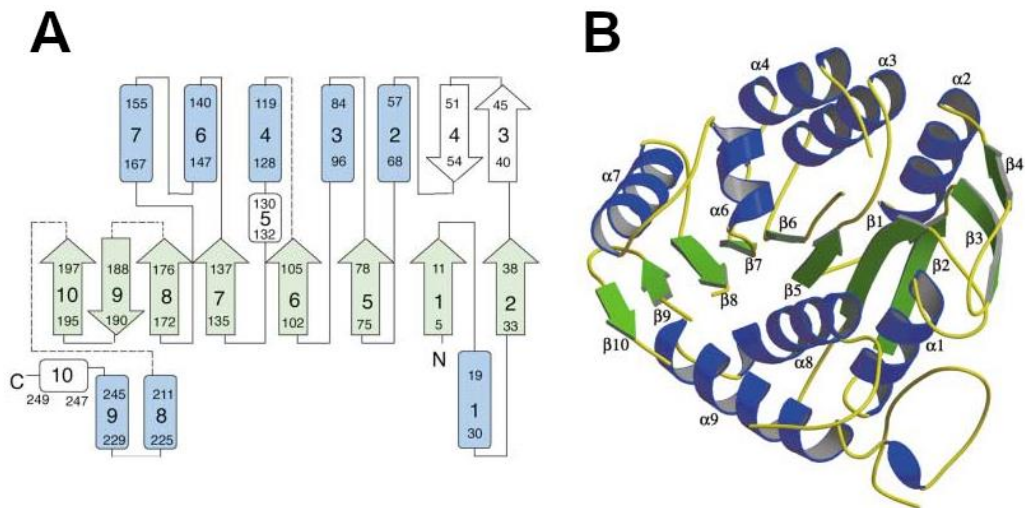
#### 1.4. Enzimas hidroximetil pirimidina fosfato quinasa (ThiD-HMPPK)

En bacterias, las 4-amino-5-hidroximetil-2-metilpirimidina fosfato quinasa (HMPPK) son enzimas codificadas por el gen *thiD* que catalizan la fosforilación de HMP a HMP-PP a expensas de dos moléculas de ATP, generando HMP-P como intermediario. Inicialmente, estas enzimas se describieron solo con actividad HMPK (fosforilación de HMP a HMP-P) pero en el trabajo de Park y col., 2004, se describió que la proteína codificada por el gen *yjbV* de *Bacillus subtilis* (la cual por similitud de secuencia en realidad corresponde a una ThiD-HMPPK) es una enzima bifuncional, con actividad HMPK y con actividad 4-amino-5-hidroximetil-2-metilpirimidina fosfato quinasa (HMP-P quinasa o HMPPK, EC 2.7.4.7). Esto significa que la misma enzima es capaz de realizar 2 fosforilaciones sobre el HMP generando finalmente 4-amino-5-hidroximetil-2-metilpirimidina pirofosfato (HMP-PP) como producto. Esta característica es compartida con las enzimas de *Salmonella typhimurium* (StHMPPK) y *Thermus thermophilus* (TtHMPPK) codificadas por el gen *thiD*, donde se ha visto por ensayos de titulación de productos que consumen dos moles de ATP por mol de HMP suministrado (Cea y col., 2020). Además, los autores identificaron, por separación cromatográfica por intercambio iónico, los componentes de reacción, correspondientes a tres picos de elución, siendo: **1.** HMP (sustrato), **2.** HMP-P (producto primera fosforilación) y **3.** HMP-PP (producto segunda fosforilación), donde se ve que en el transcurso temporal de la reacción en una primera etapa hay un consumo de HMP concomitante con la generación de HMP-P, en una segunda etapa hay una acumulación de HMP-P con la aparición de HMP-PP y en una tercera etapa hay una desaparición de HMP, con un consumo de HMP-P y una acumulación de HMP-PP, lo que indicaría que las dos fosforilaciones son independientes y consecutivas. Además, las HMPPK son altamente específicas por HMP y no muestran

actividad con precursores de piridoxal fosfato (PLP) como el piridoxal (PL), piridoxamina (PM) o piridoxina (PN) (Cea y col., 2020).

La capacidad de fosforilar HMP-P a HMP-PP hace a las enzimas codificadas por el gen *thiD* (HMPPK) únicas dentro de la superfamilia de las riboquinasas, dado que fosforilan tanto HMP como su contraparte fosforilada (HMP-P), por lo cual, el grupo aceptor de fosforilo entre ambas reacciones es distinto, correspondiendo a un grupo hidroximetil ( $R-CH_2-OH$ ) en la primera fosforilación, y a un grupo metil-fosfato ( $R-CH_2-PO_4$ ) en la segunda reacción de fosforilación, lo que podría suponer un mecanismo de reacción distinto. Además, estas enzimas no presentan actividad por PL a diferencia de las PdxK-PLK/HMPPK que son sus homólogas cercanas capaces de catalizar una única fosforilación tanto en el PL como en el HMP. Esto es sumamente llamativo, dado que pequeñas modificaciones en los sitios activos de estas enzimas pueden llevar a que una enzima evolucione desde una bifuncionalidad a nivel de sustrato a una bifuncionalidad a nivel de mecanismo de catálisis o viceversa.

Para las ThiD-HMPPK hay información estructural, encontrándose disponibles las estructuras cristalográficas de *Salmonella typhimurium*, *Thermus thermophilus*, *Clostridium difficile* y *Acinetobacter baumannii*, principalmente en la forma de enzima APO. En el caso de estructuras cristalográficas co-cristalizadas con ligandos, solo está disponible la proteína de *Salmonella typhimurium* en complejo con HMP (**Tabla suplementaria 2**), por lo que no existe información estructural del complejo ternario HMP/Mg-ATP ni del complejo ternario HMP-P/Mg-ATP. Por otro lado, tampoco hay información estructural del complejo ternario HMP/Mg-ATP en las PdxK-PLK/HMPPK dado que solo se ha descrito completamente la unión de PL al bolsillo de unión a sustrato y los cambios en conformación de loops del sitio activo (Nodwell y col., 2014).



**Figura 5: Estructura tridimensional de la ThiD-HMPPK de *Salmonella typhimurium*.** (A) Topología de SthMPPK. (B) Estructura cristalográfica de SthMPPK (PDBID: 1JXH). Modificado de Cheng y cols., 2002.

En las estructuras cristalográficas de SthMPPK y TthMPPK (**Figura 5**), hay varios loops del bolsillo de unión que no se encuentran resueltos, pero en general se aprecia que presentarían la topología conservada para la familia, incluyendo las PdxK-PLK/HMPPK. Estos loops no resueltos, se debe a la carencia de densidad electrónica, por lo que se sugiere que presentarían un alto grado de desorden y movilidad, lo cual también se ha visto en la estructura cristalográfica de SaPLK/HMPPK en conformación APO. Sin embargo, algunos de estos loops se reordenarían cuando la molécula de nucleótido y/o piridoxal están unidas en el bolsillo de unión de ligandos al comparar estructuras para la misma proteína cristalizada en presencia de distintos ligandos (Nodwell y col., 2014). Debido a la alta movilidad de estos loops y su capacidad de reorganización, se ha sugerido la existencia de un mecanismo diferencial de unión a sustrato entre una fosforilación y la siguiente por la capacidad de reorganización que tendrían estos loops.

En el trabajo de Cheng y col., 2002, se propone que la actividad HMP-P quinasa de las HMPPK se debería a un mecanismo de unión diferente para el HMP y el HMP-P, el cual involucraría una reorganización de loops adyacentes a estos sustratos, aunque no

aportan evidencia experimental que valide dicha hipótesis. Además, los autores proponen que el residuo Lys176 sería fundamental para la actividad HMP-P quinasa, pero dicha afirmación tampoco fue evaluada experimentalmente.

Por otro lado, en el trabajo de Cea y col., 2020 se ha reportado una filogenia para las quinasas de vitaminas (**Figura 4**), en la cual se aprecia que las PLK/HMPPK, gen *pdxK-tipo-HMPPK* aparecen como una novedad evolutiva del grupo de las HMPPK (gen *thiD*) y por lo tanto se identifica un ancestro común. En dicho trabajo también se realizaron simulaciones de dinámica molecular del complejo ternario (MgATP-HMP) de las estructuras depositadas en la base de datos PDB para la ThiD-HMPPK de *S. typhimurium* (PDBID: 1JXI) y la PdxK-PLK/HMPPK de *S. aureus* (PDBID: 4C5J), donde el análisis de residuos del sitio activo y su conservación entre estas ThiD-HMPPK y PdxK-PLK/HMPPK muestran que el residuo Lys176 (numeración de *St*HMPK) es conservada, mostrando que esta lisina también podría tener un rol en la fosforilación de HMP/PL y no un rol específico en la fosforilación de HMP-P. Por otro lado, en este análisis se identificó una diferencia conservada de residuo Thr211 (en enzimas ThiD-HMPPK) que es reemplazado por alanina en las enzimas PdxK-PLK/HMPPK, por lo que este residuo de treonina sí podría ser importante y tener un rol específico para la actividad HMPPK, sumado a una casi nula conservación de la horquilla  $\beta$ 3- $\beta$ 4 entre ambos tipos de enzimas.

Por lo tanto, para responder la pregunta de cómo se llevaría a cabo la segunda fosforilación del HMP (fosforilación de HMP-P) en las HMPPK y a la vez comprender el rol que podrían tener los residuos Lys176 y Thr211 junto a la horquilla  $\beta$ 3- $\beta$ 4, este proyecto de tesis plantea identificar los determinantes estructurales para la fosforilación de HMP-P en las ThiD-HMPPK y trazar la historia evolutiva de la actividad HMP-P

quinasa (EC 2.7.4.7) en las quinasas de vitaminas dependientes de ATP de bacteria, determinando durante qué momento en la evolución de esta familia de quinasas surgió o se perdió esta actividad enzimática. Para lograr este objetivo se utilizará un enfoque estructural y evolutivo empleando cristalografía de proteínas y la reconstrucción de secuencias ancestrales (ASR), de modo de realizar un estudio vertical de evolución de proteínas (Schupfner y coll., 2020). Además, las proteínas ancestrales son un buen modelo de trabajo para realizar ensayos de cristalización por sus altos rendimientos de purificación y su alta estabilidad en diversas condiciones (Munoz y col., 2021). Por ello, se utilizará como objeto de estudio en esta tesis la enzima ancestral AncEnHMPPK la cual corresponde al último ancestro en común de todas las secuencias de ThiD-HMPPK del orden *Enterobacteriales*. Esta enzima ya ha sido reconstruida, sintetizada y purificada en nuestro laboratorio, presentando buenos rendimientos de purificación, además de que ha sido posible identificar condiciones de cristalización para la enzima en su forma APO de la cual ya hay depositada una estructura cristalográfica (PDBID: 7R8Z). Considerando estos antecedentes, se realizarán ensayos de cristalización de la enzima ancestral AncEnHMPPK en presencia de HMP o HMP-P y análogos no hidrolizables de ADP y ATP. La determinación de la estructura tridimensional de distintos complejos ternarios permitirá identificar residuos claves para ambas fosforilaciones, junto a esclarecer el rol de los residuos Lys176, Thr211 y la horquilla  $\beta 3$ - $\beta 4$  en la unión de ligandos, además de los posibles cambios conformacionales en el bolsillo de unión de estas enzimas entre la primera y segunda fosforilación de HMP.

Además, se realizará la inferencia filogenética y reconstrucción de secuencias ancestrales (ASR) para analizar las sustituciones claves que se habrían dado en el bolsillo de unión de las ThiD-HMPPK durante la evolución, las que serían fundamentales

para la pérdida de la segunda reacción de fosforilación del HMP y en la aparición de la actividad PLK hacia la rama de las PdxK-PLK/HMPPK en la familia de quinasas de vitaminas dependientes de ATP de Bacterias.

## **1.5. Hipótesis**

La actividad HMP-P quinasa es un rasgo ancestral de las ThiD-HMPPK, que desaparece concomitantemente con la aparición de la actividad PLK en las enzimas PdxK-PLK/HMPPK como consecuencia de mutaciones en la horquilla  $\beta 3$ - $\beta 4$  y de la mutación T211A.

## **1.6. Objetivos**

### **1.6.1. Objetivo general**

Caracterizar las principales sustituciones en la ruta evolutiva de enzimas ThiD-HMPPK hacia las enzimas PdxK-PLK/HMPPK en la familia de quinasas de vitaminas dependientes de ATP de bacterias para identificar los determinantes estructurales responsables de la actividad HMP-P quinasa en las enzimas ThiD-HMPPK.

### **1.6.2. Objetivos específicos**

1. Expresar, purificar y cristalizar la proteína ancestral AncEnHMPPK en distintos complejos ternario con HMP/HMP-P y análogos no hidrolizables de ADP o ATP.
2. Determinar estructuras cristalográficas de la proteína ancestral AncEnHMPPK en complejo ternario con HMP/HMP-P y análogos no hidrolizables de ADP o ATP e identificar residuos claves y específicos para la fosforilación de HMP-P.
3. Analizar las sustituciones claves en el bolsillo de unión a sustratos para la aparición de la actividad PLK y desaparición de la actividad HMPPK en la trayectoria evolutiva desde las ThiD-HMPPK hacia las PdxK-PLK/HMPPK.

## **2. Materiales y Métodos**

### **2.1. Materiales**

#### **2.1.1. Vectores**

Como vector de expresión se utilizó el plásmido pET-TEV-15b (NOVAGEN), derivado del plásmido pET-15b el que incluye un marcador de selección por resistencia al antibiótico ampicilina. El gen de la proteína ancestral AncEnHMPPK fue sintetizado la compañía comercial (GenScript USA Inc., Piscataway, NJ, Estados Unidos). La secuencia codificante (872 pb) se posiciono entre los sitios de restricción NcoI y BamHI, incluyendo un tag de poli histidina (Hexa péptido) en el extremo amino terminal y un sitio de corte reconocido por la proteasa del virus del tabaco (TEV) entre el tag y el primer aminoácido de la secuencia.

#### **2.1.2. Medios de cultivos**

El medio Luria-Bertani (LB) se preparó utilizando 10 g/L de triptona, 5 g/L de extracto de levadura y 10 g/L de NaCl. Se aforó al volumen requerido con agua destilada y se autoclavó.

Para el medio sólido Luria-Bertani, empleado para seleccionar las transformantes, se utilizó medio LB suplementado con 15 g/L de agar, aforándose a un volumen de 400 mililitros empleando agua destilada para luego autoclavar, tras lo cual se enfrió a 50°C y se suplementó con kanamicina o ampicilina según correspondía.

#### **2.1.3. Soluciones generales**

Solución tampón *TG-SDS 10X*: Un litro de solución contiene 30,2 g de Tris, 11,4 ml de ácido acético glacial y 10,0 g de dodecilsulfato de sodio (SDS), previo aforar se ajusta a pH 8,4 a temperatura ambiente.

Solución tampón *TAE 50X*: Un litro de solución contiene 30,2 g de Tris, 11,4 ml de ácido acético glacial y 10,0 g de ácido etilendiaminotetraacético (EDTA), previo aforar se ajusta a pH 8,4 a temperatura ambiente.

#### 2.1.4. Reactivos químicos generales

**Tabla 1: Reactivos químicos utilizados.**

| Proveedor                    | Insumos  | Grado/Código  |
|------------------------------|--|---|
| Biorad <sup>1</sup>          | Bradford<br>Gel Filtration Standard  | 5000002<br>1511901  |
| Enamine Ltd <sup>2</sup>     | HMP-P  | ≥95% (HPLC)   |
| Finetech <sup>3</sup>        | HMP  | FT-0647258  |
| GE Healthcare <sup>4</sup>   | HisTrap HP<br>HiPrep 16/60 Sephacryl S-<br>200 HR<br>Superdex 75 Increase 10/300<br>GL | 175248<br>17116601<br>29148721  |
| Jena Bioscience <sup>5</sup> | ADP <sub>βS</sub><br>AMP-PCP<br>AMP-PNP  | ≥85% (HPLC)<br>≥95% (HPLC)<br>≥95% (HPLC)   |
| Merck <sup>6</sup>           | ATP <sub>γS</sub><br>Etanol<br>NaCl<br>MgCl <sub>2</sub><br>Tryptona                   | ≥90% (HPLC)<br>Para cromatografía, >99%<br>Para análisis, >99,5%<br>Para análisis, >99% |
| Sigma-Aldrich <sup>7</sup>   | 2-mercaptoetanol<br>HEPES  | >99%<br>>99,5% (Titration)  |
| Winkler <sup>8</sup>         | Extracto de levadura<br>GlicinaTris Base<br>AccuRuler Prestained Protein<br>Agar       | >99%<br>>98,5%<br>>99,5%  |

<sup>1</sup> Biorad, Hercules, California, USA

<sup>2</sup> Enamine Ltd, Chervonotkatska St., Kyiv, Ukraine

<sup>3</sup> Fine Tech Ind Ltd, <http://www.finetechnology-ind.com>

<sup>4</sup> GE Healthcare, Westborough, Massachusetts, USA

<sup>5</sup> Jena Bioscience, Löbstedter Str, Jena, Deutschland

<sup>6</sup> Merck, Kenilworth, NJ, Estados Unidos

<sup>7</sup> Sigma-Aldrich Corp., St. Louis, MO, Estados Unidos

<sup>8</sup> Winkler, Valle Grande, Lampa, Santiago de Chile



## **2.2. Métodos**

### **2.2.1. Cuantificación de proteínas**

La cuantificación de proteínas se realizó de dos maneras: **1.** Para mezclas heterogéneas se empleó el método de Bradford (Bradford y col., 1976) en un volumen de 1 mL, empleando 0,2 mL de una solución comercial de Coomassie G-250 (BioRAD) y de acuerdo con el coeficiente de extinción determinado experimentalmente a partir de una curva de calibración preparada con albúmina de suero bovino (BSA). **2.** Para soluciones de proteínas puras, la cuantificación de proteínas se realizó a partir de la ley de Lambert-Beer, empleando la absorbancia a 280 nm de la muestra y el coeficiente de extinción molar calculado teóricamente a partir del servidor web ProtParam de Expasy (Gasteiger y col., 2005). Para la proteína AncEnHMPPK el coeficiente de extinción molar fue de  $29575 \text{ M}^{-1}\text{cm}^{-1}$ , y de la misma herramienta se obtuvo el peso molecular teórico de 30.669 Da, esto considerando el Hexa péptido de histidina y el sitio de corte TEV (**2.1.1**).

### **2.2.2. Electroforesis de proteínas en condiciones desnaturalantes**

Las muestras se prepararon en tampón de carga (SDS 1%, glicerol 10%, Tris-HCl pH 6,8 10 mM, 2-mercaptoetanol 1 mM, y azul de bromofenol) y se calentaron a 95°C por 10 minutos. La electroforesis se realizó en un gel de poliacrilamida al 12% a un voltaje constante de 150V en presencia de solución tampón TG-SDS 1X. Una vez finalizada la electroforesis se retiró el gel de la cámara y se dejó tiñendo por toda la noche en solución azul de Coomassie al 0,25 % p/v, para luego ser desteñido en una solución compuesta por metanol 40%, ácido acético 40% y agua.

### **2.2.3. Purificación de AncEnHMPPK**

Se creció un cultivo de bacterias transformadas con el plásmido respectivo, en 50 mL de medio Luria Bertani (LB) con 100 µg/mL de ampicilina durante la noche, para

posteriormente inocular 2 L de medio LB con 100 µg/mL de ampicilina manteniendo el cultivo a 37°C hasta alcanzar una OD<sub>600</sub> de 0,6. La expresión de proteína se indujo con 1 mM de IPTG durante toda la noche a 30°C. Las bacterias fueron centrifugadas por 10 minutos a 6.130 g y resuspendidas en 50 mL de solución tampón para IMAC A (solución tampón A, Tris-HCl pH 7,8 50 mM, NaCl 500 mM, MgCl<sub>2</sub> 5 mM, imidazol 20 mM, 2-mercaptoetanol 5 mM), suplementado con PMSF 1 mM. Posteriormente, las bacterias resuspendidas se sonicaron con un régimen de 18 pulsos de 30 segundos de duración, con un 40% de amplitud y descansos de 30 segundos (Digital Sonifier 450, Branson). A continuación, el lisado se centrifugó a 38.360 g por 10 minutos, recuperándose la fracción soluble, la cual fue filtrada usando filtros para jeringa de 0,45 µm para posteriormente cargar en una columna de afinidad por níquel (HisTrap de 5 mL GE Healthcare), empleando una bomba peristáltica Bio rad con un flujo de 3 mL/min. Tras cargar la proteína, se realizó un lavado de 50 mL con solución tampón A. Posteriormente, la elución de la proteína se realizó con un gradiente de 20 a 500 mM de imidazol, usando un flujo de 3 mL/min y se colectaron las fracciones del del pico de elución empelando un sistema de purificación ÄKTA Prime (GE Healthcare). Tras la elución, se dializó la proteína en solución tampón para cristalización (HEPES pH 8,0 50 mM, NaCl 150 mM, 2-mercaptoetanol 5 mM) suplementado con 10% v/v glicerol. Posteriormente tras la diálisis el pool de fracciones se suplemento con 10% v/v de glicerol adicional y se prepararon alícuotas de 1,5mL, las cuales se congelaron con nitrógeno líquido para luego almacenar a -80°C. El proceso de purificación se evaluó mediante electroforesis de acrilamida en condiciones desnaturantes (SDS-PAGE).

#### **2.2.4. Determinación del estado oligomérico de AncEnHMPPK mediante HPLC**

La determinación de la masa molar de AncEnHMPPK se realizó mediante cromatografía de exclusión molecular por HPLC empleando una columna Superdex 75 Increase 10/300 GL previamente equilibrada en solución tampón para cristalización en el equipo HPLC Jasco equipado con una bomba PU-4180, un autosampler AS-4050, un detector UV/Visible UV-4070 y una interfase LC-NetII/ADC. La elución de la proteína se realizó mediante un flujo isocrático de 1 mL/min. La determinación de los volúmenes de elución se realizó mediante el “*software*” del fabricante del equipo (ChromNAV Control Center), mientras que la determinación de la masa molar se realizó interpolando el volumen de elución experimental a una curva estándar realizada previamente bajo las mismas condiciones experimentales utilizando un estándar de filtración en gel de marca Biorad.

#### **2.2.5. Determinación del estado oligomérico de AncEnHMPPK mediante SAXS**

Los ensayos de SAXS se realizaron en la línea de rayos X a bajo ángulo SAXSpoint 2.0 de Anton Paar (Anton Para, Strasse 20, Graz, Austria) equipado con un detector Eiger R 1M y una fuente de rayos Primux 100 Copper, ubicado en la Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile. Los experimentos se realizaron a 25°C y una presión de 1,98 mbar, se empleó una distancia al detector de 550 mm y un “*beam stopper*” de 1,5 mm. Realizándose mediciones de 30 minutos, dentro de las cuales se tomaron capturas cada cinco minutos (seis capturas por medición), evaluándose un total de 5 muestras (agua, solución tampón y tres concentraciones ascendentes de proteína). El tratamiento de los patrones de dispersión (normalización e integración) se realizó mediante el “*software*” del fabricante del equipo, mientras que el análisis del estado de

agregación, masa molar y calidad de los datos se realizó mediante el software ATSAS 3.0 (Manalastas-Cantos y col., 2021).

#### **2.2.6. Separación de poblaciones homogéneas y cristalografía de proteínas para la realización de ensayos de cristalización**

Para la realización de ensayos de cristalización, la separación de poblaciones homogéneas de AncEnHMPPK se realizó mediante una cromatografía de exclusión molecular (FPLC) empleando una columna HiPrep 16/60 Sephacryl S-200 HR previamente equilibrada en solución tampón para cristalización. La elución de la proteína se realizó mediante un flujo isocrático de 1 mL/min, empleando un sistema de purificación ÄKTA Prime (GE Healthcare). Las fracciones del pico mayoritario fueron colectadas y concentradas a 30 mg/mL para realizar los ensayos de cristalización.

Se identificaron condiciones favorables de cristalización mediante el método de difusión de vapor por gota sentada utilizando baterías comerciales de condiciones de cristalización y un equipo automatizado Crystal Gryphon 620-1000- 10 de Art Robins Instruments Rigaku. En total se evaluaron 9 condiciones de ligandos (indicadas en **Tabla 2**) utilizando 2 kits comerciales de cristalización (INDEX HT, Hampton Research, Journey Aliso Viejo, CA USA y JCSG plus, Molecular Dimensions, Trust Drive, Suite B, Holland, OH 43528 USA).

**Tabla 2: Condiciones de cristalización AncEnHMPPK.**

| Ronda cristalización | Gota | Concentración de proteína | Ligandos                                       |
|----------------------|------|---------------------------|--|
| 1                    | 1    | 25 mg/mL                  | HMP 1mM, ADP $\beta$ S 1mM, MgCl $_2$ 3mM      |
|                      | 2    | 15 mg/mL                  | HMP 1mM, AMP-PCP 1mM, MgCl $_2$ 3mM            |
|                      | 3    | 25 mg/mL                  | HMP 1mM, AMP-PCP 1mM, MgCl $_2$ 3mM            |
| 2                    | 1    | 25 mg/mL                  | HMP-P 1mM, ATP $\gamma$ S 10mM, MgCl $_2$ 15mM |
|                      | 2    | 25 mg/mL                  | HMP-P 1mM, AMP-PNP 10mM, MgCl $_2$ 15mM        |
|                      | 3    | 25 mg/mL                  | HMP-P 1mM, AMP-PCP 10mM, MgCl $_2$ 15mM        |
|                      | 4    | 23,3 mg/mL                | HMP 1mM, AMP-PNP 1mM, MgCl $_2$ 6mM            |
|                      | 5    | 23,3 mg/mL                | HMP 1mM, AMP-PCP 1mM, MgCl $_2$ 6mM            |
|                      | 6    | 23,3 mg/mL                | HMP 1mM, ATP $\gamma$ S 1mM, MgCl $_2$ 6mM     |

Los cristales se crecieron al menos 72 horas a 18°C. El almacenamiento de los cristales se realizó pasando los cristales de proteína de interés en soluciones crio-protectoras (solución madre de cristalización suplementada con Glicerol o PEG400 según el caso) y se montaron en “cryo-loops” los cuales se preservaron directamente en nitrógeno líquido para su posterior difracción.

Para algunos cristales de la “ronda 1” se realizaron ensayos de “soaking” que constan en reemplazar el ligando del cristal, embebiendo este en solución madre de cristalización suplementada con el ligando a incorporar más los co-ligandos en caso de ser necesario. Esto se efectuó para reemplazar el ligando HMP por HMP-P, tras lo cual se efectuó el proceso de almacenamiento descrito en el párrafo anterior.

### 2.2.7. Difracción de rayos X de cristales de AncEnHMPPK y análisis de datos

La colecta de los datos de difracción de rayos X para los cristales de AncEnHMPPK se llevaron a cabo en dos instancias, en dos fuentes de rayos X distintas: **1.** En la línea 23-B del APS empleando un detector Eiger 2M, en el Laboratorio Nacional de Argonne, Chicago, USA. **2.** En la línea MANACÁ del sincrotrón Sirius empleando un detector PILATUS M2, en el Laboratorio Brasileño de Luz Sincrotrón (LNLS), São Paulo, Brasil.

Los datos de difracción de rayos X fueron indexados, integrados y escalados utilizando el “software” XDS (Kabsch, 2010) y fue fusionada utilizando Aimless de la plataforma CCP4 (Evans & Murshudov, 2013). Las fases fueron resueltas por reemplazo molecular con Molrep (Vagin & Teplyakov, 2010). Para los datos obtenidos del Laboratorio Nacional de Argonne, se utilizó la estructura de la HMPPK de *Salmonella typhimurium* en complejo con HMP (PDB 1JXI, Cheng y col., 2002), para encontrar las fases aproximadas de los datos cristalográficos. Para los datos obtenidos en el LNLS se utilizó una estructura de AncEnHMPPK co-cristalizada en complejo ternario con HMP y Mg-AMP-PCP para encontrar la solución de las fases. El refinamiento de las estructuras se realizó de manera iterativa mediante COOT (Emsley & Cowtan, 2004) y Phenix (Adams y col., 2010). Las moléculas de ligandos, iones y agua fueron colocadas usando COOT y refinadas con restricciones generadas en ReadySet/eLBOW en Phenix. La validación de la estructura fue realizada con el servidor web Molprobit (Williams, y col., 2018).

#### **2.2.8. Reconstrucción de los complejos Mg-ATP de las estructuras cristalográficas en presencia de HMP y HMP-P**

Se reconstruyeron los complejos Mg-ATP para las estructuras cristalográficas de AncEn HMPPK en presencia de HMP y HMP-P, utilizando COOT. Para esto se utilizó la estructura de AncEnHMPPK co-cristalizada en complejo ternario con HMP y Mg-AMP-PCP como molde. Para el complejo ternario HMP/Mg-ATP se utilizó la cadena B de la estructura cristalográfica de AncEnHMPPK co-cristalizada con HMP y Mg-AMP-PCP, y se reemplazó la molécula de AMP-PCP por ATP, refinándose de manera local la molécula de ATP dentro de la densidad electrónica, conservando la geometría y distancias de coordinación del ión magnesio, además se conservaron las aguas de cristalización a un radio de 4,0 Å de la cadena polipeptídica.

En el caso del complejo ternario HMP-P/Mg-ATP, se superpusieron estructuralmente en COOT: **1.** La cadena B de la estructura cristalográfica de AncEnHMPPK co-cristalizada con HMP y Mg-AMP-PCP; **2.** La cadena B de la estructura cristalográfica de AncEnHMPPK co-cristalizada con HMP-P y ATP<sub>γS</sub> a 2,20 Å; **3.** La cadena B de la estructura cristalográfica de AncEnHMPPK co-cristalizada con HMP-P y ATP<sub>γS</sub> a 2,53 Å. Para la reconstrucción del complejo Mg-ATP, se utilizó el ión magnesio con sus aguas de coordinación de la estructura (1), con la molécula de ATP<sub>γS</sub> de la estructura (3) que se encontraba en similar conformación que la molécula de AMP-PCP de (1), dejándose la cadena polipeptídica, la molécula de ligando HMP-P y las aguas de cristalización a un radio de 4,0 Å de la estructura (2), finalmente se reemplazó la molécula de ATP<sub>γS</sub> por ATP, refinándose de manera local la molécula de ATP dentro de la densidad electrónica, conservando la geometría y distancias de coordinación del ión magnesio. La posterior visualización y análisis de los complejo se realizó mediante Pymol.

### **2.2.9. Recolección de secuencias de quinasas de vitaminas dependientes de ATP de bacteria**

Utilizando como templado las secuencias de los PDB 4C5L y 2I5B se realizaron psi-blast a partir de la base de datos no redundante de proteínas del NCBI, empleando una matriz BLOSUM45 en parámetros de puntuación y excluyendo secuencias provenientes de muestras ambientales. Se efectuaron 3 iteraciones de la matriz de puntuación, posteriormente las secuencias fueron filtradas y agrupadas por identidad utilizando la herramienta web CD-hit (Huang y col., 2010) con un corte del 95% de identidad para eliminar secuencias con alta similitud.

### **2.2.10. Generación del perfil evolutivo e Inferencia filogenética de las quinasas de vitaminas dependientes de ATP de bacteria**

Se realizó un perfil evolutivo de las quinasas de vitaminas dependientes de ATP, usando como referencia un alineamiento múltiple de secuencia de ThiD-HMPPK, junto las nuevas secuencias colectadas a partir de la base de datos de NCBI como se detalla en el apartado anterior. Primero se realizó un alineamiento basado en estructura utilizando la herramienta “stamp” (Russell & Barton, 1992) en “multiseq” del “software” VMD (Roberts y col., 2006), usando como templatado las estructuras de la HMPPK de *Salmonella typhimurium* (PDBID:1JXI), la HMPPK de *Thermus thermophilus* (PDBID:1UB0), la PdxK-PLK/HMPPK de *Bacillus subtilis* (PDBID:2I5B) y la PdxK-PLK/HMPPK de *Staphylococcus aureus* (PDBID:4C5L). Como segundo paso se alinearon las secuencias colectadas hacia el alineamiento estructural descrito en el párrafo anterior mediante ClustalW (Thompson y col., 1994), finalmente se realizó un alineamiento entre grupos del alineamiento resultante con el de las ThiD-HMPPK mediante ClustalW. De este último alineamiento con VMD, se tomó un conjunto no redundante de secuencias con un corte del 85% de identidad, incorporando en su totalidad las secuencias de ThiD-HMPPK del alineamiento base.

A partir del conjunto no redundante del alineamiento múltiple de secuencias entre las ThiD-HMPPK y las PdxK-PLK/HMPPK, se eligió el modelo evolutivo que explica mejor el conjunto de datos utilizando el programa ProtTest 3.0 y discriminando según el criterio de información de Akaike corregido (Darriba y col., 2011). El árbol filogenético fue construido por método de máxima verosimilitud empleando el programa PhyML (Guindon y col., 2010) utilizando un modelo de sustitución aminoacídica LG con corrección por sitios invariantes y con variación entre sitios de acuerdo a una distribución gamma, la



robustez de los nodos se evaluó mediante una prueba de razones de verosimilitud basado en  $\chi^2$ .

### **2.2.11. Reconstrucción de secuencias ancestrales**

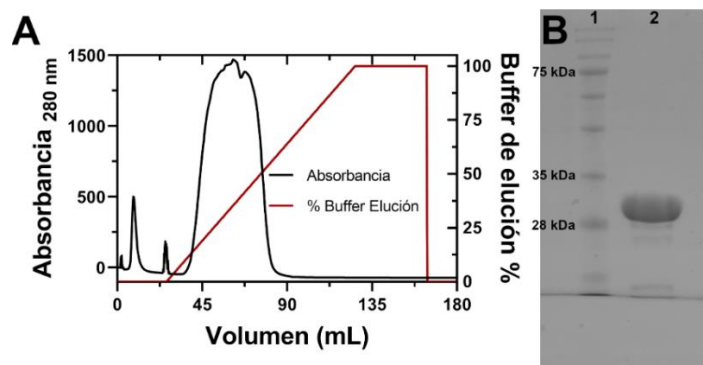
Para la inferencia de secuencias ancestrales, el árbol por máxima verosimilitud inferido con PhyML se analizó en el visor Figtree y se enraizó con las secuencias de ThiD-HMPPK de bacterias del orden *Deinococcus-Thermus*, el cual es considerado uno de los filos más profundos y ancestrales en bacterias. Posteriormente, se realizó la reconstrucción de secuencias ancestrales por máxima verosimilitud en dos pasos, donde se seleccionaron los nodos de interés en la trayectoria evolutiva que tuvieran soporte de nodo mayor o igual al 80%. Primero, empleando el alineamiento múltiple de secuencias y la filogenia enraizada se estimó el aminoácido más probable para cada posición mediante el método de Bayes empírico con el programa PAML 4.4 (Yang, 2007) y utilizando la herramienta Lazarus de acuerdo con lo descrito por Hanson-Smith y col. (Hanson-Smith y col., 2010),.

Como segundo paso, se realizó la corrección de los espacios (o *gaps*) de las secuencias inferidas utilizando el programa Mesquite 3.61 (Maddison & Maddison, 2019). Para esto, el alineamiento de secuencias múltiples de aminoácidos se transformó en un alineamiento binario de caracteres mediante la herramienta BioEdit 7.2.5 (Hall, 2007), haciendo la equivalencia del carácter “1” a cualquier aminoácido en el alineamiento y el carácter “0” a un espacio o *gap*. La reconstrucción de los estados ancestrales para rasgos discretos se realizó por máxima verosimilitud empleando un modelo estocástico de evolución en base a la distribución de los caracteres en el alineamiento binario y a la filogenia enraizada previamente utilizada.

### 3. Resultados

#### 3.1. Expresión y purificación de la proteína ancestral AncEnHMPPK

La proteína ancestral AncEnHMPPK fue expresada en *E. coli* cepa BL21, utilizando el vector pET-TEV-15b. Se realizó un único paso de purificación en una columna de afinidad a ión níquel, tras lo cual las fracciones del pico mayor de la cromatografía (**Figura 6**, panel A) fueron colectadas y dializadas en tampón de cristalización suplementado con glicerol al 20% v/v para posteriormente ser almacenadas a -80°C. La pureza de la proteína se analizó mediante SDS-PAGE (**Figura 6**, panel B), obteniéndose 200 mg de proteína altamente pura (sobre el 90%) desde 2 litros de cultivo LB.



**Figura 6: Purificación de AncEnHMPPK.**

(A) Cromatograma de la elución de AncEnHMPPK en columna de afinidad de níquel. En el eje derecho se indica la concentración de imidazol empleada durante la cromatografía indicado como porcentaje, correspondiendo 0% a 20mM, mientras que el 100% corresponde a 500 mM de imidazol. (B) SDS-PAGE de la purificación de AncEnHMPPK. 1. Estándares de peso molecular, 2. AncEnHMPPK tras cromatografía de afinidad de níquel, se cargaron 15µg de proteína.

#### 3.2. Determinación del estado oligomérico de AncEnHMPPK

Dada la poca información en literatura sobre el estado de oligomerización de las proteínas ThiD-HMPPK y algunas inconsistencias para las enzimas PdxK-PLK/HMPPK como mezclas de monómero y dímero en solución (Newman y cols., 2006a, 2006b), analizamos el estado oligomérico del ancestro de las ThiD-HMPPK de *Enterobacteriales*

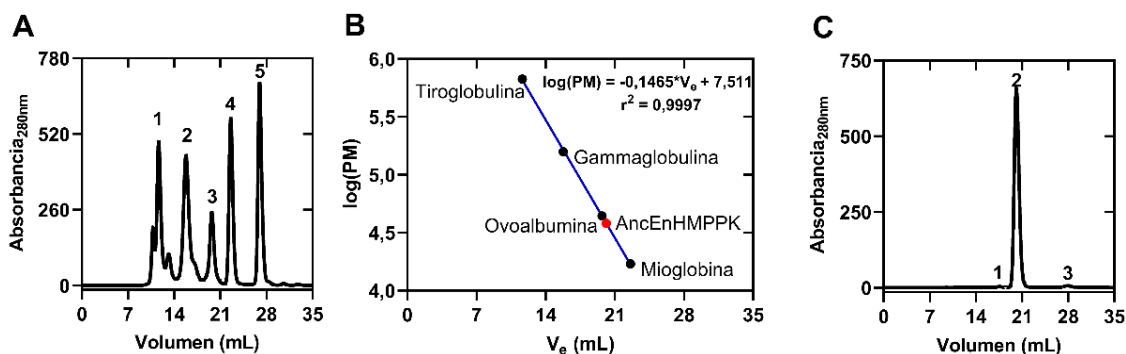
(AncEnHMPPK). Para ello, se realizó una separación cromatográfica por tamaño acoplada a HPLC, de donde se interpoló el volumen de elución experimental de AncEnHMPPK en una curva de calibración construida a partir de un estándar de tamaños para cromatografía de exclusión molecular, bajo las mismas condiciones experimentales, descritas en la sección de materiales y métodos. Los resultados de la cromatografía de exclusión molecular (**Figura 7**, panel A) muestran una correcta separación de las proteínas del estándar, apreciándose picos discretos de los cuales se determinó el volumen de elución para cada estándar (**Tabla suplementaria 3**). Posteriormente, se evaluó la correlación lineal entre el volumen de elución ( $V_e$ ) de las proteínas del estándar y el logaritmo de su masa molar ( $\log (PM)$ ) (**Figura 7**, panel B), tras lo cual se realizó un ajuste lineal determinando la ecuación de dicha recta.

Para el cromatograma de AncEnHMPPK se aprecia la presencia de un único pico de elución (**Figura 7**, panel C), pero al realizar el análisis del volumen de elución, el software fue capaz de detectar tres picos de elución, de los cuales el identificable visualmente constituye al 98% del área integrada de los tres picos de elución, por lo cual los otros dos no fueron considerados. El pico mayoritario de AncEnHMPPK tiene un  $V_e$  de 20,137 mL, lo que interpolando a la curva de calibración indica un  $\log (PM)$  de 4,581 y al aplicar antilogaritmo una masa molar de 38,1 kDa, la cual no coincide con el tamaño esperado para el monómero dado que se espera un valor teórico de 30,7 kDa de acuerdo con lo determinado a partir de la secuencia aminoacídica. Esto sugiere que la proteína puede tener una forma que muy diferente a una esfera, por lo que el monómero eluye a volúmenes mayores o que de acuerdo a lo propuesto por Newman y col., 2006b para BsPLK/HMPPK, la proteína podría estar en una mezcla de monómero dímero en

solución y que su cinética de equilibrio monómero-dímero sea más rápida que la cinética que toma el equilibrio de la separación por tamaño de la resina de exclusión molecular.

**Tabla 3: Volúmenes de elución para AncEnHMPPK.**

| Pico | $V_e$ (mL) | Área (%) |
|------|------------|----------|
| 1    | 17,522     | 0,401    |
| 2    | 20,137     | 98,477   |
| 3    | 28,005     | 1,152    |



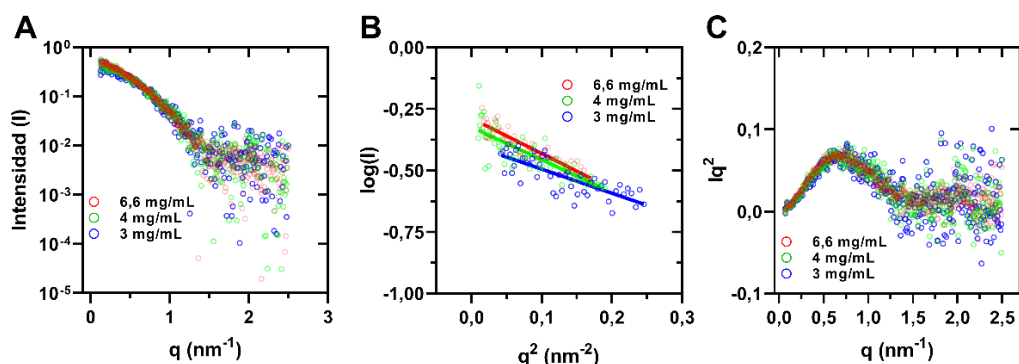
**Figura 7: Cromatogramas Superdex 75 Increase 10/300.**

(A) Cromatograma exclusión molecular para estándar Biorad. En el eje y se indica la absorbancia a 280 nm registrada respecto el volumen de elución (eje x). Los componentes del estándar se indican con números, correspondiendo: 1. Tiroglobulina, 2. Gammaglobulina, 3. Ovoalbúmina, 4. Mioglobina, 5. Vitamina B12. (B) Curva de calibración. En eje y se indica el logaritmo del peso molecular respecto el volumen de elución (eje x). (C) Cromatograma de elución para AncEnHMPPK. En el eje y se indica la absorbancia a 280 nm registrada respecto el volumen de elución (eje x).

Para analizar si la forma de AncEnHMPPK es la responsable de la tardía elución de la cromatografía de exclusión o si es un equilibrio dímero-monómero, se realizaron experimentos de dispersión de rayos X a bajo ángulo (SAXS, **Figura 8**, panel A y **Figura suplementaria 1**) a distintas concentraciones de proteína (3, 4 y 6,6 mg/mL). Si hay un equilibrio dímero-monómero que permite que las poblaciones cambien con el cambio de la concentración de proteína, la masa molar promedio estimada aumentará cuando se aumente la concentración de proteína.

El análisis de Guinier (**Figura 8**, panel B) muestra un comportamiento lineal de las soluciones de proteínas, lo que indica que no hay agregación de las muestras y tampoco interacción inter-partículas, mientras que el análisis de Kratky (**Figura 8**, panel C) muestra una curva tipo Gaussiana para las tres soluciones, lo que es consistente con una proteína compacta en solución no habiendo indicios de regiones intrínsecamente desordenadas.

Al analizar los radios de giro promedio ( $R_g$ ) y los estimados bayesianos de masas molares de las tres soluciones (**Tabla 4**), se evidencia un aumento del tamaño promedio de las partículas con el aumento de concentración, lo que es consistente con la hipótesis de un equilibrio dímero-monómero. Posteriormente, se calculó una curva de distribución de probabilidades de distancias a partir de la dispersión de la muestra de 6,6 mg/mL (**Figura 9**, panel B), muestra que presenta un tamaño coherente con un dímero. Esta distribución de distancias permite estimar modelos *ab initio* de la partícula (**Figura 9**, paneles C y D), siendo el modelo consenso muy coherente con la superposición de un dímero de la estructura cristalográfica de AncEnHMPPK, lo cual se aprecia en la correlación de los ajustes a los datos experimentales (**Figura 9**, panel A). En resumen, estos resultados muestran que la proteína ancestral en solución es una proteína globular compacta que se encuentra en un equilibrio dímero-monómero y que probablemente su cinética de equilibrio sea más rápida que el equilibrio de separación en la cromatografía de exclusión molecular, provocando que en dicha cromatografía el tiempo de retención sea intermedio al de un monómero y dímero.

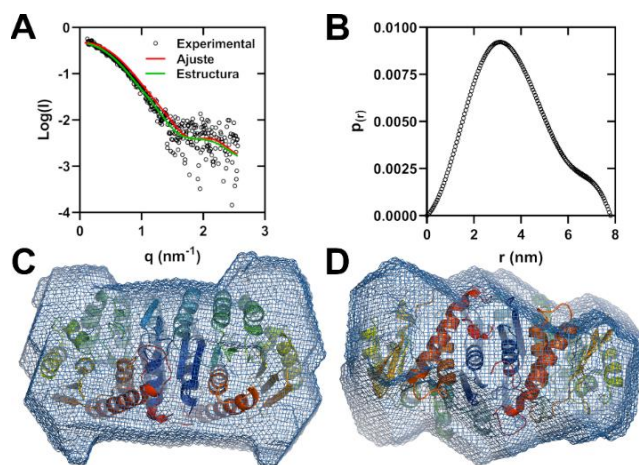


**Figura 8: Analisis por SAXS para AncEnHMPPK.**

(A) Grafico de intensidad respecto vector de dispersión. (B) Gráfico de Guinier. (C) Gráfico de Kratky.

**Tabla 4: Estimación bayesiana de masa molar de AncEnHMPPK mediante SAXS.**

| Concentración (mg/mL) | $R_g$ (Å)      | Masa molar promedio (kDa) | Probabilidad promedio (%) | Intervalo masa molar (kDa) | Probabilidad intervalo (%) |
|-----------------------|----------------|---------------------------|---------------------------|----------------------------|----------------------------|
| 3,0                   | $26,2 \pm 1,4$ | 40,2                      | 31,35                     | 37,3 – 54,9                | 91,01                      |
| 4,0                   | $29,7 \pm 1,2$ | 58,1                      | 31,40                     | 54,9 – 63,1                | 90,45                      |
| 6,6                   | $31,2 \pm 0,9$ | 62,3                      | 33,10                     | 54,9 – 67,9                | 93,08                      |



**Figura 9: Modelo *ab initio* de AncEnHMPPK en solución a una concentración de 6,6 mg/mL.**

(A) Grafico del logaritmo de la intensidad respecto el vector de dispersión, para datos experimentales, ajuste de acuerdo a distribución de probabilidades de distancia (línea roja) y dispersión teórica de la estructura cristalografica del dimerio (línea verde). (B) Distribución de probabilidad de distancias para datos de SAXS a 6,6 mg/mL. Se ajusto con un  $\alpha$  de 4,254,  $r_{min}$  de 0 nm,  $r_{max}$  de 7,80 nm,  $p(r_{min})$  de 0 y  $p(r_{max})$  de 0. (C-D) Se muestra la superficie del modelo *ab initio* como “*mesh*”, mientras que la estructura tridimensional del modelo se presenta como “*cartoon*”, cada monómero esta coloreado de acuerdo con un espectro de color azul a rojo en orientación N a C terminal. Se presentan dos vistas con rotaciones de 270° respecto al eje X.

### 3.3. Ensayos de co-cristalización de AncEnHMPPK con HMP, HMP-P y análogos no hidrolizables de ATP

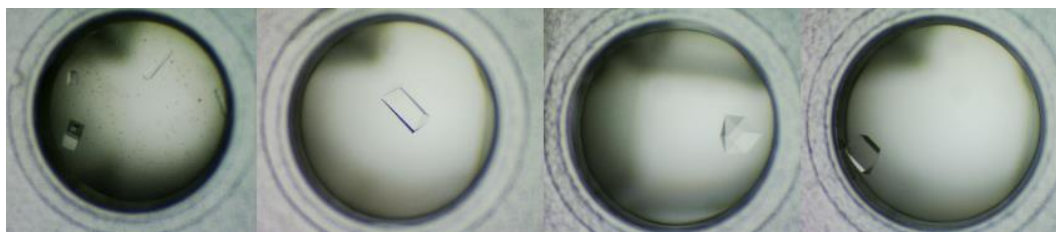
Con el objetivo de obtener estructuras cristalográficas de la proteína ancestral con los distintos sustratos e identificar posibles residuos críticos para la actividad HMPK y HMPPK se realizaron ensayos de cristalización a gran escala, mediante el método de difusión de vapor por gota sentada utilizando baterías comerciales de condiciones de cristalización, empleando una población homogénea tras separación por cromatografía de exclusión molecular y análogos no hidrolizables de nucleótidos (**Figura suplementaria 3** y **Figura suplementaria 4**). Se obtuvo una gran cantidad de condiciones de cristalización para AncEnHMPPK en las baterías comerciales INDEX HT (Hampton Research) y JCSG plus (Molecular Dimensions), teniendo cristales de proteína en presencia de todas las condiciones de ligando ensayadas (**Tabla 5**).

**Tabla 5: Condiciones de cristalización de AncEnHMPPK en presencia de ligandos.**

| Concentración de proteína (mg/mL) | Condición de ligandos   | Kit                 |
|-----------------------------------|---|---------------------|
| 15                                | HMP (1 mM), AMP-PCP (1 mM) MgCl <sub>2</sub> (3 mM)               | INDEX HT, JCSG Plus |
| 23,3                              | HMP (1 mM), AMP-PCP (1 mM) MgCl <sub>2</sub> (6 mM)               | INDEX HT, JCSG Plus |
|                                   | HMP (1 mM), AMP-PNP (1 mM) MgCl <sub>2</sub> (6 mM)               | INDEX HT, JCSG Plus |
| 25                                | HMP (1 mM), ATP <sub>γS</sub> (1 mM) MgCl <sub>2</sub> (6 mM)     | INDEX HT, JCSG Plus |
|                                   | HMP (1 mM), AMP-PCP (1 mM) MgCl <sub>2</sub> (3 mM)               | INDEX HT, JCSG Plus |
|                                   | HMP (1 mM), ADP <sub>βS</sub> (1 mM) MgCl <sub>2</sub> (3 mM)     | INDEX HT, JCSG Plus |
|                                   | HMP-P (1 mM), AMP-PCP (10 mM) MgCl <sub>2</sub> (15 mM)           | INDEX HT, JCSG Plus |
|                                   | HMP-P (1 mM), AMP-PNP (10 mM) MgCl <sub>2</sub> (15 mM)           | INDEX HT, JCSG Plus |
|                                   | HMP-P (1 mM), ATP <sub>γS</sub> (10 mM) MgCl <sub>2</sub> (15 mM) | INDEX HT, JCSG Plus |

Se obtuvieron cristales de distinta forma y tamaño, de los cuales se seleccionaron los cristales de mayor tamaño y forma regular (Ejemplos, **Figura 10**). Los cristales seleccionados para los experimentos de difracción fueron preservados en nitrógeno líquido utilizando una solución crio-protectora (solución madre de cristalización más agente como glicerol o PEG400) para evitar la formación de hielo el cual interfiere

posteriormente en el experimento de difracción. Los cristales almacenados fueron enviados a dos fuentes de rayos X distintas (sincrotrones Argonne-APS y LNLS-Sirius), donde se realizaron los experimentos de difracción según lo indicado en la sección de materiales y métodos.



**Figura 10: Cristalización AncEnHMPPK.**

Ejemplos de cristales obtenidos para AncEnHMPPK a partir de baterías comerciales mediante difusión de vapor por gota sentada en placas de 96 pocillos.

En la **Tabla 6** se presenta un resumen de los cristales difractados de los cuales se obtuvieron patrones de difracción procesables, y fue posible generar un modelo tridimensional para refinar. Cabe destacar que para cristales de la condición HMP/Mg-ADP<sub>βS</sub> y HMP/Mg-AMP-PCP se realizaron experimentos de “*soaking*” o remojo, que consta en la adición o reemplazo de ligandos en un cristal de proteína. En este caso, se buscó reemplazar el ligando HMP por HMP-P, para lo cual los cristales de proteína se incubaron en solución madre (o solución de cristalización) suplementada con el ligando a incorporar (HMP-P) y el coligando (análogo no hidrolizable de nucleótido) y se dejaron incubando al menos cinco minutos esperando lograr el reemplazo.



**Tabla 6: Condiciones de cristalización para AncEnHMPPK.**

| Kit   | Condición  | Ligandos  | Crio (% v/v)   |
|-------|--|---|----------------|
| JCSG  | CHES 100mM pH 9,5 20% p/v PEG 8000                                       | HMP Mg-AMP-PCP<br>("soaking" con HMP-P)           | 15% PEG<br>400 |
| INDEX | 0,8M Ácido succínico pH 7,0  | HMP Mg-ADP <sub>βS</sub><br>("soaking" con HMP-P) | 25% PEG<br>400 |
| JCSG  | 0,1 M Citrato de sodio pH 5,5 20% p/v PEG 300                            | HMP-P ATP <sub>γS</sub>                           | 15% Glicerol   |
| INDEX | 0,1 M TRIS pH 8,5 0,2 M NaCl 25% p/v PEG 3350                            | HMP-P Mg-AMP-PNP                                  | 25% Glicerol   |
| INDEX | 0,5% p/v MPEG 5000 0,8 M Sal de Seignette tetrahidrato 0,1 M Tris pH 8,5 | HMP-P Mg-AMP-PCP                                  | 25% Glicerol   |
| INDEX | 0,5% p/v MPEG 5000 0,8 M Sal de Seignette tetrahidrato 0,1 M Tris pH 8,5 | HMP-P Mg-AMP-PNP                                  | 25% Glicerol   |
| INDEX | 35% v/v Tacsimato pH 7,0   | HMP-P ATP <sub>γS</sub>                           | 25% Glicerol   |
| INDEX | 1% p/v MPEG 2000 1 M ácido succínico 0,1 M HEPES pH 7,0                  | HMP-P ATP <sub>γS</sub>                           | 25% Glicerol   |

### 3.4. Difracción de rayos X y análisis preliminar de los datos

Se realizaron dos rondas de difracción de rayos X, la primera en la línea 23-B del "Advanced Photon Source" (APS) del Laboratorio nacional Argonne, Chicago USA, donde se disponía de un detector Eiger 2M, y como estrategia de colecta de datos se usó una longitud de onda de 1,03 Å y un rango de oscilación de 0,1°; colectando un total 3600 cuadros (se presenta un patrón de difracción representativo en la **Figura suplementaria 5**). De los 17 cristales enviados se obtuvieron tres patrones de difracción analizables con una resolución en torno a 2,3 a 2,8 Å. De estos tres cristales, a dos se les había realizado ensayos de "soaking", el cual solo fue satisfactorio para la condición HMP-P/Mg-ADP<sub>βS</sub> donde se apreció densidad electrónica para el grupo metil fosfato de la molécula de HMP-P, mientras que en la condición HMP-P/Mg-AMP-PCP (2,8 Å) no se apreció densidad electrónica para el grupo metil fosfato de la molécula de HMP-P, por lo cual no se siguió con el refinamiento de esta estructura dado que se constaba de un conjunto de datos de difracción a mayor resolución para el complejo HMP/Mg-AMP-PCP (2,3 Å).

La segunda ronda de difracción se realizó en la línea MANACÁ del sincrotrón Sirius la que dispone de un detector PILATUS M2 y se encuentra en el Laboratorio Nacional de Luz Sincrotrón (LNLS), São Paulo, Brasil. Como estrategia de colecta de datos se usó una longitud de onda de 0,977 Å y un rango de oscilación de 0,1°, colectando un total 3600 cuadros. En dicha instancia se enviaron 30 cristales de los cuales se difractaron 14 (tras pruebas de difracción), obteniéndose finalmente 6 patrones de difracción analizables con una capacidad de resolución en torno a 2,2 a 3,0 Å.

La información de difracción de rayos X fue indexada, integrada y escalada utilizando XDS. Para todos los datos tratados, se determinó que el grupo espacial es P 21 21 21 y el rango de resolución final para los set de datos fue entre 2,2 a 3,0 Å teniendo en cuenta los valores de  $I/\sigma$  y los factores R (R-merge, R-meas y Rpim), los cuales se utilizan para evaluar la consistencia y calidad de la información obtenida del experimento de difracción. En la siguiente sección se muestran las estadísticas para cada conjunto de datos (Tabla 8 y Tabla 9).

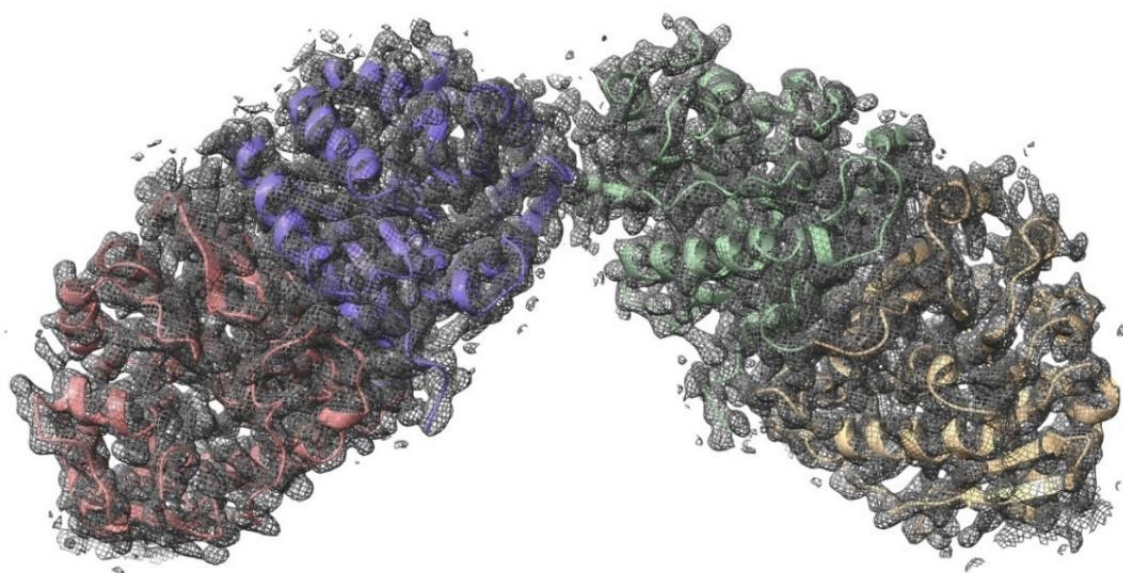
### **3.5. Reemplazo molecular y refinamiento**

Para los datos de cristalografía obtenidos en el APS la resolución de las fases se realizó mediante reemplazo molecular utilizando la estructura de la HMPPK de *Salmonella typhimurium* (StHMPPK) co-cristalizada con HMP en su sitio activo (PDBID: 1JXI, 2,64 Å). Para los datos obtenidos en la línea MANACÁ se utilizó la estructura de AncEnHMPPK en complejo ternario con HMP y Mg-AMP-PCP (2,3 Å) obtenida de la difracción en APS-Argonne para la estimación de las fases.

Durante el proceso de estimación de las fases se determinó el coeficiente de Matthews ( $V_M$ ) para todos los conjuntos de datos tratados. El  $V_M$  permite determinar la cantidad de macromoléculas por unidad asimétrica dado que establece una relación entre el volumen

de la celda unitaria y el peso molecular de la macromolécula, obteniendo un valor cuya probabilidad está dada por un rango observado empíricamente, que tiene el valor de resolución como información adicional (Kantardjieff & Rupp, 2003). En el caso de las estructuras de AncEnHMPPK, se utilizó la plataforma CCP4 para calcular el valor de  $V_M$ , el cual estuvo en un rango de 2,69 y 3,21 para cuatro moléculas de AncEnHMPPK en la unidad asimétrica, y de 2,15 a 2,57 para 5 moléculas con un porcentaje de solvente asociado de 42,53% y 61,69% respectivamente, con una probabilidad final en el intervalo de 12 a 27% para cuatro moléculas y de 41 a 57% para cinco moléculas de AncEnHMPPK tomando en cuenta las resoluciones de las estructuras. Por lo tanto, se esperaba observar entre cuatro a cinco moléculas de AncEnHMPPK en la unidad asimétrica. Al efectuar el reemplazo molecular, se observa la presencia de cuatro moléculas en la unidad asimétrica (**Figura 11**) para todas las estructuras.

Las estructuras resultantes se refinaron en un intervalo de resolución de 2,20 a 3,00 Å con valores de R-work/R-free de 0,18/0,20 y 0,20/0,21 para la mayor y menor resolución respectivamente. En las **Tabla 8** y **Tabla 9** se muestra la información del refinamiento y validación para las ocho estructuras determinadas, mientras que en la **Figura 11** se muestra el resultado de reemplazo molecular para una estructura representativa de AncEnHMPPK, donde se aprecian las cuatro monómeros por celda unitaria.



**Figura 11: Modelo representativo de los 4 monómeros en la unidad asimétrica de la estructura de AncEnHMPPK.**

Mapa 2mFo-DFc de densidad electrónica obtenido por reemplazo molecular representativo para AncEnHMPPK, contorneado a  $1\sigma$ , coloreado de color gris. Se identificaron cuatro cadenas principales de AncEnHMPPK representadas como “cartoon”, las cuales se colorearon de rojo, morado, verde y amarillo para su distinción.

**Tabla 7: Resumen de condiciones de cristalización de las estructuras refinadas de AncEnHMPPK.**

| Cristal                                | Resolución (Å) | Condición  | Crio         | Kit   |
|--|----------------|--|--------------|-------|
| AncEn/HMP-P/Mg-ATP <sub>γ</sub> s (01) | 2,53           | 0,1 M Citrato de sodio pH 5,5 20% p/v PEG 300                            | 10% PEG400   | JCSG  |
| AncEn/HMP-P/Mg-ATP <sub>γ</sub> s (02) | 2,60           | 35% v/v Tacsimato pH 7,0   | 25% Glicerol | INDEX |
| AncEn/HMP-P/Mg-ATP <sub>γ</sub> s (03) | 2,20           | 1% p/v MPEG 2000 1 M ácido succínico 0,1 M HEPES pH 7,0                  | 25% Glicerol | INDEX |
| AncEn/HMP-P/AMP-PNP (04)               | 2,88           | 0,1 M TRIS pH 8,5 0,2 M NaCl 25% p/v PEG 3350                            | 15% Glicerol | INDEX |
| AncEn/HMP-P/ADP (05)                   | 2,30           | 0,5% p/v MPEG 5000 0,8 M Sal de Seignette tetrahidrato 0,1 M Tris pH 8,5 | 25% Glicerol | INDEX |
| AncEn/HMP-P/AMP-PCP (06)               | 3,00           | 0,5% p/v MPEG 5000 0,8 M Sal de Seignette tetrahidrato 0,1 M Tris pH 8,5 | 25% Glicerol | INDEX |
| AncEn/HMP-P/ADPBs (07)                 | 2,40           | 0,8M Ácido succínico pH 7,0  | 25% PEG400   | INDEX |
| AncEn/HMP/AMP-PCP (08)                 | 2,30           | CHES 100mM pH 9,5 20% p/v PEG 8000                                       | 15% PEG400   | JCSG  |

**Tabla 8: Información cristalográfica y estadísticas del refinamiento de las estructuras de AncEnHMPPK co-cristalizado con distintos ligandos.**

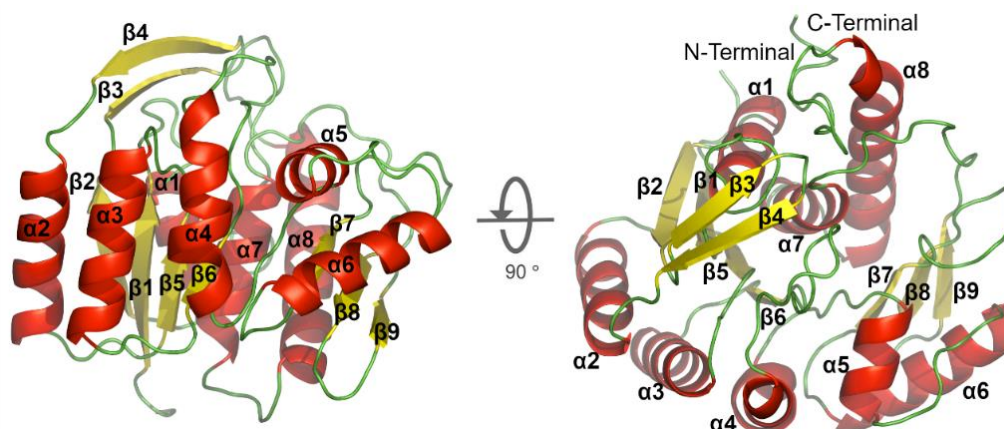
|   | AncEn/HMP-P/<br>ATP $\gamma$ s (01)  | AncEn/HMP-P/<br>ATP $\gamma$ s (02)  | AncEn/HMP-P/<br>ATP $\gamma$ s (03)  | AncEn/HMP-<br>P/AMP-PNP (04)   |
|---|--|--|--|--|
| Longitud de onda (Å)                      | 0,9772   | 0,9772   | 0,9772   | 0,9772   |
| Grupo espacial                            | P 21 21 21   | P 21 21 21   | P 21 21 21   | P 21 21 21   |
| Unidades de celda                         | a = 71,222<br>b = 132,001<br>c = 150,137<br>$\alpha = \beta = \gamma = 90$ (°) | a = 70,715<br>b = 130,294<br>c = 152,001<br>$\alpha = \beta = \gamma = 90$ (°) | a = 67,472<br>b = 147,074<br>c = 159,133<br>$\alpha = \beta = \gamma = 90$ (°) | a = 67,457<br>b = 148,63<br>c = 158,64<br>$\alpha = \beta = \gamma = 90$ (°) |
| Rango de resolución (Å)                   | 46,80 – 2,53   | 47,57 – 2,60   | 45,75 – 2,20   | 48,57 – 2,88   |
| Refracciones totales/ Refracciones únicas | 324956 / 48034   | 639901 / 49524   | 1061143 / 81182  | 455332 / 36908   |
| Multiplicidad                             | 6,8 (7,1)  | 12,9 (13,1)  | 13,1 (13,4)  | 12,2 (10,7)  |
| Complejitud (%)                           | 99,95 (99,92)  | 99,96 (100,0)  | 99,80 (99,88)  | 99,94 (99,97)  |
| I Promedio/sigma(I)                       | 11,85 (1,90)   | 16,98 (2,08)   | 22,65 (3,23)   | 13,12 (2,15)   |
| R-merge                                   | 0,1708 (1,111)   | 0,1158 (1,235)   | 0,06982 (0,7164)   | 0,2321 (1,225)   |
| R-meas                                    | 0,1851 (1,198)   | 0,1206 (1,286)   | 0,07274 (0,7447)   | 0,2422 (1,287)   |
| R-pim                                     | 0,07062 (0,446)  | 0,03347 (0,3529)   | 0,02019 (0,2025)   | 0,06822 (0,3849)   |
| CC 1/2                                    | 0,996 (0,711)  | 1,00 (0,93)  | 0,999 (0,902)  | 0,999 (0,930)  |
| R-work/R-free                             | 0,1871/ 0,2320   | 0,1989 / 0,2348  | 0,1819 / 0,2097  | 0,2069/ 0,2517   |
| Factor-B de Wilson                        | 41,13  | 61,87  | 47,78  | 58,60  |
| Promedio de B-factor (Å <sup>2</sup> )    | 38,95  | 68,76  | 57,60  | 53,35  |
| Ligandos                                  | 45,28  | 70,33  | 53,47  | 61,72  |
| Solventes                                 | 39,88  | 57,99  |  | -  |
| RMS (enlaces) (Å)                         | 0,008  | 0,006  | 0,005  | 0,005  |
| RMS (ángulos) (°)                         | 1,030  | 0,860  | 0,740  | 0,750  |
| Gráfico de Ramachandran                   |  |  |  |  |
| Favorecido (%)                            | 96,60  | 95,75  | 96,79  | 95,57  |
| Permitidos (%)                            | 3,30   | 4,06   | 3,11   | 4,34   |
| Fuera de Ramachandran (%)                 | 0,09   | 0,19   | 0,09   | 0,09   |
| Rotáremos fuera (%)                       | 3,47   | 1,12   | 1,87   | 4,83   |
| Clash score                               | 6,34   | 7,40   | 4,56   | 6,84   |

**Tabla 9: Información cristalográfica y estadísticas del refinamiento de las estructuras de AncEnHMPPK co-cristalizado con distintos ligandos (continuación).**

|   | AncEn/HMP-P/<br>ADP (05)               | AncEn/HMP-P/<br>AMP-PCP (06)             | AncEn/HMP-P/<br>ADPBs (07)             | AncEn/HMP/<br>Mg-AMP-PCP (08)            |
|---|--|--|--|--|
| Longitud de onda (Å)                      | 0,9772                                 | 0,9772                                   | 1,033167                               | 1,033167                                 |
| Grupo espacial                            | P 21 21 21                             | P 21 21 21                               | P 21 21 21                             | P 21 21 21                               |
| Unidades de celda                         | a = 66,471<br>b = 148,63<br>c = 158,64 | a = 68,174<br>b = 147,071<br>c = 161,377 | a = 67,243<br>b = 146,78<br>c = 159,37 | a = 70,715<br>b = 130,294<br>c = 152,001 |
|   | $\alpha = \beta = \gamma = 90$ (°)     | $\alpha = \beta = \gamma = 90$ (°)       | $\alpha = \beta = \gamma = 90$ (°)     | $\alpha = \beta = \gamma = 90$ (°)       |
| Rango de resolución (Å)                   | 46,91 - 2,30                           | 49,09 - 3,00                             | 46,77 - 2,40                           | 41,19 - 2,30                             |
| Refracciones totales/ Refracciones únicas | 797780/ 59487                          | 442676 / 33284                           | 856263 / 62409                         | 856263 / 62952                           |
| Multiplicidad                             | 13,4 (13,9)                            | 13,3 (14,1)                              | 13,7 (13,8)                            | 13,7 (13,8)                              |
| Compleitud (%)                            | 99,95 (99,97)                          | 99,92 (99,97)                            | 99,75 (99,64)                          | 99,58 (99,20)                            |
| I Promedio/sigma(I)                       | 15,75 (2,68)                           | 16,58 (2,27)                             | 24,96 (3,57)                           | 28,98 (3,81)                             |
| R-merge                                   | 0,1878 (1,329)                         | 0,1218 (1,202)                           | 0,07159 (0,7561)                       | 0,07191 (0,8463)                         |
| R-meas                                    | 0,1953 (1,38)                          | 0,1267 (1,247)                           | 0,07441 (0,7852)                       | 0,0747 (0,8783)                          |
| R-pim                                     | 0,05297 (0,3688)                       | 0,03463 (0,3297)                         | 0,02006 (0,2105)                       | 0,02003 (0,2337)                         |
| CC 1/2                                    | 0,998 (0,833)                          | 0,999 (0,803)                            | 0,999 (0,905)                          | 1 (0,912)                                |
| R-work/R-free                             | 0,1870 / 0,2134                        | 0,2005 / 0,2198                          | 0,1932/ 0,2321                         | 0,1875 / 0,2210                          |
| Factor-B de Wilson                        | 35,02                                  | 84,77                                    | 66,36                                  | 47,38                                    |
| Promedio de B-factor (Å <sup>2</sup> )    | 37,16                                  | 83,51                                    | 72,70                                  | 55,06                                    |
| Ligandos                                  | 39,82                                  | 83,29                                    | 73,63                                  | 59,03                                    |
| Solventes                                 | 40,71                                  | -  | 62,77                                  | 54,42                                    |
| RMS (enlaces) (Å)                         | 0,0011                                 | 0,003                                    | 0,015                                  | 0,011                                    |
| RMS (ángulos) (°)                         | 0,96                                   | 0,600                                    | 1,56                                   | 1,30                                     |
| Gráfico de Ramachandran                   |  |  |  |  |
| Favorecido (%)                            | 97,81                                  | 95,00                                    | 97,90                                  | 97,33                                    |
| Permitidos (%)                            | 2,19                                   | 4,81                                     | 2,10                                   | 2,58                                     |
| Fuera de Ramachandran (%)                 | 0,00                                   | 0,19                                     | 0,00                                   | 0,10                                     |
| Rotámeros fuera (%)                       | 0,91                                   | 0,37                                     | 0,25                                   | 2,48                                     |
| Clash score                               | 5,68                                   | 5,48                                     | 7,82                                   | 6,30                                     |

### 3.6. Análisis estructurales de AncEnHMPPK con HMP, HMP-P y análogos no hidrolizables de ATP

Para analizar la topología de la estructura de AncEnHMPPK, las estructuras tridimensionales determinadas en presencia de distintos ligandos se representaron como “cartoon” (**Figura suplementaria 6**), para posteriormente realizar una inspección visual, identificándose los elementos de estructura secundaria junto a su distribución espacial, donde se observa el típico plegamiento tipo Rossmann representativo de esta familia de proteínas, identificándose nueve hebras- $\beta$ , de las cuales siete forman parte de la sábana- $\beta$  central, las cuales están rodeadas de ocho  $\alpha$ -hélices, mientras que la horquilla  $\beta$ 3- $\beta$ 4 forma una tapa hacia el bolsillo de unión a ligandos de AncEnHMPPK. No se encontraron diferencias en la topología en las ocho estructuras refinadas, (**Figura 12**), por lo cual se realizó la superposición estructural utilizando una cadena polipeptídica representativa para cada conjunto de datos.



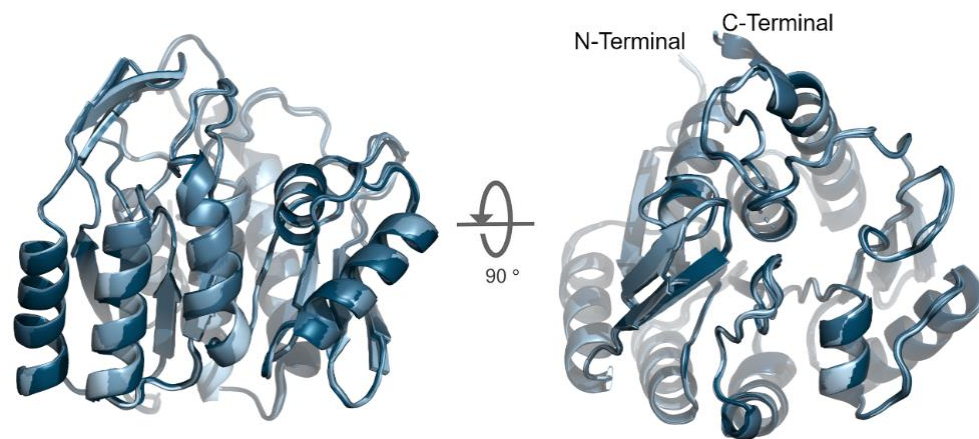
**Figura 12: Estructura cristalográfica representativa de AncEnHMPPK.**

Se muestran dos orientaciones para AncEnHMPPK (rotación  $90^\circ$  respecto eje X). Se presenta la estructura como “cartoon”, donde se colorearon en rojo las  $\alpha$ -hélices, en amarillo las hebras- $\beta$  en verde y en verde los Loop. Los elementos de estructura secundaria se numeraron desde extremo N a C terminal.

Para analizar si los distintos ligandos generaban cambios conformacionales que pudiesen observarse en las estructuras cristalográficas, se realizó la superposición

estructural de un monómero de cada estructura considerando como referencia la estructura de AncEnHMPPK en complejo ternario con HMP y Mg-AMP-PCP (**Figura 13**). Visualmente, no se encontraron diferencias significativas, obteniéndose un RMSD-C $\alpha$  promedio de 0,367 Å con una desviación estándar de 0,151 Å (**Tabla 10**), lo cual indica un alto grado de sobreposición en la disposición espacial de los C $\alpha$  de la cadena principal de las estructuras refinadas. Por lo tanto, no existirían diferencias en la conformación entre las estructuras, independiente de los ligandos presentes en el sitio activo.

Adicionalmente, se realizó la comparación de la topología de AncEnHMPPK con proteínas homólogas, utilizando la herramienta ENDscript (Robert & Gouet, 2014). El alineamiento múltiple de secuencias con homólogos de AncEnHMPPK con estructuras tridimensionales reportadas (**Figura suplementaria 7**), complementado con los elementos de estructura secundaria para las estructuras alineadas, muestra una conservación en la distribución de los elementos de estructura secundaria, indicando que no habría variaciones de topología dentro de la familia.



**Figura 13: Superposición estructural de las estructuras cristalográficas de la proteína ancestral AncEnHMPPK.**

Se representan como “*cartoon*” una única cadena representativa para cada complejo ternario obtenido (**Tabla 10**), Se realizó una superposición estructural usando como punto de comparación la estructura en complejo con HMP/AMP-PCP, de lo cual no se aprecian diferencias significativas entre las estructuras. Se colorearon las distintas estructuras con diferentes tonos de azul.



**Tabla 10: Comparación por RMSD de las conformaciones de las estructuras cristalográficas de los complejos ternarios de AncEnHMPPK.**

|                            | HMP<br>AMP-PCP | HMP-P<br>ADP <sub>βS</sub> | HMP-P<br>ADP<br>CSD213 | HMP-P<br>AMP-PCP | HMP-P<br>AMP-PNP | HMP-P<br>ATP <sub>γS</sub> |
|----------------------------|----------------|----------------------------|------------------------|------------------|------------------|----------------------------|
| HMP<br>AMP-PCP             | -              | 0,405                      | 0,573                  | 0,299            | 0,311            | 0,326                      |
| HMP-P<br>ADP <sub>βS</sub> |                | -                          | 0,589                  | 0,319            | 0,242            | 0,248                      |
| HMP-P<br>ADP<br>CSD213     |                |                            | -                      | 0,550            | 0,533            | 0,544                      |
| HMP-P<br>AMP-PCP           |                |                            |                        | -                | 0,206            | 0,197                      |
| HMP-P<br>AMP-PNP           |                |                            |                        |                  | -                | 0,168                      |
| HMP-P<br>ATP <sub>γS</sub> |                |                            |                        |                  |                  | -                          |

### 3.7. Análisis del bolsillo de unión a sustratos de las estructuras de AncEnHMPPK

Para identificar residuos de aminoácidos con un rol clave tanto en la unión de los sustratos como en la catálisis, se analizaron los bolsillos de unión de los distintos complejos ternarios obtenidos para AncEnHMPPK (**Figura 14**) centrándose en la primera capa de coordinación de los ligandos, esto es, residuos de aminoácidos con algún átomo a una distancia no mayor a 5 Å de los ligandos (**Tabla 11**).

De la inspección visual de los bolsillos de unión representativos (**Figura 14**), se identificaron los siguientes residuos que mediarían la unión de HMP y HMP-P: glutamina 44, lisina 111 e histidina 209, de los cuales los residuos de lisina e histidina podrían mediar la estabilización del grupo metil fosfato de la molécula de HMP-P en el bolsillo de unión, mientras que la glutamina 44 interactuaría con el anillo de pirimidina del HMP/HMP-P.

Por otro lado, en el caso del nucleótido (análogos no hidrolizables de ATP y ADP) se identificaron los siguientes residuos: valina 203 que interacciona con el anillo de adenina, glicina 178, serina 185, aspártico 187 y lisina 237 que interactúan con la ribosa, y finalmente los residuos lisina 111, asparagina 139, lisina 176, histidina 179, treonina 208, treonina 211 y glicina 212 que interactuarían con los grupos fosfato  $\alpha$ ,  $\beta$  y  $\gamma$  del nucleótido.

**Tabla 11: Residuos a 5 Å de los ligandos HMP/HMP-P y análogo no hidrolizable del nucleótido.**

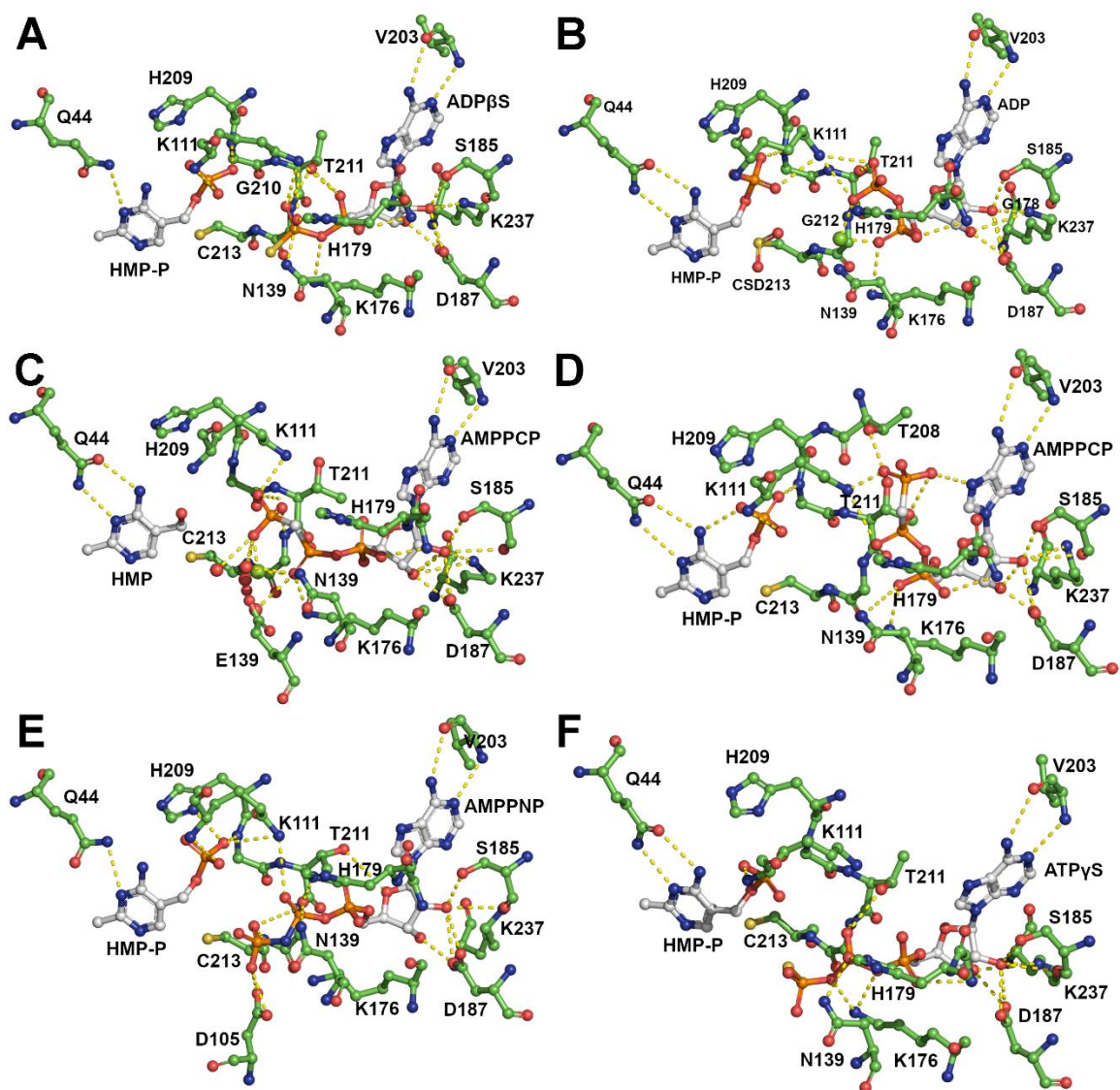
| Ligando  | HMP/AMP-PCP       |             | HMP-P/Análogo ATP |               | HMP-P/ADP $\beta$ S ADP |               |
|--|-------------------|-------------|-------------------|---------------|-------------------------|---------------|
|  | HMP/HMP-P         | Gly11       | Lys111            | Gly11         | Lys111                  | Gly11         |
| Thr12  |                   | Leu116      | Thr12             | <b>Ser112</b> | Thr12                   | <b>Ser112</b> |
| Asp13  |                   | His209      | Asp13             | Leu116        | Asp13                   | Leu116        |
| Ala18  |                   | Gly210      | Ala18             | <b>Thr208</b> | Ala18                   | <b>Thr208</b> |
| Gly19  |                   | Cys213      | Gly19             | His209        | Gly19                   | His209        |
| Val42  |                   | AMP-        | Val42             | Gly210        | Val42                   | Gly210        |
| Gln44  |                   | PCP         | Gln44             | <b>Thr211</b> | Gln44                   | <b>Thr211</b> |
| Met80  |                   |             | Met80             | Cys213        | Met80                   | <b>Gly212</b> |
| Val107 $\ddagger$  |                   |             | Val107 $\ddagger$ | Análogo       | Val107                  | Cys213        |
| Leu109   |                   |             | Leu109            |               | Leu109                  | <b>Ile256</b> |
| Ala110   |                   |             | Ala110            |               | Ala110                  | ADP $\beta$ S |
| Análogo no hidrolizable (AMP-PCP, AMP-PNP, ATP $\gamma$ S o ADP $\beta$ S) | Asp105 $\ddagger$ | Arg202      | Lys111            | Lys237        | Asp105                  | Ala204        |
|  | Thr106 $\ddagger$ | Val203      | Gln139 $\dagger$  | Leu240        | Val107                  | Thr205        |
|  | Val107 $\ddagger$ | Ala204      | Lys176            | Ser241        | Lys111                  | Thr208        |
|  | Leu109            | Thr205      | Gly177            | Leu244        | Gln139 $\dagger$        | Gly210        |
|  | Lys111            | Thr208      | Gly178            | HMP-P         | Glu142 $\dagger$        | Thr211        |
|  | Thr137            | His209      | His179            |               | Lys176                  | Gly212        |
|  | Gln139 $\dagger$  | Gly210 $\#$ | Ser185            |               | Gly177                  | Cys213        |
|  | Glu142 $\dagger$  | Thr211 $\#$ | Pro186            |               | Gly178                  | Leu215        |
|  | Lys176            | Gly212 $\#$ | Asp187            |               | His179                  | Lys237        |
|  | Gly177            | Cys213 $\#$ | Pro201            |               | Ser185                  | Leu240        |
|  | Gly178            | Leu215      | Arg202            |               | Pro186                  | Ser241        |
|  | His179            | Lys237      | Val203            |               | Asp187                  | Leu244        |
|  | Ser185            | Leu240      | Thr208            |               | Ala200                  | HMP-P         |
|  | Pro186            | Ser241      | Thr211            |               | Pro201                  |               |
|  | Asp187            | Leu244      | Gly212            |               | Arg202                  |               |
|  | Pro201            | HMP         | Leu215            |               | Val203                  |               |

$\ddagger$  Residuo del motivo DPV

$\dagger$  Residuo del motivo NXXE

$\#$  Residuo del motivo GXGD

**En negrita se destacan residuos que solo se encontraría presente en el bolsillo de unión para HMP-P**



**Figura 14: Bolsillo de unión de nucleótido de AncEnHMPPK en presencia de HMP y HMP-P.**

Se presentan los bolsillos de unión de los distintos complejos ternarios obtenidos para AncEnHMPPK. Los residuos y los ligandos del sitio activo se presentan como “stick balls” coloreados según átomos: Los átomos de carbono se coloraron en verde (proteína) y blanco (ligandos), los átomos de nitrógeno en azul, los átomos de oxígeno en rojo y los átomos de azufre en amarillo. Las líneas punteadas amarillas representan puentes de hidrógeno. Se muestran los bolsillos de unión de los siguientes complejos: **(A)** HMP-P y  $ADP_{\beta S}$ . **(B)** HMP-P y ADP. **(C)** HMP y Mg-AMP-PCP. **(D)** HMP-P y AMP-PCP. **(E)** HMP-P y AMP-PNP. **(F)** HMP-P y  $ATP_{\gamma S}$ .

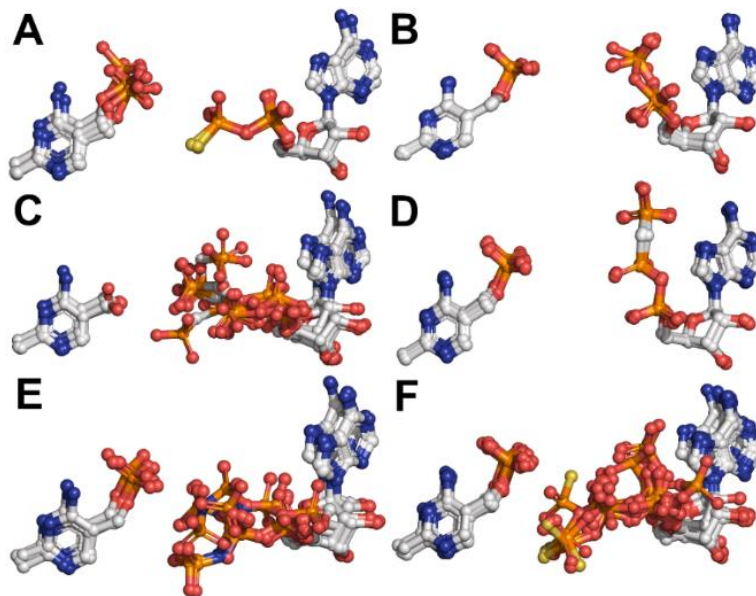
Interesantemente no se identificaron cambios estructurales importantes en el bolsillo de unión entre las estructuras cristalográficas de los complejos ternarios con HMP y HMP-P, ni tampoco reorientación de loops del sitio activo. Esto contrasta con lo propuesto por

Cheng y col, quienes proponen que las ThiD-HMPPK presentan una reorganización de los loops adyacentes al bolsillo de unión entre la primera y segunda fosforilación de la molécula de HMP.

Sin embargo, sorprendentemente se encontró que la molécula de nucleótido (análogos de ATP) adopta distintas conformaciones en el bolsillo de unión, tanto entre distintas cadenas de un mismo cristal como entre cadenas de cristales distintos, mientras que las moléculas de HMP y HMP-P siempre presentan las mismas conformaciones. Estas distintas conformaciones de la molécula de nucleótido (**Figura 15**) varían en la distribución espacial de los fosfatos  $\beta$  y  $\gamma$  encontrándose conformeros que podrían ser de alta energía debido a la torsión entre los distintos grupos fosfatos. Dentro de las distintas conformaciones se aprecian cambios en la orientación del fosfato  $\gamma$  y la ausencia del ión magnesio, dado que no se aprecia densidad electrónica para este ión, sumado a que en las cercanías de los fosfato  $\beta$  y  $\gamma$  no se aprecian moléculas de agua altamente ordenadas que estarían mediando la capa de coordinación del ión magnesio. Por otro lado, la ausencia de magnesio en complejo con la molécula de nucleótido podría contribuir a la presencia de estas distintas conformaciones de nucleótido encontradas en las estructuras cristalográficas.

De estas distintas conformaciones de la molécula de nucleótido se pueden distinguir tres grupos, correspondiendo a: **1.** Conformación extendida (típica o clásica), en la cual los fosfatos  $\alpha$ ,  $\beta$  y  $\gamma$  están en paralelo respecto a la ribosa. **2.** Conformación vertical, en la cual los fosfatos  $\alpha$ ,  $\beta$  y  $\gamma$  están en perpendicular respecto a la ribosa. **3.** Conformación atípica del fosfato  $\gamma$  donde el fosfato  $\gamma$  del nucleótido se ordena saliendo del plano en el

cual se distribuyen los fosfatos  $\alpha$  y  $\beta$ . En la **Figura suplementaria 8** se muestra un ejemplo representativo de cada conformación.



**Figura 15: Conformaciones de los análogos de ADP y ATP en el bolsillo de unión a sustrato de AncEnHMPPK.**

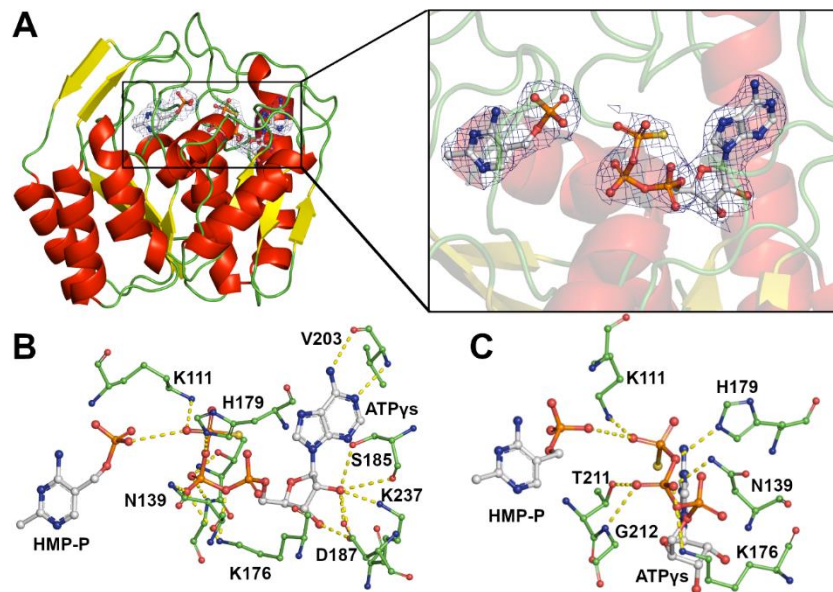
Se presentan distintas conformaciones encontradas en los bolsillos de unión a sustrato de AncEnHMPPK. **(A)** Complejo HMP-P y  $\text{ADP}_{\beta\text{S}}$  ( $n=3$ , correspondientes a tres cadenas polipeptídicas). **(B)** Complejo HMP-P y ADP ( $n=4$ , correspondientes a cuatro cadenas polipeptídicas). **(C)** Complejo HMP y Mg-AMP-PCP ( $n=8$ , correspondientes a cuatro cadenas polipeptídicas y dos conformaciones alternativas por cadena). **(D)** Complejo HMP-P y AMP-PCP ( $n=2$ , correspondientes a dos cadenas polipeptídicas). **(E)** Complejo HMP-P y AMP-PNP ( $n=4$ , correspondientes a cuatro cadenas polipeptídicas). **(F)** Complejo HMP-P y  $\text{ATP}_{\gamma\text{S}}$  ( $n=12$ , correspondientes a tres cristales distintos con cuatro cadenas polipeptídicas por cristal). Las moléculas de ligandos se presentan como “stick balls” coloreados según átomos: Los átomos de carbono se coloraron en blanco ligandos, los átomos de nitrógeno en azul, los átomos de oxígeno en rojo y los átomos de azufre en amarillo.

La ausencia del ión magnesio en el sitio activo es algo de sumo interés dado que el verdadero sustrato de estas enzimas es el complejo Mg-ATP (Flachner y col., 2004). Por lo tanto, es esperable apreciar este complejo en las estructuras cristalográficas de mayor resolución, por lo que es posible que las condiciones en las que se realizó el almacenamiento de los cristales influyeron en la difusión del ion magnesio del complejo con los análogos no hidrolizables de ATP, dado que para la estructura de AncEnHMPPK

en complejo ternario con HMP y Mg-AMPCP de la primera ronda de colección de datos se aprecia el ion magnesio en tres de los cuatro monómeros. Donde para la pesca y almacenamiento de los cristales se suplementó con ligandos la solución crio protectora, a diferencia del proceso de almacenamiento de la segunda ronda de colección de datos, donde no se suplementó la solución crio-protectora con ligandos.

Respecto a las conformaciones atípicas del fosfato  $\gamma$  del ATP $_{\gamma S}$ , se aprecia una en particular que se ilustra en la **Figura 16**, donde se observa un alejamiento de los grupos fosfato  $\gamma$  del análogo de ATP y el metil fosfato del HMP-P (6,2 Å entre los átomos de fosforo de ambos fosfatos y 3,5 Å entre los átomos de oxígeno más próximos entre ambos grupos fosfatos). Este alejamiento podría ser favorable para la segunda fosforilación impidiendo que estos grupos se repelan entre sí, por lo tanto, sería una conformación de menor energía. Además, esta conformación atípica del fosfato  $\gamma$  estaría estabilizada en el bolsillo de unión por los residuos lisina 111, asparagina 139, lisina 176, histidina 179, treonina 211 y glicina 212, por interacciones electrostáticas y formación de puentes de hidrógeno.

Este complejo ternario tiene dos particularidades relevantes: **1.** la ausencia del ión magnesio junto a la molécula de ATP $_{\gamma S}$  ya que para esta familia de quinasas es el verdadero sustrato es el complejo Mg-ATP. **2.** Una gran distancia (9,4 Å) entre el grupo -SH de la cisteína 213 y el átomo de fosforo del fosfato  $\gamma$  del ATP $_{\gamma S}$  dado que este residuo ha sido reportado como el catalítico para esta familia de enzimas, lo que sugiere que la segunda fosforilación podría ser mediada por un mecanismo catalítico diferente al de la primera fosforilación.

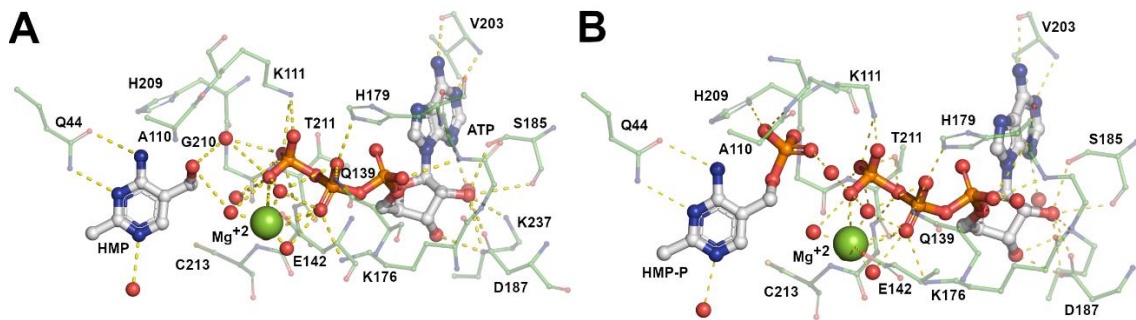


**Figura 16: Bolsillo de unión a sustrato de AncEnHMPPK unido a HMP-P y ATP $\gamma$ S en conformación atípica.**

(A) Modelo tridimensional de AncEnHMPPK en complejo ternario con HMP-P y ATP $\gamma$ S, junto a un acercamiento al sector de los ligandos donde se presenta el mapa 2mFo-DFc para los ligandos, contorneado a  $1\sigma$ , coloreado de color gris. (B-C) Bolsillo de unión para ATP $\gamma$ S. Se muestran residuos a 5 Å de la molécula de ATP $\gamma$ S, además se presentan residuos que presentan interacciones con la conformación atípica del fosfato  $\gamma$  y de la molécula de ATP $\gamma$ S. Las moléculas de ligandos y residuos de aminoácidos se presentan como “stick balls” coloreados según átomos: Los átomos de carbono se coloraron en blanco (ligandos) y en verde (proteínas), los átomos de nitrógeno en azul, los átomos de oxígeno en rojo y los átomos de azufre en amarillo.

En base a la información obtenida del modelo tridimensional de AncEnHMPPK en complejo ternario con HMP y Mg-AMP-PCP, fue posible reconstruir los complejos Mg-ATP mediante el software de visualización “*coot*” (Figura 17). Donde no se aprecian diferencias significativas entre los complejos ternarios con HMP y HMP-P respecto a la coordinación de los ligandos, encontrándose leves diferencias en las distancias de residuos como Gln44, Lys111, His179 y Thr211, los cuales se destacan en las **Tabla suplementaria 6** y **Tabla suplementaria 7**. De estos complejos ternarios se observa que el residuo Cys213 se encuentra alejado respecto al dador y aceptor de fosforilo para ambas fosforilaciones, estando el grupo -SH a una distancia de 5,6 y 6,9 Å respecto al HMP (-OH aceptor del fosforilo) y ATP (-PO $_4\gamma$ ) respectivamente, en el complejo HMP/Mg-

ATP y de 6,1 y 6,8 Å, respecto al HMP-P (-PO<sub>4</sub> aceptor del fosforilo) y ATP (-PO<sub>4</sub>γ), lo cual podría indicar que la cisteína 213 no tendría participación relevante en la segunda fosforilación del HMP, además se aprecia que los residuos His179 y Thr211 presentan un cambio de orientación acercando sus cadenas laterales a los grupos fosfato β y γ de la molécula de ATP.



**Figura 17: Reconstrucción de los complejos Mg-ATP para los bolsillos de unión de AncEnHMPPK en complejo ternario con HMP/Mg-AMPCP y HMP-P/ATP<sub>γs</sub>.**

(A) Bolsillo de unión para el complejo Mg-ATP en presencia de HMP. (B) Bolsillo de unión para el complejo Mg-ATP en presencia de HMP-P. Se muestran residuos a 5 Å de las moléculas de ligandos. Las moléculas de ligandos y residuos de aminoácidos se presentan como “stick balls” coloreados según átomos: Los átomos de carbono se coloraron en blanco (ligandos) y en verde (proteínas), los átomos de nitrógeno en azul, los átomos de oxígeno en rojo y los átomos de azufre en amarillo.

### 3.8. Reconstrucción filogenética por máxima verosimilitud

Para identificar mutaciones claves que surgieron durante la divergencia y evolución de las ThiD-HMPPK y las PdxK-PLK/HMPPK se realizó un alineamiento múltiple de secuencias entre estos grupos de enzimas, utilizándose como templates para la búsqueda y colecta de secuencias las PdxK-PLK/HMPPK de *Bacillus subtilis* y *Staphylococcus aureus*, reduciéndose el número de secuencias a alinear filtrando por identidad y tamaño. Las secuencias restantes se alinearon con ClustalW y el alineamiento resultante se alineo con un MSA de ThiD-HMPPK como se describe en la sección de materiales y métodos. Finalmente, se eliminaron secuencias redundantes



filtrando por identidad al 85% en VMD, resultando un MSA con 160 secuencias, las cuales se listan con su código de acceso y especie en la **Tabla suplementaria 8**. Con el alineamiento de secuencias múltiples se identificaron posiciones invariantes entre ambos grupos de enzimas (**Tabla 12**), las cuales corresponden a residuos de aminoácidos que están o son próximos al bolsillo de unión de sustratos de estas enzimas los cuales se indican en la **Tabla 12**. Además, algunas de estas posiciones invariantes franquean o constituyen algunos de los motivos conservados de esta familia de proteínas (**Figura 18**).

**Tabla 12: Posiciones invariantes obtenidas del alineamiento múltiple de secuencias de las quinasas de vitaminas dependientes de ATP de Bacteria.**

| Residuos* |         |         |
|-----------|---------|---------|
| Thr8      | Gln63   | Gly169  |
| Ala10     | Lys77   | Lys176# |
| Gly11#    | Gly79   | Gly177# |
| Gly16     | Asp105# | Gly178# |
| Ala18     | Val107# | Asp187# |
| Gly19     | Leu129  | Gly210# |
| Ala22     | Pro131  | Gly212# |
| Asp23     | Thr137# | Cys213# |
| Leu24     | Pro138  | Ala243  |
| Lys25     | Asn139# | Gly257  |
| Phe27     | Glu142# |         |
| Gly34     | Ala143  |         |

\* Numeración respecto secuencia de AncEnHMPPK

# Residuos de la primera capa de coordinación de ligandos en el bolsillo de unión de AncEnHMPPK.

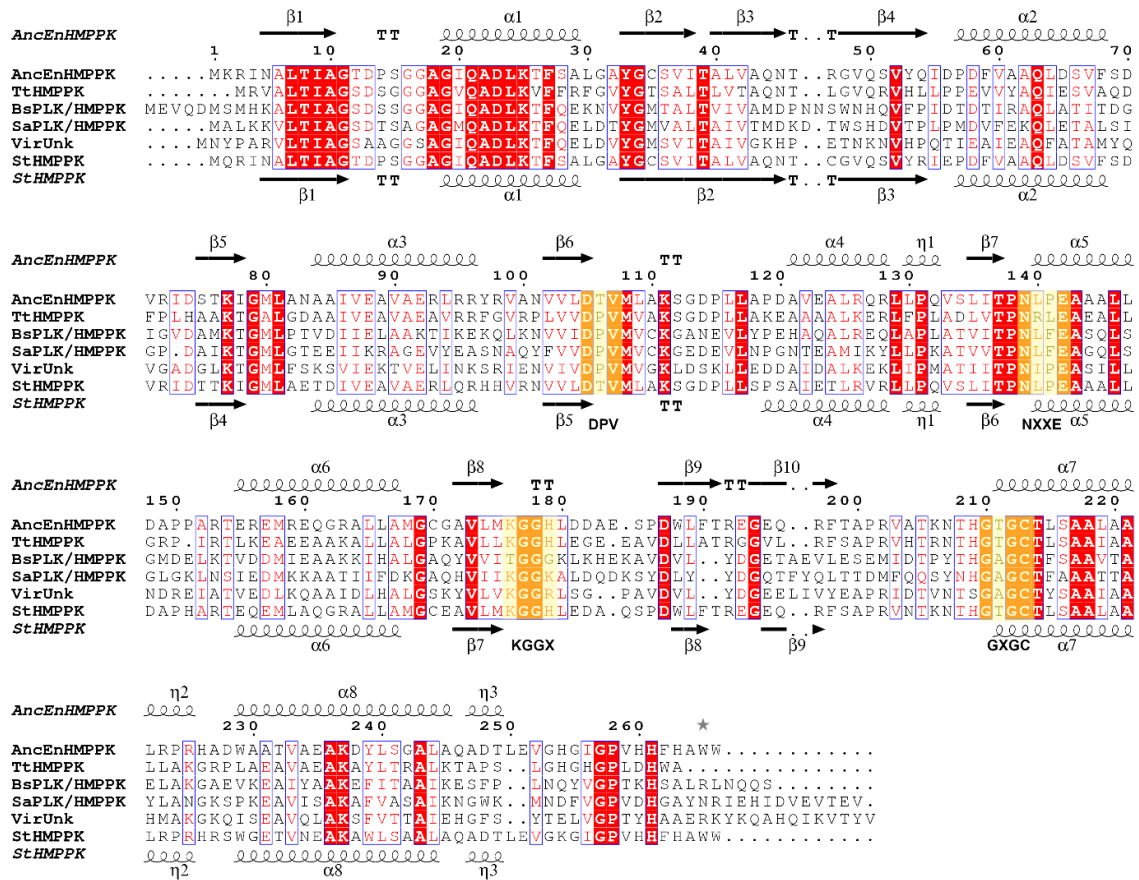
En la **Figura 18** se ilustra un alineamiento representativo para la familia de las quinasas de vitaminas dependientes de ATP de Bacteria donde se destacan los motivos conservados. Estos motivos son, DPV y NXXE involucrados en la coordinación del catión divalente en el sitio activo (ión  $Mg^{2+}$ ), y el motivo catalítico GXGC, el que presenta variaciones entre las ThiD-HMPPK y PdxK-PLK/HMPPK encontrándose como GTGC y GAGC respectivamente. Adicionalmente, fue posible identificar un nuevo motivo conservado en la familia, KGGX, involucrado en la estabilización de la molécula de ATP

en el bolsillo de unión. Este motivo está presente en el 100% de las secuencias colectadas, y en las ThiD-HMPPK se encuentra como KGGH, siendo este último residuo equivalente a la His179 (numeración AncEnHMPPK) que, en base a la información cristalográfica, interaccionaría con el fosfato  $\beta$  y  $\gamma$  de la molécula de ATP (**Figura 14** y **Figura 16**).

Mediante MSA se analizó en que elementos estructurales se encuentran los residuos conservados del bolsillo de unión de ligandos. En su mayoría se encuentran ubicados en regiones correspondientes a loops de los modelos tridimensionales de AncEnHMPPK (**Figura 12** y **Figura suplementaria 7**). Por ejemplo, el residuo Gln44 involucrado en la unión de HMP-P se encuentra en el Loop  $\beta 3$ - $\beta 4$ , los motivos involucrados en la unión del catión divalente DPV y NXXE se encuentran en el Loop  $\beta 6$ - $\alpha 4$  y en la  $\alpha$ -hélice  $\alpha 5$  respectivamente, mientras que el motivo KGGX involucrado en la unión de nucleótido se encuentra en el Loop  $\beta 8$ - $\beta 9$  y finalmente el motivo catalítico junto al residuo His209 se encuentran en el Loop  $\beta 10$ - $\alpha 7$ .

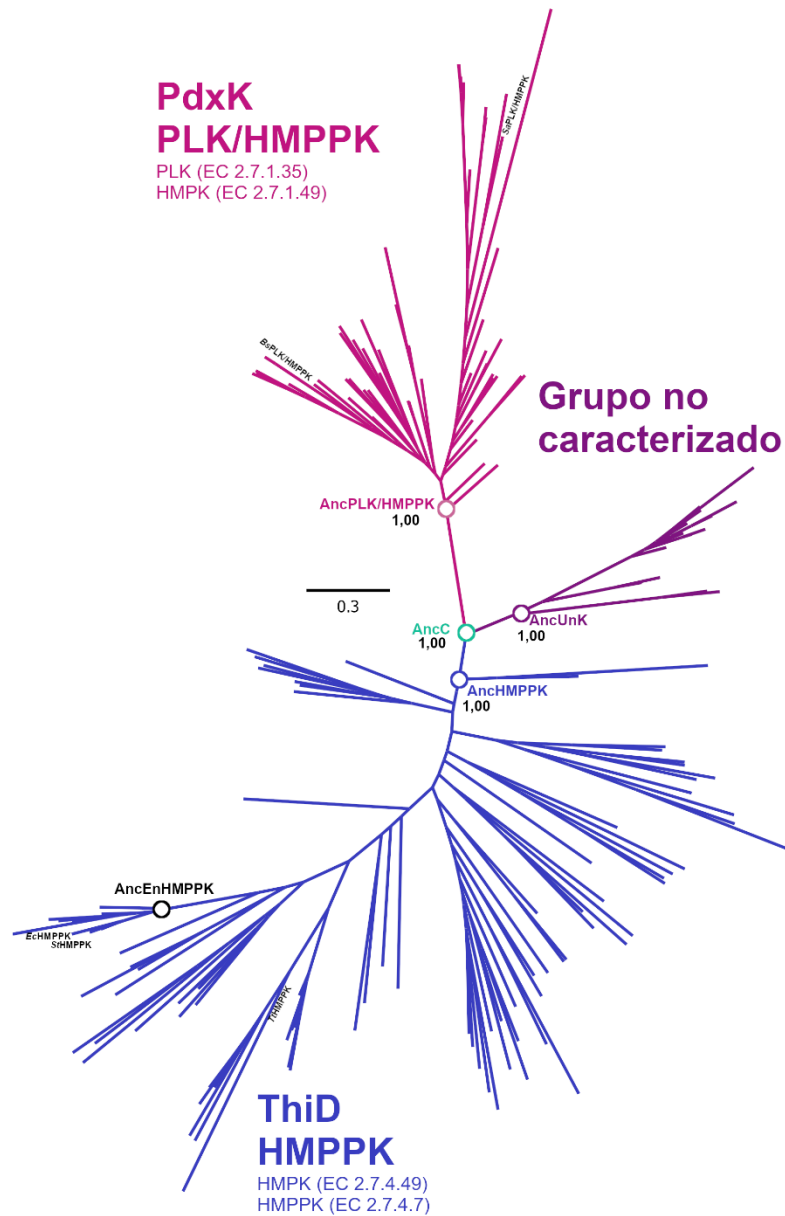
Para la inferencia filogenética de esta familia de proteínas, se evaluó el modelo evolutivo que se ajusta mejor al conjunto de secuencias con Prottest3 (Darriba y col., 2011). Considerando todas las matrices disponibles junto a parámetros de frecuencia, sitios invariantes y distribución gamma, se escogió el modelo evolutivo con el menor valor de Akaike y máxima verosimilitud. De los modelos evaluados se muestran los cinco mejores en la **Tabla suplementaria 9**, siendo el mejor modelo evolutivo el LG +G +I, el cual corresponde al uso de una matriz de reemplazo LG (Le & Gascuel, 2008), empleando una distribución gamma discreta de las tasas de cambio (+G), combinado con una categoría de sitios invariantes (+I). Finalmente, tras la selección del mejor modelo

evolutivo para este conjunto de secuencias, se infirió una filogenia por máxima verosimilitud utilizando PhyML3.1 (Guindon y col., 2010) utilizando el script indicado en la sección anexo (7.4.3.1).



**Figura 18: Alineamiento representativo de quinasas de vitaminas dependientes de ATP de bacterias.**

El alineamiento se realizó con homólogos de la familia quinasas de vitaminas dependientes de ATP de bacterias, encontrándose coloreado de acuerdo con el grado de similitud de las secuencias. En la línea superior se muestran los elementos de estructura secundaria de AncEnHMPPK, mientras que en la línea inferior se muestran los elementos de estructura secundaria de la HMPPK de *Salmonella typhimurium* (PDB ID: 1JX1), calculadas con DSSP. En el alineamiento se muestran las secuencias de quinasas de vitaminas dependientes de ATP de bacteria, representativas de los grupos: 1.- ThiD-HMPPK: La HMPPK ancestral del orden Enterobacteriales (AncEnHMPPK), la HMPPK de *Salmonella typhimurium* (StHMPPK) y la HMPPK de *Thermus thermophilus* (TtHMPPK). 2.- PdxK-PLK/HMPPK: La PLK/HMPPK de *Bacillus subtilis* (BsPLK/HMPPK) y la PLK/HMPPK de *Staphylococcus aureus* (SaPLK/HMPPK). 3.- Un grupo no reportado anteriormente sin caracterización (Unk) con un representante de *Virgibacillus sp.* SK37 (VirUnk). De estos grupos se presentan motivos conservados en recuadros de color amarillo. Este alineamiento fue construido con la plataforma web ENDscript (Robert & Gouet, 2014).



**Figura 19: Relaciones filogenéticas de las quinasas de vitaminas dependientes de ATP de bacterias.**

Árbol filogenético por máxima verosimilitud no enraizado. Los colores representan diferentes grupos filogenéticos, correspondiendo el negro a las ThiD-HMPPK, verde a las PdxK-PLK/HMPPK, fucsia a un grupo sin caracterización reportada. Se presentan en la filogenia la enzima ancestral AncEnHMPPK y se indican las enzimas actuales *Bs*PLK/HMPPK, *Ec*HMPPK, *Sa*PLK/HMPPK, *St*HMPPK y *Th*HMPPK, que tienen caracterizaciones bioquímicas reportadas, junto la proteína ancestral AncEnHMPPK. Se utilizó el modelo evolutivo LG+I+G de acuerdo con la selección de modelos por el factor de información de Akaike (AIC) (Tabla suplementaria 9). Los números de las ramas representan los valores de soporte calculados mediante aLRT paramétrico (basados en  $\chi^2$ ) utilizando PhyML (versión 3.1) para las secuencias ancestrales a reconstruir.

De la filogenia resultante (**Figura 19** y **Figura suplementaria 11**), se puede apreciar la distinción de dos grupos monofiléticos claramente identificables, los cuales corresponderían a las ThiD-HMPPK y las PdxK-PLK/HMPPK, en concordancia con lo descrito por Cea y col., 2020. Para enraizar el árbol y darle una direccionalidad, se utilizó la secuencia de la ThiD-HMPPK de *Deinococcus geothermalis*, dado que presenta la menor tasa de sustitución y la cual posteriormente se utilizó como grupo externo para la reconstrucción de secuencias ancestrales.

Al analizar en detalle la topología del árbol, se aprecia un nuevo grupo de secuencias emparentado a las PdxK-PLK/HMPPK. Al rastrear estas secuencias en el MSA, se observó que dichas secuencias no comparten gran parte de los residuos del sitio activo descritos para las ThiD-HMPPK y PdxK-PLK/HMPPK. Además, los residuos involucrados en la unión de HMP-P identificados en esta tesis también difieren en este nuevo grupo. Por lo tanto, como estas secuencias no presentan una anotación clara, debido a que aparecen anotadas como piridoxal, piridoxamina y piridoxina quinasas en las bases de datos, pero no hay ningún reporte experimental sobre actividades enzimáticas para alguna de estas secuencias, nos referiremos a este nuevo grupo como “Grupo no caracterizado” (**Figura 19**). Luego de la clasificación de grupos, se procedió a seleccionar los nodos de interés para realizar la reconstrucción de secuencias ancestrales, correspondiendo estos nodos a: **1.** El último ancestro en común de las ThiD-HMPPK que da origen a las PdxK-PLK/HMPPK (AncHMPPK). **2.** El último ancestro en común entre las PdxK-PLK/HMPPK y el grupo no caracterizado (AncC). **3.** El último ancestro en común del grupo no caracterizado (AncUnk). **4.** El último ancestro en común de las PdxK-PLK/HMPPK. Dichos nodos presentan un buen soporte estadístico con aLRT paramétrico basado en  $\chi^2$  igual a 1,0, por lo cual se espera que independiente de

errores en el alineamiento, la reconstrucción de estos ancestros sea robusta. La reconstrucción de las secuencias de los nodos indicados se aborda con mayor detalle en la siguiente sección.

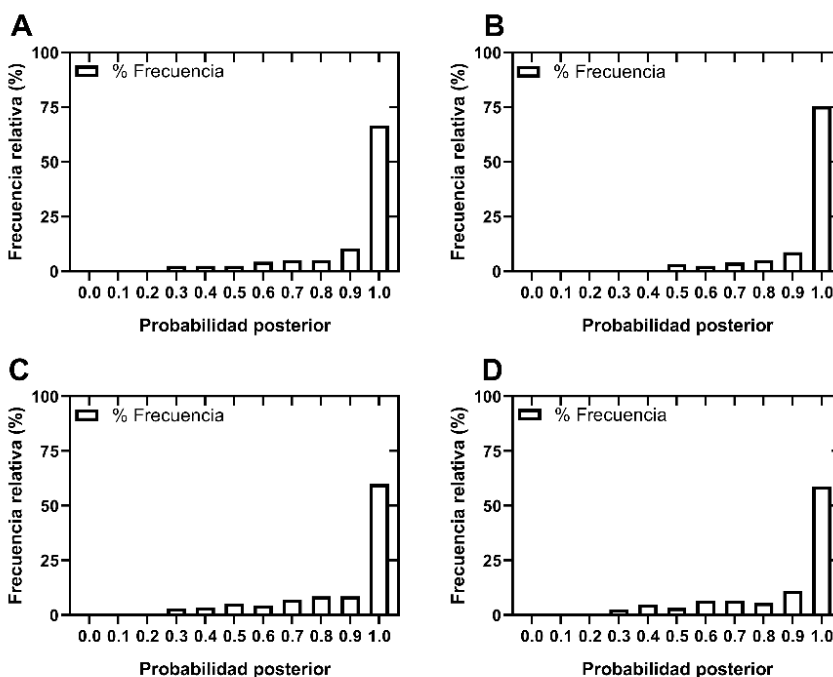
### **3.9. Reconstrucción de secuencias ancestrales de la familia de quininas de vitaminas dependientes de ATP de bacteria**

Se realizó la inferencia de las secuencias ancestrales de los últimos ancestros en común de los tres grupos identificados, ThiD-HMPPK, PdxK-PLK/HMPPK, y el grupo no caracterizado, junto al último ancestro en común entre los grupos de PdxK-PLK/HMPPK y el grupo no caracterizado (AncUnk). Para inferir los estados ancestrales se corrigieron los espacios del alineamiento mediante máxima verosimilitud, interpretando la presencia o ausencia de un espacio a lo largo de alineamiento como rasgos discretos, utilizando el programa Mesquite 3.61 (Maddison & Maddison, 2019). Las secuencias ancestrales reconstruidas presentan un largo entre 268 a 271 aa, con un promedio de probabilidad posterior (PP) mayor a 0,860 y con al menos un 58,60% de las posiciones con una probabilidad posterior igual a 1,0. La **Tabla 13** muestra la comparación de la robustez y calidad de las secuencias reconstruidas. Además, el análisis general de las probabilidades posteriores por posición se muestra en la **Figura suplementaria 9**, mientras que el análisis estadístico mediante histogramas se muestra en la **Figura 20**, el cual ilustra que al menos el 75% de las posiciones presenta una PP mayor a 0,8 para ese aminoácido en las cuatro secuencias. Estos análisis sugieren que la secuencia de AncPLK/HMPPK presenta la mayor robustez de reconstrucción (PP promedio de 0,938), mientras que la secuencia de AncUnk presenta la menor (PP promedio de 0,860). Los residuos con menor PP se encuentran en las posiciones más variables del alineamiento, que corresponden principalmente a loops y a los extremos N y C terminal de las

secuencias. Las secuencias reconstruidas para AncHMPPK, AncC, AncUnk y AncPLK/HMPPK se encuentran disponibles en el anexo 7.4.3.3.

**Tabla 13: Comparación de las secuencias ancestrales reconstruidas.**

| Secuencia    | Largo de secuencia (aa) | PP = 1,0 | PP promedio |
|--------------|-------------------------|----------|-------------|
| AncHMPPK     | 268                     | 66,79%   | 0,897       |
| AncC         | 269                     | 59,85%   | 0,867       |
| AncUnk       | 271                     | 58,67%   | 0,860       |
| AncPLK/HMPPK | 270                     | 75,56%   | 0,938       |



**Figura 20: Histograma de distribución de probabilidades posteriores para la reconstrucción de secuencias ancestrales.**

(A) AncthiD. (B) AncPLK/HMPPK. (C) AncC. (D) AncUnk. Graficas muestran la frecuencia relativa en porcentaje para intervalos discretos de probabilidad posterior.

### 3.10. Historia evolutiva de la actividad HMP-P quinasa de bacterias

Para trazar los posibles cambios en la especificidad por sustratos y en la pérdida de la capacidad de catalizar la fosforilación de HMP-P de las enzimas ThiD-HMPPK, analizamos en conjunto la información proveniente del estudio filogenético de esta familia y del estudio estructural referente al bolsillo de unión a ligando de la proteína

ancestral AncEnHMPPK de las estructuras cristalográficas determinadas. Esto permitió identificar las principales sustituciones que habrían ocurrido durante la divergencia de esta familia de proteínas a lo largo de las distintas secuencias ancestrales reconstruidas.

Para complementar este análisis se verificó la probabilidad posterior de los residuos involucrados en la unión de HMP-P en el bolsillo de unión de AncEnHMPPK, (**Tabla 14**), donde se aprecia que la mayoría de estos presentan una probabilidad posterior mayor a 0,5. De este análisis se destaca que el último ancestro en común de las ThiD-HMPPK presenta todos los residuos involucrados en la fosforilación del HMP-P a HMP-PP, por lo cual debería ser capaz de realizar la segunda fosforilación del HMP. Por otra parte, al seguir la ruta evolutiva (**Figura 21 y Figura suplementaria 10**) solo se aprecia la sustitución del residuo Thr211 (numeración respecto AncEn HMPPK) a alanina en la proteína ancestral ancC, la que se conserva a lo largo de la evolución hacia las PdxK-PLK/HMPPK. Al proseguir esta ruta evolutiva se dan varias sustituciones en simultáneo, como Q44M, A110C y H179S hacia la rama de las PdxK-PLK/HMPPK (AncPLK/HMPPK) y Q44R, A110G y H179R hacia la rama del grupo no caracterizado (AncUnk), mientras que hacia la divergencia a las PdxK-PLK/HMPPK se da la inserción de dos residuos en el Loop  $\beta$ 3- $\beta$ 4.

Estos resultados sugieren que la proteína ancestral AncC podría, catalizar la segunda fosforilación del HMP a HMP-PP, mientras que esta actividad se perdería hacia la divergencia a los ancestros AncPLK/HMPPK y AncUnk. Esto indicaría que durante la evolución la capacidad de fosforilar PL apareció repentinamente y podría haberle conferido una ventaja adaptativa a la bacteria por lo que se habría mantenido durante la evolución.

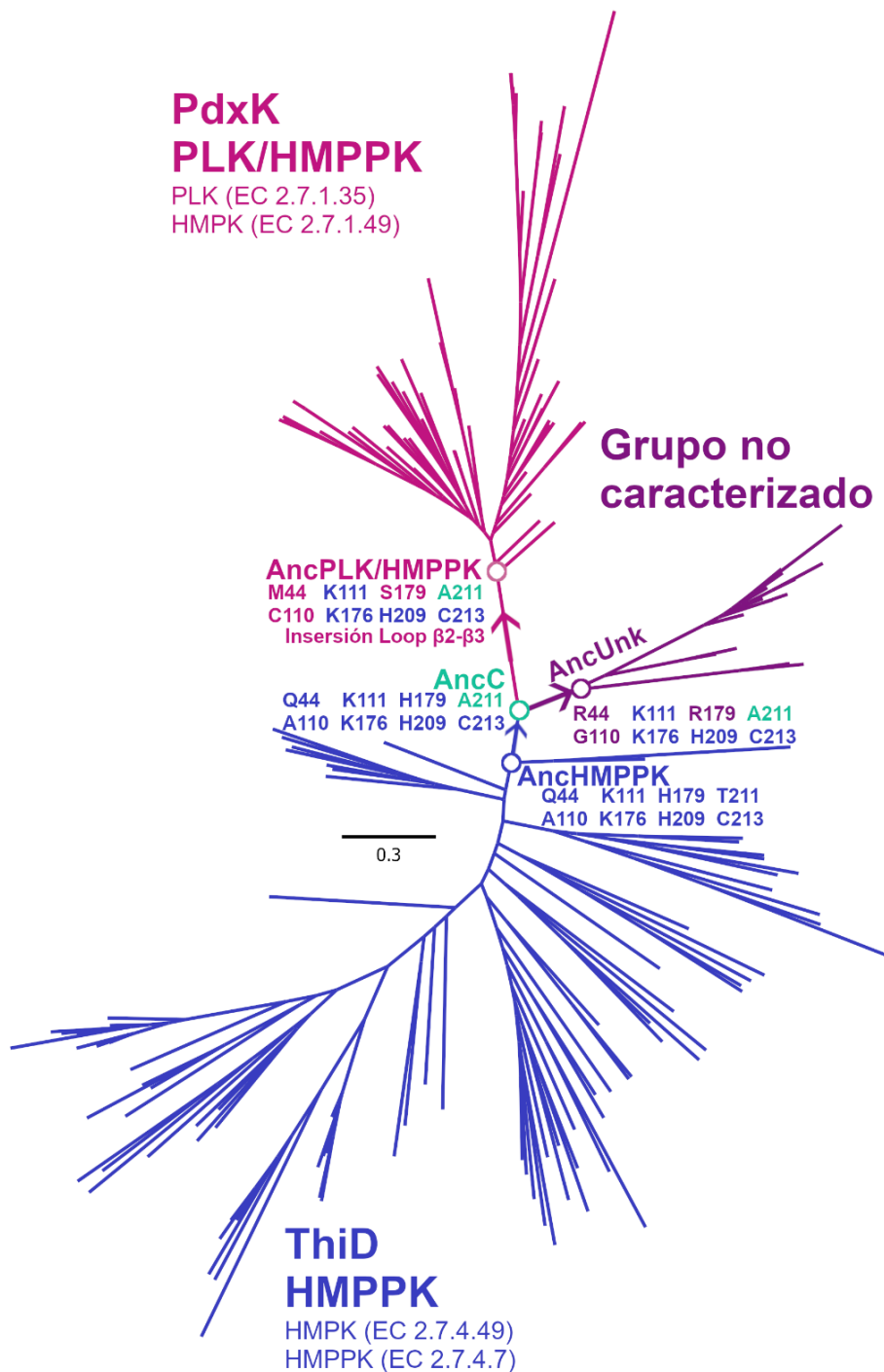


**Tabla 14: Probabilidades posteriores de residuos del sitio activo de las secuencias ancestrales reconstruidas involucrados en la unión de HMP-P.**

| Secuencia ancestral | Residuo       | PP            |
|---------------------|---------------|---------------|
| AncHMPPK            | Q44           | 1,000         |
|                     | A110          | 0,997         |
|                     | K111          | 1,000         |
|                     | G112          | 1,000         |
|                     | H179          | 0,999         |
|                     | T208          | 1,000         |
|                     | H209          | 1,000         |
|                     | G210          | 1,000         |
|                     | T211          | 0,999         |
|                     | G212          | 1,000         |
|                     | C213          | 1,000         |
|                     | V254          | 0,623         |
| AncC                | Q44           | 0,862         |
|                     | A110          | 0,760         |
|                     | K111          | 1,000         |
|                     | G112          | 0,993         |
|                     | H179          | 0,507         |
|                     | T208          | 1,000         |
|                     | H209          | 0,999         |
|                     | G210          | 1,000         |
|                     | A211          | 0,985         |
|                     | G212          | 1,000         |
|                     | C213          | 1,000         |
|                     | V254          | 0,848         |
| AncUnk              | R44           | 0,466         |
|                     | G110          | 0,995         |
|                     | K111          | 1,000         |
|                     | M112          | 0,989         |
|                     | R179          | 0,606         |
|                     | T208          | 1,000         |
|                     | S209          | 0,918         |
|                     | G210          | 1,000         |
|                     | A211          | 1,000         |
|                     | G212          | 1,000         |
|                     | C213          | 1,000         |
|                     | V254          | 0,867         |
| AncPLK/HMPPK        | M44           | 1,000         |
|                     | C111          | 1,000         |
|                     | K112          | 1,000         |
|                     | G113          | 1,000         |
|                     | S180          | 0,997         |
|                     | T210          | 1,000         |
|                     | H211          | 1,000         |
|                     | G212          | 1,000         |
|                     | A213          | 1,000         |
|                     | G214          | 1,000         |
|                     | C215          | 1,000         |
|                     | V256          | 0,977         |
|                     | Inserción K48 | 1,000*/0,727# |
|                     | Inserción N49 | 0,999*/0,682# |

\* Probabilidad posterior de que haya un aminoácido en dicha posición.

# Probabilidad posterior del aminoácido seleccionado



**Figura 21: Ruta evolutiva de las quinasas de vitaminas dependientes de ATP de bacterias.**

Se muestra filogenia por máxima verosimilitud de las quinasas de vitaminas dependientes de ATP de bacterias no enraizada. Se encuentran coloreados los distintos grupos identificados (ThiD-HMPPK en azul, PdxK-PLK/HMPPK en rosado y el grupo no caracterizado en morado). Los ancestros reconstruidos en esta tesis se presentan como círculos, mostrándose las sustituciones del bolsillo de unión a sustratos que se dieron durante la divergencia de los grupos.

## 4. Discusión

### 4.1. Estado oligomérico quinasas de vitaminas dependientes de ATP de bacterias

El estado de oligomerización de enzimas de la familia de las quinasas de vitaminas dependientes de ATP de bacterias ha sido motivo de discusión en varios trabajos de la literatura. Se han reportado distintos estados de oligomerización, encontrando desde monómeros a tetrámeros y también se ha propuesto la mezcla de distintos estados de oligomerización en solución, lo cual se resumen en la **Tabla 15**.

**Tabla 15: Masa molar y estados de oligomerización reportados para enzimas ThiD-HMPPK y PdxK-PLK/HMPPK.**

| Proteína            | Masa molar teórica (kDa) | Masa molar determinada (kDa)* | Estado de oligomerización reportado*           |
|---------------------|--------------------------|-------------------------------|--|
| <i>St</i> HMPPK     | 30,9                     | 33,9                          | Monómero <sup>1</sup> – Dímero <sup>#2</sup>   |
| <i>Tt</i> HMPPK     | 29,1                     | 30,7                          | Monómero <sup>1</sup>                          |
| <i>Ec</i> HMPPK     | 28,6                     | 31,0                          | Tetrámero <sup>3</sup> - Monómero <sup>4</sup> |
| <i>Bs</i> PLK/HMPPK | 29,0                     | ~40,0                         | Monómero o dímero en solución <sup>5</sup>     |

\* Cromatografía de exclusión molecular

# Análisis de interfaz de dimerización de una estructura cristalográfica

1 Cea P., 2019

2 Cheng y col., 2002

3 Reddick y col., 1998

4 Mizote y col., 1999

5 Newman y col., 2006a

Tanto para ThiD-HMPPK y PdxK-PLK/HMPPK se han reportado inconsistencias en el estado de oligomerización; por ejemplo, las enzimas ThiD-HMPPK de *Salmonella typhimurium* (*St*HMPPK) y *Thermus thermophilus* (*Tt*HMPPK) han sido reportadas como monómeros en experimentos de separación cromatográfica por exclusión molecular (Cea, 2019), no obstante, Cheng y col., 2002 ;proponen de modo teórico que *St*HMPPK sería una enzima dimérica basándose en los contactos cristalográficos, de acuerdo a la presencia de interfases de oligomerización identificadas en el servidor PISA (Krissinel & Henrick, 2007). Por otra parte, la ThiD-HMPPK de *Escherichia coli* se ha reportado como

una proteína tetramérica lo cual indicaría que las proteínas de la familia de quinasas de vitaminas presentarían propensión a oligomerizar. Las inconsistencias reportadas para las proteínas ThiD-HMPPK podría ser explicado por lo propuesto por Newman y colaboradores para BsPLK/HMPPK. Esta proteína presenta un tamaño ligeramente mayor al del monómero, pero menor al de un dímero, por lo que los autores sugieren que la proteína existiría en una mezcla de monómero-dímero en solución donde la cinética del equilibrio monómero-dímero sería más rápida que la cinética que toma el equilibrio de la separación por tamaño en la resina de exclusión molecular (Newman y col., 2006a).

Para AncEnHMPPK, los resultados de esta tesis sugieren que presenta dos poblaciones en solución (dímero y monómero, **Figura 9, Tabla 3 y Tabla 4**), favoreciéndose la formación del dímero a mayores concentraciones de proteína, y siendo posible la separación de una población semejante a un monómero mediante cromatografía de exclusión molecular, lo que sugiere una cinética rápida de dimerización. Además, analizando las estructuras cristalográficas de AncEnHMPPK en el servidor PISA, se encontraron interfases de dimerización análogas a las de S $\dagger$ HMPPK (**Figura suplementaria 2 y Tabla suplementaria 4**), las cuales también están en homólogos de la rama de las PdxK-PLK/HMPPK y PdxK PLK canónicas (que son homólogos más lejanos, **Tabla suplementaria 5**). Sin embargo, sorpresivamente esta interfaz de dimerización no se encuentra en las enzimas tiazol quinasas (THZK), las cuales han sido reportadas como trímeros en solución, como el caso de la THZK de *Staphylococcus aureus* (Drebes y col., 2016), donde mediante SAXS se reconstruyó y ajustó un modelo *ab initio* para un trímero.

Lo anterior es sumamente interesante dado que durante la evolución de esta familia de proteínas se presentaron cambios, tanto en el estado de oligomerización como en la especificidad por sustrato, por lo cual una posible línea de trabajo a futuro sería evaluar a lo largo de la evolución cuales fueron las variaciones en factores estructurales que conllevaron al cambio del estado oligomérico entre las tiazol quinastas y la rama de las piridoxal quinastas e hidroximetil pirimidina quinastas.

En base a los resultados de esta tesis y a lo propuesto por Newman y colaboradores, se puede concluir que AncEnHMPPK se encuentra como una mezcla monómero-dímero en solución, sugiriendo una cinética rápida de dimerización, la cual favorece la formación del dímero al incrementar la concentración de proteína en solución. Lo anterior explicaría las diferencias de tamaños determinados por cromatografía de exclusión molecular respecto a las estimaciones teóricas para un monómero o dímero de esta enzima.

#### **4.2. Estructuras cristalinas de AncEnHMPPK en presencia de sustratos**

La información cristalográfica de las enzimas de la familia de las quinastas de vitaminas dependientes de ATP es muy variada, donde se ha podido describir la arquitectura de estas proteínas, la cual está acorde a lo reportado para la superfamilia riboquinasa. En el caso de las PdxK-PLK/HMPPK (**Figura 4**), se han reportado estructuras cristalográficas co-cristalizadas con diferentes ligandos, pero en el caso de la rama de las ThiD-HMPPK no existía información estructural disponible de complejos ternarios con HMP y HMP-P, estando solo disponibles estructuras en conformación APO y co-cristalizadas con HMP.

En esta tesis se reportan las primeras estructuras cristalinas para una enzima ThiD-HMPPK co-cristalizada en complejo ternario con HMP y HMP-P junto a análogos no hidrolizables de ADP y ATP (**Figura suplementaria 6**). Las estructuras cristalográficas

aquí reportadas presentan la arquitectura común reportada para quinasas de vitaminas dependientes de ATP (**Figura suplementaria 7**), la cual está ampliamente conservada a nivel de superfamilia.

Por otro lado, el uso de análogos de moléculas de nucleótidos permite capturar distintas etapas de la catálisis o conformaciones dependiendo el tipo de nucleótido usado (análogos de ADP o ATP) como se ha reportado en la revisión de Lacabanne et al., 2020 y en particular para el caso de la DnaB helicasa y diversos transportadores ABC. Para AncEnHMPPK se obtuvieron estructuras cristalográficas de distintos complejos ternarios que dan cuenta de diferentes etapas de reacción para ambas actividades enzimáticas (HMPK y HMPPK). Por ejemplo, el complejo HMP/Mg-AMP-PCP da cuenta de la etapa pre-catalítica de la reacción HMPK (complejo enzima-sustratos), mientras que el complejo HMP-P/ADP<sub>βS</sub> da cuenta de la etapa post-catalítica (complejo enzima-productos), análogamente los complejos HMP-P/análogos de ATP dan cuenta de la etapa pre-catalítica de la actividad HMPPK (complejo enzima-sustratos).

Para tener una idea completa de todas las etapas de reacción para ambas actividades enzimáticas (HMPK/HMPPK) faltarían estructuras cristalográficas de complejos ternarios con análogos del estado de transición de la reacción. Para ello, se puede emplear Mg-ADP<sub>βS</sub>/AlF<sub>3</sub> como análogo de nucleótido para ambas actividades enzimáticas, donde la molécula de AlF<sub>3</sub> es un análogo trigonal plano del fosfato  $\gamma$  que se transfiere desde el ATP hacia el sustrato aceptor de fosforilo. Además, faltaría el complejo enzima-producto para la actividad HMPPK para el cual debería utilizarse HMP-PP y ADP<sub>βS</sub> como ligandos.

Por tanto, como una proyección respecto a los resultados descritos en este capítulo sería intentar obtener las estructuras cristalinas de los complejos análogos del estado de

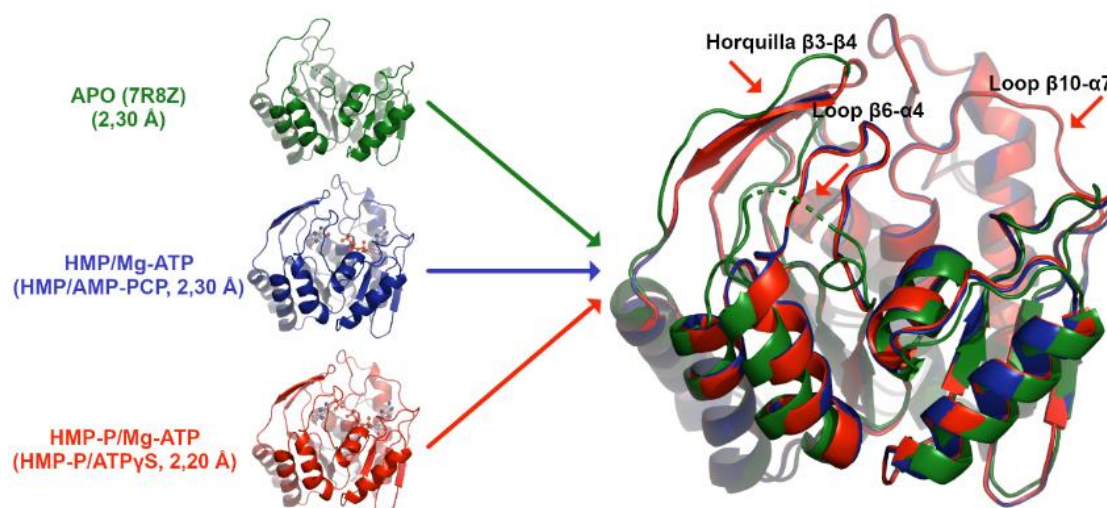
transición de ambas reacciones y del complejo enzima-productos para la proteína AncEnHMPPK, con lo cual será posible completar las etapas de la transferencia del grupo fosfato hacia el grupo hidroximetil o metil fosfato como aceptores de fosforilo. De esta manera, se podrían validar los aminoácidos identificados tanto en el proceso de transferencia del fosforilo, como los residuos que estabilizarían al grupo pirofosfato de la molécula de HMP-PP en sitio activo de las enzimas ThiD-HMPPK.

#### **4.3. Reordenamiento del sitio activo de quinasas de vitaminas dependientes de ATP y segunda fosforilación de HMP**

La capacidad de las enzimas ThiD-HMPPK de catalizar la fosforilación de un grupo metil fosfato es algo que no ha sido esclarecido respecto a su estructura como su mecanismo. Cheng y col proponen que la enzima StHMPPK presentarían una reorganización de loops del sitio activo entre las actividades HMPK y HMPPK, existiendo dos conformaciones del sitio activo. Lo anterior lo sugieren a partir de la información obtenida de la estructura co-cristalizada con HMP y a la presencia de iones fosfato en el sitio activo, por lo cual, estas distintas conformaciones del sitio activo serían la clave para que ambas reacciones de fosforilación ocurran en un mismo andamio estructural.

En esta tesis se determinaron las estructuras cristalográficas de una enzima ThiD-HMPPK ancestral en complejos ternarios con HMP y HMP-P junto a análogos no hidrolizables de ATP (**Tabla 8, Tabla 9 y Figura suplementaria 6**). Al comparar estas estructuras cristalográficas con la conformación APO de la enzima ancestral (PDBID:7R8Z, **Figura 22**), se observa que hay cambios conformacionales entre la enzima APO y la enzima en complejos ternarios, donde se destacan principalmente un reordenamiento de la horquilla  $\beta 3$ - $\beta 4$  y un ordenamiento de los loops  $\beta 6$ - $\alpha 4$  y  $\beta 10$ - $\alpha 7$ , loops que no son posible resolver totalmente en la enzima APO. Sorprendentemente, no

hay diferencias entre los complejos enzima-sustrato (HMP/Mg-ATP y HMP-P/Mg-ATP) o todas las estructuras enzimas-sustrato aquí reportadas para AncEnHMPPK (**Figura suplementaria 6 y Figura 13**), estando los elementos de estructura secundaria y loops en una misma conformación para todas las estructuras de complejos ternarios aquí determinadas.



**Figura 22: Cambios conformacionales inducidos por unión a ligandos en AncEnHMPPK.**

Se presentan tres estructuras de AncEnHMPPK: **1.** Conformación APO (verde, PDBID: 7R8Z). **2.** En complejo ternario con HMP y Mg-ATP (azul). **3.** En complejo ternario con HMP-P y Mg-ATP (rojo). Las principales diferencias en conformaciones de elementos de estructura secundaria se señalan con flechas de color rojo.

Por lo tanto, tras comparar las distintas estructuras de ThiD-HMPPK, se aprecia que no hay cambios conformacionales de elementos del sitio activo entre las actividades que catalizan esta rama de la familia (HMPK, HMPPK), por lo cual la propuesta de distintas conformaciones del sitio activo propuesta por Cheng no es consistente con la evidencia aquí presentada.

Entonces, otra posible explicación para la capacidad de las ThiD-HMPPK de fosforilar un grupo metil fosfato podría ser la reorientación de cadenas laterales de residuos del

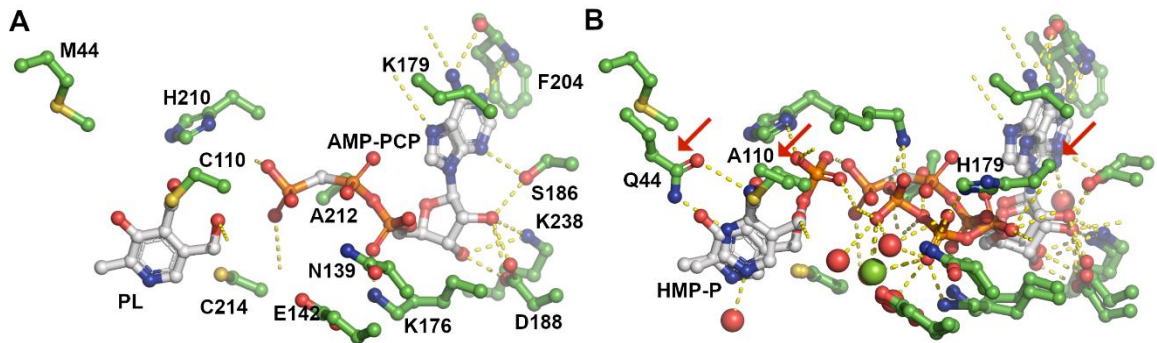


sitio activo que contribuyan aportando interacciones que estabilicen los cuatro grupos fosfato (uno del HMP-P y tres del ATP), esto en ausencia de cambios conformacionales de loops cercanos al sitio activo como demostrado más arriba. Para tener una idea más amplia de como ocurriría la actividad HMPPK en las ThiD-HMPPK, realizamos la comparación de los bolsillos de unión de AncEnHMPPK en complejo ternario con HMP y HMP-P. encontrando que: **1.** Hay varios residuos en conformaciones similares como: Lys111, Gln139, Glu142, Lys176, Ser185, Asp187, His209, Cys213 y Lys237, lo que sugiere que estarían cumpliendo funciones análogas en ambas reacciones enzimáticas. **2.** Los ligandos HMP/HMP-P, análogos de ATP y ión magnesio se encuentran orientados de manera equivalente en los sitios activos y con conformaciones similares, en el caso de la molécula de ATP en ambos casos evaluados la ribosa se encuentra en conformación C2'-endo y C3'-exo (Banfalvi, 2020; Evich y col., 2017), Por otra parte, el ión magnesio en complejo con ATP (HMP/Mg-AMP-PCP) se encuentra en configuración “*bidentate*” o C2 (Buelens y col., 2021), lo que hace que los grupos fosfato  $\beta$  y  $\gamma$  aporten átomos de oxígeno a la coordinación del ion magnesio. Para el complejo ternario con HMP-P/Mg-análogo de ATP no fue posible asignar el ión magnesio junto a la molécula de ATP, pero en base a la información estructural del complejo HMP/AMP-PCP fue posible reconstruir el complejo Mg-ATP en configuración “*bidentate*” o C2 como se indicó en la sesión de materiales y métodos (2.2.8). **3.** Interesantemente, las diferencias entre el sitio activo en presencia de HMP y HMP-P son la reorientación de las cadenas laterales de los residuos His179 y Thr211 que se acercan a los grupos fosfatos del ATP en el complejo con HMP-P/Mg-ATP, lo cual permitiría estabilizar los cuatro grupos fosfatos en el sitio activo para la actividad HMPPK, siendo esta dupla de residuos

probablemente muy importante para la actividad HMPPK. Para verificar esto será necesario realizar las mutaciones puntuales de estos residuos.

Para dilucidar los factores estructurales asociados a la actividad PLK, analizamos el sitio activo de la PdxK-PLK/HMPPK de *S. aureus* (SaPLK/HMPPK, PDBID: 4C5N, **Figura 23**, panel A), para compararlo con el bolsillo de unión de una ThiD-HMPPK en complejo con HMP-P y validar los residuos mencionados previamente. Del análisis, encontramos que los residuos de aminoácidos Lys111, Gln139, Glu142, Lys176, Ser185, Asp187, His209, Cys213 y Lys237 (numeración de AncEnHMPPK) están conservados entre PdxK-PLK/HMPPK y ThiD-HMPPK, cumpliendo roles análogos respecto a la catálisis y unión de PL y HMP-P, respectivamente

Por otra parte, los residuos His179 y Thr211 serían fundamentales para la actividad HMPPK, dado que estas interacciones no están presentes en SaPLK/HMPPK, donde el residuo His179 de AncEnHMPPK se encuentra sustituido por una lisina que apunta en una dirección contraria a donde se localizan los grupos fosfatos del ATP (**Figura 23**, panel B). Así también, la Thr211 de AncEnHMPPK se encuentra reemplazada por una alanina en SaPLK/HMPPK.



**Figura 23: Bolsillo de unión a sustrato de SaPLK/HMPPK.**

**(A)** Bolsillo de unión a sustratos de SaPLK/HMPPK en presencia de piridoxal y AMP-PCP. **(B)** Superposición de los bolsillos de unión a ligandos de SaPLK/HMPPK y AncEnHMPPK en complejo ternario con HMP/HMP-P y Mg-ATP. Las principales diferencias se marcan como flechas de color rojo.

#### 4.4. Historia evolutiva de la actividad HMP-P quinasa de bacterias

Al analizar la familia de quinazas de vitaminas (**Figura 4**) se aprecia que la actividad PLK aparece dos veces durante la evolución de esta familia (grupo de PdxK/PdxY PLK y PdxK-PLK/HMPPK), según lo propuesto por Cea y col., 2020, habiendo una convergencia de función, con residuos distintos en los sitios activos, pero con cierta analogía dada por: **1.** la formación de un intermediario covalente con el grupo aldehído del piridoxal. En el caso de las PdxK-PLK/HMPPK se generaría un hemitioacetal con un residuo de cisteína (Nodwell y col., 2014), mientras que en las PdxK PLK se generaría una base de Shift entre un residuo de lisina y el aldehído del PL (Deka y col., 2019). **2.** el residuo del sitio activo que actúa como base catalítica corresponde a un residuo de cisteína en las PdxK-PLK/HMPPK y a un aspartato en las PdxK PLK.

Al trazar la ruta evolutiva de la actividad PLK en la rama de las PdxK-PLK/HMPPK mediante la reconstrucción de secuencias ancestrales (**Sección 3.9, Figura 21**) pudimos identificar sustituciones claves que habrían dado cuenta de la aparición de la actividad piridoxal quinasa, lo que fue concomitante con la pérdida de la actividad HMPPK.

Siguiendo la ruta de aparición de actividad PLK, el último ancestro en común de las ThiD-HMPPK (AncHMPPK) presenta todos los residuos del sitio activo de las ThiD-HMPPK comúnmente descritas, en cambio su descendiente directo, AncC, solo presenta la sustitución T211A en su sitio activo, rasgo que es heredado al siguiente descendiente, AncPLK/HMPPK. Este ancestro presenta las siguientes sustituciones respecto al AncC: Q44M, A110C y H179S, junto a la inserción de dos residuos en la horquilla  $\beta$ 3- $\beta$ 4. También es importante destacar que AncPLK/HMPPK presenta todos los residuos importantes descritos para las PdxK-PLK/HMPPK actuales.

De los residuos mencionados anteriormente, la sustitución Q44M fue evaluada en el trabajo de Gonzalez-Ordenes y col, (2021) donde mostraron que dicha sustitución en ancC provoca un aumento en la afinidad por PL (disminución en  $K_M$ ) apoyando su rol en la aparición de la actividad PLK esta ruta evolutiva. También, la sustitución A110C ha sido evaluada en la proteína actual SaPLK/HMPPK, donde la mutante C111A (sustitución reversa) genera una enzima inactiva con PL como sustrato, pero la actividad con HMP como sustrato no se ve afectada (Nodwell y col., 2014).

En conclusión, podemos decir que corroboramos la primera parte de la hipótesis de trabajo. A través de la caracterización de las relaciones evolutivas entre ThiD-HMPPK y PdxK-PLK/HMPPK, encontramos que las PdxK-PLK/HMPPK habrían divergido a partir de las ThiD-HMPPK, lo que indica que la actividad HMPPK habría sido el rasgo ancestral de la familia. Además, la pérdida de la actividad HMPPK sería concomitante con la aparición de la actividad PLK en las enzimas PdxK-PLK/HMPPK, debido a la pérdida de los residuos críticos para la actividad HMPPK (H179S y T211A) concomitante con la aparición de los residuos críticos para la actividad PLK (Q44M y A110C). Por otro lado, la segunda parte de la hipótesis es correcta, ya que la sustitución T211A habría estado

involucrada en la pérdida de la actividad HMPPK, mientras que mutaciones en la horquilla  $\beta$ 3- $\beta$ 4 (Ej. Q44M e inserciones) habrían sido claves para la aparición de la actividad PLK. Sin embargo, inicialmente la hipótesis no consideró la relevancia del residuo H179, cuya sustitución también sería clave para la pérdida de la actividad HMPPK. No obstante, es necesario realizar trabajos de caracterización cinética de mutantes sitio-dirigidas para los residuos claves aquí identificados, para comprender en profundidad como habría sido la pérdida de la actividad HMPPK y la aparición de la actividad PLK en esta familia de enzimas.

## 5. Conclusiones

- La proteína AncEnHMPPK está en un equilibrio monómero-dímero en solución, el que depende de la concentración de proteína.
- En la enzima AncEnHMPPK la fosforilación de la molécula de HMP-P no involucra un cambio conformacional de los loops adyacentes al bolsillo de unión.
- Los residuos His179 y Thr211 de AncEnHMPPK serían cruciales y específicos para la fosforilación de HMP-P en las ThiD-HMPPK, los cuales presentan un cambio en orientación en presencia de HMP-P en el sitio activo, estabilizando los fosfatos  $\beta$  y  $\gamma$  de la molécula de ATP.
- El análisis evolutivo concuerda con lo propuesto sobre que la segunda fosforilación en la molécula de HMP sería el rasgo ancestral en esta rama de la familia, siendo la fosforilación por PL la novedad evolutiva en las PdxK-PLK/HMPPK.
- La reconstrucción de secuencias ancestrales (ARS) evidenció mutaciones relevantes durante la evolución de la rama de las PdxK-PLK/HMPPK. Las sustituciones T211A y H179S estarían involucradas en la pérdida de la actividad HMP-P quinasa, mientras que cambios en la hebra beta3-beta4, la sustitución Gln44Met y la inserción de dos residuos, sumado a las sustituciones A110C y T211A, estarían involucradas en la aparición de la actividad PLK en las PdxK-PLK/HMPPK.

## 6. Bibliografía

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., & Zwart, P. H. (2010). PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, 66(2), 213–221. <https://doi.org/10.1107/S0907444909052925>

Banfalvi, G. (2020). Ribose Selected as Precursor to Life. *DNA and Cell Biology*, 39(2), 177–186. <https://doi.org/10.1089/dna.2019.4943>

Bradford, M. M., Dong, Y., Xu, L., Liu, S., Bai, X., Bradford, M., Briskin, D., Leonard, R., Hodges, T., Buchanan, B., Gruissem, W., Jones, R., Bush, D., Cakmak, I., Marschner, H., Chen, W., Yang, J., Qin, C., Jin, C., ... Cao, Y. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, 72(1–2), 248–254. [https://doi.org/10.1016/0003-2697\(76\)90527-3](https://doi.org/10.1016/0003-2697(76)90527-3)

Buelens, F. P., Leonov, H., De Groot, B. L., & Grubmüller, H. (2021). ATP-Magnesium Coordination: Protein Structure-Based Force Field Evaluation and Corrections. *Journal of Chemical Theory and Computation*, 17(3), 1922–1930. <https://doi.org/10.1021/acs.jctc.0c01205>

Castro-Fernandez, V., Bravo-Moraga, F., Herrera-Morande, A., & Guixe, V. (2014). Bifunctional ADP-dependent phosphofructokinase/glucokinase activity in the order Methanococcales - Biochemical characterization of the mesophilic enzyme from *Methanococcus maripaludis*. *FEBS Journal*, 281(8), 2017–2029.

<https://doi.org/10.1111/febs.12757>

Castro-Fernandez, V., Bravo-Moraga, F., Ramirez-Sarmiento, C. A., & Guixe, V. (2014). Emergence of pyridoxal phosphorylation through a promiscuous ancestor during the evolution of hydroxymethyl pyrimidine kinases. *FEBS Letters*, *588*(17), 3068–3073. <https://doi.org/10.1016/j.febslet.2014.06.033>

Cea, P. A. (2019) Caracterización bioquímica de la 4-amino-5-hidroximetil-2-metilpirimidina quinasa de *Salmonella typhimurium* y *Thermus thermophilus*. Seminario de Título Ingeniero en Biotecnología Molecular. Santiago, Universidad de Chile, Facultad de Ciencias.

Cea, P. A., Araya, G., Vallejos, G., Recabarren, R., Alzate-Morales, J., Babul, J., Guixé, V., & Castro-Fernandez, V. (2020). Characterization of hydroxymethylpyrimidine phosphate kinase from mesophilic and thermophilic bacteria and structural insights into their differential thermal stability. *Archives of Biochemistry and Biophysics*, *688*(March), 108389. <https://doi.org/10.1016/j.abb.2020.108389>

Cheng, G., Bennett, E. M., Begley, T. P., & Ealick, S. E. (2002). Crystal structure of 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate kinase from *Salmonella typhimurium* at 2.3 Å resolution. *Structure*, *10*(2), 225–235. [https://doi.org/10.1016/S0969-2126\(02\)00708-6](https://doi.org/10.1016/S0969-2126(02)00708-6)

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2011). ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics*, *27*(8), 1164–1165. <https://doi.org/10.1093/bioinformatics/btr088>

Deka, G., Kalyani, J. N., Jahangir, F. B., Sabharwal, P., Savithri, H. S., & Murthy, M. R. N. (2019). Structural and functional studies on *Salmonella typhimurium* pyridoxal kinase:



the first structural evidence for the formation of Schiff base with the substrate. *FEBS Journal*, 286(18), 3684–3700. <https://doi.org/10.1111/febs.14933>

Dick, T., Manjunatha, U., Kappes, B., & Gengenbacher, M. (2010). Vitamin B6 biosynthesis is essential for survival and virulence of *Mycobacterium tuberculosis*. *Molecular Microbiology*, 78(4), 980–988. <https://doi.org/10.1111/j.1365-2958.2010.07381.x>

Drebes, J., Künz, M., Windshügel, B., Kikhney, A. G., Müller, I. B., Eberle, R. J., Oberthür, D., Cang, H., Svergun, D. I., Perbandt, M., Betzel, C., & Wrenger, C. (2016). Structure of ThiM from Vitamin B1 biosynthetic pathway of *Staphylococcus aureus* - Insights into a novel pro-drug approach addressing MRSA infections. *Scientific Reports*, 6(October 2015), 1–10. <https://doi.org/10.1038/srep22871>

Emsley, P., & Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*, 60(12 I), 2126–2132. <https://doi.org/10.1107/S0907444904019158>

Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography*, 69(7), 1204–1214. <https://doi.org/10.1107/S0907444913000061>

Evich, M., Spring-Connell, A. M., & Germann, M. W. (2017). Impact of modified ribose sugars on nucleic acid conformation and function. *Heterocyclic Communications*, 23(3), 155–165. <https://doi.org/10.1515/hc-2017-0056>

Flachner, B., Kovári, Z., Varga, A., Gugolya, Z., Vonderviszt, F., Náray-Szabó, G., & Vas, M. (2004). Role of Phosphate Chain Mobility of MgATP in Completing the 3-Phosphoglycerate Kinase Catalytic Site: Binding, Kinetic, and Crystallographic Studies

with ATP and MgATP. *Biochemistry*, 43(12), 3436–3449.  
<https://doi.org/10.1021/bi035022n>

Gonzalez-Ordenes, F., Bravo-Moraga, F., Gonzalez, E., Hernandez-Cabello, L., Alzate-Morales, J., Guixé, V., & Castro-Fernandez, V. (2021). Crystal structure and molecular dynamics simulations of a promiscuous ancestor reveal residues and an epistatic interaction involved in substrate binding and catalysis in the ATP-dependent vitamin kinase family members. *Protein Science*, 30(4), 842–854.  
<https://doi.org/10.1002/pro.4040>

Grubman, A., Phillips, A., Thibonnier, M., Kaparakis-Liaskos, M., Johnson, C., Thiberge, J. M., Radcliff, F. J., Ecobichon, C., Labigne, A., de Reuse, H., Mendz, G. L., & Ferrero, R. L. (2010). Vitamin B6 is required for full motility and virulence in *Helicobacter pylori*. *MBio*, 1(3), 1–9. <https://doi.org/10.1128/mBio.00112-10>

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, 59(3), 307–321.  
<https://doi.org/10.1093/sysbio/syq010>

Hall T. (2007). BioEdit 7.0.5.3 Department of Microbiology, North Carolina State University. Available from <Http://Www.Mbio.Ncsu.Edu/BioEdit/Bioedit.Html>.  
<http://ci.nii.ac.jp/naid/10024040332/en/>

Hanson-Smith, V., Kolaczkowski, B., & Thornton, J. W. (2010). Robustness of ancestral sequence reconstruction to phylogenetic uncertainty. *Molecular Biology and Evolution*, 27(9), 1988–1999. <https://doi.org/10.1093/molbev/msq081>

Huang, Y., Niu, B., Gao, Y., Fu, L., & Li, W. (2010). CD-HIT Suite: A web server for

clustering and comparing biological sequences. *Bioinformatics*, 26(5), 680–682.  
<https://doi.org/10.1093/bioinformatics/btq003>

Kabsch, W. (2010). research papers XDS research papers. *Acta Crystallographica Section D: Biological Crystallography*, 66(2), 125–132.  
<https://doi.org/10.1107/S0907444909047337>

Kantardjieff, K. A., & Rupp, B. (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Science*, 12(9), 1865–1871. <https://doi.org/10.1110/ps.0350503>

Krissinel, E., & Henrick, K. (2007). Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology*, 372(3), 774–797.  
<https://doi.org/10.1016/j.jmb.2007.05.022>

Lacabanne, D., Wiegand, T., Wili, N., Kozlova, M. I., Cadalbert, R., Klose, D., Mulkidjanian, A. Y., Meier, B. H., & Böckmann, A. (2020). ATP Analogues for Structural Investigations: Case Studies of a DnaB Helicase and an ABC Transporter. *Molecules (Basel, Switzerland)*, 25(22). <https://doi.org/10.3390/molecules25225268>

Le, S. Q., & Gascuel, O. (2008). An improved general amino acid replacement matrix. *Molecular Biology and Evolution*, 25(7), 1307–1320.  
<https://doi.org/10.1093/molbev/msn067>

Lünse, C. E., Scott, F. J., Suckling, C. J., & Mayer, G. (2014). Novel TPP-riboswitch activators bypass metabolic enzyme dependency. *Frontiers in Chemistry*, 2(JUL), 1–8.  
<https://doi.org/10.3389/fchem.2014.00053>

Maddison, W. P., & Maddison, D. R. (2019). Mesquite: a modular system for evolutionary

analysis, v. 3.61. Available from <Http://Mesquiteproject.Org>.

Manalastas-Cantos, K., Konarev, P. V., Hajizadeh, N. R., Kikhney, A. G., Petoukhov, M. V., Molodenskiy, D. S., Panjkovich, A., Mertens, H. D. T., Gruzinov, A., Borges, C., Jeffries, C. M., Svergun, D. I., & Franke, D. (2021). ATSAS 3.0: Expanded functionality and new tools for small-angle scattering data analysis. *Journal of Applied Crystallography*, *54*, 343–355. <https://doi.org/10.1107/S1600576720013412>

Merino, F., Rivas-Pardo, J. A., 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*(2), 516–524. <https://doi.org/10.1016/j.biochi.2011.08.021>

Mizote, T., S Smith, D. D., Nakayama, H., & Nakazawa, T. (n.d.). Cloning and characterization of the thiDIJ gene of *Escherichia coli* encoding a thiamin-synthesizing bifunctional enzyme, hydroxymethyl pyrimidine kinase/phosphomethylpyrimidine kinase. In *Microbiology* (Vol. 145).

Munoz, S. M., Fernandez, V. C., & Guixé, V. (2021). Structure of an ancestral ADP-dependent kinase with fructose-6P reveals key residues for binding, catalysis, and ligand-induced conformational changes. *Journal of Biological Chemistry*, *296*(1), 100219. <https://doi.org/10.1074/jbc.RA120.015376>

Newman, J. A., Das, S. K., Sedelnikova, S. E., & Rice, D. W. (2006a). Cloning, purification and preliminary crystallographic analysis of a putative pyridoxal kinase from *Bacillus subtilis*. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, *62*(10), 1006–1009. <https://doi.org/10.1107/S1744309106035779>

Newman, J. A., Das, S. K., Sedelnikova, S. E., & Rice, D. W. (2006b). The Crystal

Structure of an ADP Complex of *Bacillus subtilis* Pyridoxal Kinase Provides Evidence for the Parallel Emergence of Enzyme Activity During Evolution. *Journal of Molecular Biology*, 363(2), 520–530. <https://doi.org/10.1016/j.jmb.2006.08.013>

Nodwell, M. B., Koch, M. F., Alte, F., Schneider, S., & Sieber, S. A. (2014). A subfamily of bacterial ribokinases utilizes a hemithioacetal for pyridoxal phosphate salvage. *Journal of the American Chemical Society*, 136(13), 4992–4999. <https://doi.org/10.1021/ja411785r>

Park, J. H., Burns, K., Kinsland, C., & 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. *Journal of Bacteriology*, 186(5), 1571–1573. <https://doi.org/10.1128/JB.186.5.1571-1573.2004>

Ramírez-Sarmiento, C. A., Engelberger, F., & Guixé, V. (2020). An evolutionary marker of the ribokinase superfamily is responsible for zinc-mediated regulation of human pyridoxal kinase. *Catalysts*, 10(5). <https://doi.org/10.3390/catal10050555>

Reddick, J. J., Kinsland, C., Nicewonger, R., Christian, T., Downs, D. M., Winkler, M. E., & 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. In *Tetrahedron* (Vol. 54).

Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, 42(W1), 320–324. <https://doi.org/10.1093/nar/gku316>

Roberts, E., Eargle, J., Wright, D., & Luthey-Schulten, Z. (2006). MultiSeq: Unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics*, 7, 1–11. <https://doi.org/10.1186/1471-2105-7-382>

Russell, R. B., & Barton, G. J. (1992). Multiple protein sequence alignment from tertiary structure comparison: Assignment of global and residue confidence levels. *Proteins: Structure, Function, and Bioinformatics*, 14(2), 309–323. <https://doi.org/10.1002/prot.340140216>

Schupfner, M., Straub, K., Busch, F., Merkl, R., & Sterner, R. (2020). Analysis of allosteric communication in a multienzyme complex by ancestral sequence reconstruction. *Proceedings of the National Academy of Sciences of the United States of America*, 117(1), 346–354. <https://doi.org/10.1073/pnas.1912132117>

Sigrell, J. A., Cameron, A. D., Jones, T. A., & Mowbray, S. L. (1998). *1-s2.0-S0969212698000203-main.pdf*. 183–193.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22(22), 4673–4680. <https://doi.org/10.1093/nar/22.22.4673>

Vagin, A., & Teplyakov, A. (2010). Molecular replacement with MOLREP. *Acta Crystallographica Section D: Biological Crystallography*, 66(1), 22–25. <https://doi.org/10.1107/S09074444909042589>

Walker, J. M., Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). *Protein Analysis Tools on the ExPASy Server 571 571 From: The Proteomics Protocols Handbook Edited Protein Identification and Analysis Tools on*

*the ExPASy Server*. 571–607. <http://www.expasy.org/tools/>.

Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B., Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S., & Richardson, D. C. (2018). MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science*, *27*(1), 293–315. <https://doi.org/10.1002/pro.3330>

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*(8), 1586–1591. <https://doi.org/10.1093/molbev/msm088>

Zhang, Y., Dougherty, M., Downs, D. M., & Ealick, S. E. (2004). Crystal structure of an aminoimidazole riboside kinase from *Salmonella enterica*: Implications for the evolution of the ribokinase superfamily. *Structure*, *12*(10), 1809–1821. <https://doi.org/10.1016/j.str.2004.07.020>

## 7. Anexo

### 7.1. Información adicional

#### 7.1.1. Estructuras cristalográficas disponibles de ThiD-HMPPK y PdxK-PLK/HMPPK

Tabla suplementaria 1: Estructuras cristalográficas disponibles para enzimas PLK/HMPPK actuales.

| Código PDB | Especie                      | Actividad | Ligandos   |
|------------|------------------------------|-----------|------------|
| 2I5B       | <i>Bacillus subtilis</i>     | PLK/HMPK  | ADP        |
| 4C5J       | <i>Staphylococcus aureus</i> | PLK/HMPK  | APO        |
| 4C5K       | <i>Staphylococcus aureus</i> | PLK/HMPK  | ADP        |
| 4C5L       | <i>Staphylococcus aureus</i> | PLK/HMPK  | PL         |
| 4C5M       | <i>Staphylococcus aureus</i> | PLK/HMPK  | AMP-PCP    |
| 4C5N       | <i>Staphylococcus aureus</i> | PLK/HMPK  | PL-AMP-PCP |

Tabla suplementaria 2: Estructuras cristalográficas disponibles para enzimas ThiD-HMPPK actuales.

| Código PDB | Especie                        | Actividad  | Ligandos |
|------------|--------------------------------|------------|----------|
| 1JXH       | <i>Salmonella typhimurium</i>  | HMPK/HMPPK | APO      |
| 1JXI       | <i>Salmonella typhimurium</i>  | HMPK/HMPPK | HMP      |
| 1UB0       | <i>Thermus Thermophilus</i>    | HMPK/HMPPK | APO      |
| 4JJP       | <i>Clostridium difficile</i>   | HMPK/HMPPK | APO      |
| 4YL5       | <i>Acinetobacter baumannii</i> | HMPK/HMPPK | APO      |

### 7.2. Determinación del estado oligomérico de AncEnHMPPK

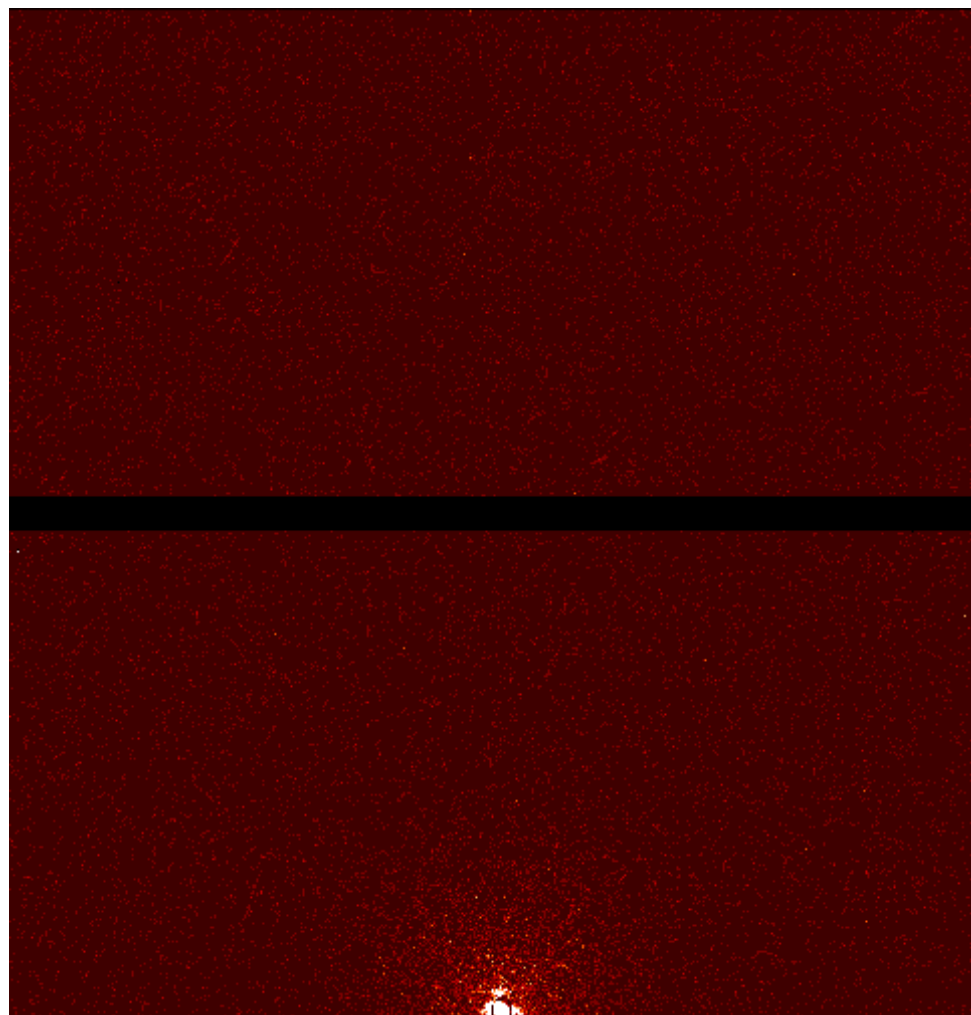
#### 7.2.1. Cromatografía de exclusión molecular mediante HPLC

Tabla suplementaria 3: Volúmenes de elución para estándar Biorad.

| Pico | Compuesto      | Masa Molar (Da) | V <sub>e</sub> (mL) | Área (%) |
|------|----------------|-----------------|---------------------|----------|
| 1    | Tiroglobulina  | 670000          | 11,647              | 14,288   |
| 2    | Gammaglobulina | 158000          | 11,787              | 21,732   |
| 3    | Ovoalbúmina    | 44000           | 19,705              | 11,082   |
| 4    | Mioglobina     | 17000           | 22,587              | 24,266   |
| 5    | Vitamina B12   | 1350            | 26,965              | 28,631   |

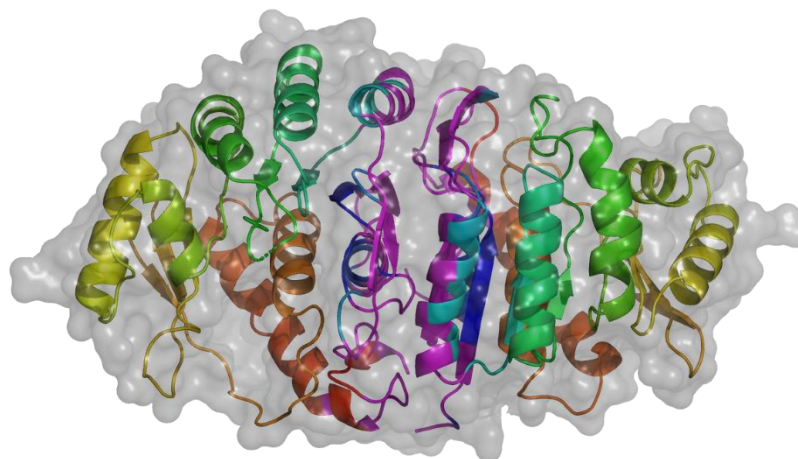


### 7.2.2. Dispersión de rayos X a bajo ángulo



**Figura suplementaria 1: Patrón de dispersión de SAXS representativo para AncEnHMPPK.**

### 7.2.3. Análisis de interfaz de dimerización mediante servidor PISA



#### Figura suplementaria 2: Interfaz de dimerización de AncEnHMPPK.

Se muestra el dímero de AncEnHMPPK como “cartoon” y su superficie. Se colorearon ambas cadenas (“cartoon”) con el espectro “rainbow” que denota direccionalidad desde extremo N-Terminal (azul) a C-Terminal (rojo), además, se destacó la interfaz de dimerización en color magenta. Los residuos que forman parte de la interfaz se detallan en la **Tabla suplementaria 4**.

#### Tabla suplementaria 4: Residuos de la interfaz de dimerización de AncEnHMPPK.

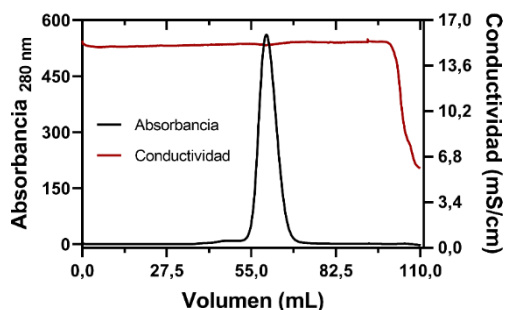
| Residuo | Cadena | Residuo | Cadena | Residuo | Cadena |
|---------|--------|---------|--------|---------|--------|
| His 0   | A-B    | Tyr 33  | A-B    | Asp 70  | A-B    |
| Met 1   | A-B    | Gly 34  | A-B    | Val 71  | A-B    |
| Lys 2   | A      | Cys 35  | A-B    | Arg 72  | A-B    |
| Arg 3   | A-B    | Ser 36  | A-B    | Gln 246 | A-B    |
| Ile 4   | A-B    | Val 37  | A-B    | Thr 249 | A-B    |
| Asn 5   | A-B    | Ile 38  | A-B    | Leu 250 | A-B    |
| Thr 12  | A-B    | Gln 50  | A-B    | Glu 251 | A-B    |
| Pro 14  | A-B    | Ser 51  | A-B    | Val 252 | A-B    |
| Ser 15  | A-B    | Val 52  | A-B    | Gly 253 | A-B    |
| Gly 16  | A-B    | Tyr 53  | A-B    | His 254 | A-B    |
| Gly 17  | A-B    | Gln 54  | A-B    | Pro 258 | A-B    |
| Ile 20  | A-B    | Ile 55  | A-B    | Val 259 | A-B    |
| Gln 21  | A-B    | Asp 56  | A-B    | His 260 | A-B    |
| Leu 24  | A-B    | Phe 59  | A-B    | His 261 | A-B    |
| Lys 25  | A-B    | Ala 62  | A-B    | Phe 262 | A-B    |
| Ser 28  | A-B    | Gln 63  | A-B    | Trp 265 | A-B    |
| Ala 29  | A-B    | Ser 66  | A-B    | Trp 266 | A-B    |
| Gly 31  | A-B    | Val 67  | A-B    |         |        |
| Ala 32  | A-B    | Ser 69  | A-B    |         |        |

**Tabla suplementaria 5: Estructuras cristalográficas con interfaz de dimerización análoga a la de AncEnHMPPK.**

| Código PDB  | Proteína            | Especie                        |
|-------------|---------------------|--------------------------------|
| 1jxi-h      | ThiD-HMPPK          | <i>Salmonella typhimurium</i>  |
| 1ub0        | ThiD-HMPPK          | <i>Thermus Thermophilus</i>    |
| 2i5b        | PdxK-PLK/HMPPK      | <i>Bacillus subtilis</i>       |
| 4c5j-n      | PdxK-PLK/HMPPK      | <i>Staphylococcus aureus</i>   |
| 4ywr / 4yl5 | ThiD-HMPPK putativa | <i>Acinetobacter baumannii</i> |
| 5zwa-b      | PdxK PLK            | <i>Salmonella typhimurium</i>  |
| 2ddw        | PdxK PLK            | <i>Escherichia coli</i>        |
| 1td2        | PdxY PLK            | <i>Escherichia coli</i>        |
| 3fhy        | PdxK PLK            | <i>Homo sapiens</i>            |
| 6yk1        | PdxK PLK            | <i>Mus musculus</i>            |
| 4e69        | Quinasa de azúcar   | <i>Oceanicola granulosus</i>   |
| 2fv7        | Riboquinasa         | <i>Homo sapiens</i>            |
| 5zwy        | Riboquinasa         | <i>Leishmania donovani</i>     |

### 7.3. Cristalografía de proteínas

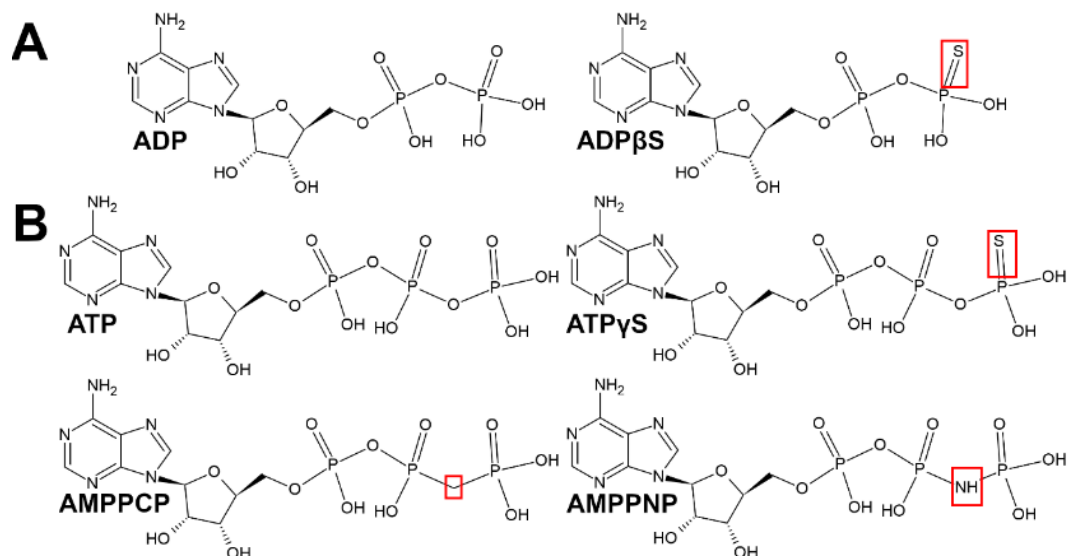
#### 7.3.1. Separación de poblaciones homogéneas para realizar ensayos de cristalización



**Figura suplementaria 3: Cromatografía de exclusión molecular preparativa de AncEnHMPPK para la separación de poblaciones homogéneas.**

En el eje izquierdo se indica la absorbancia registrada durante la cromatografía, mientras que el eje derecho indica la conductividad registrada durante el proceso.

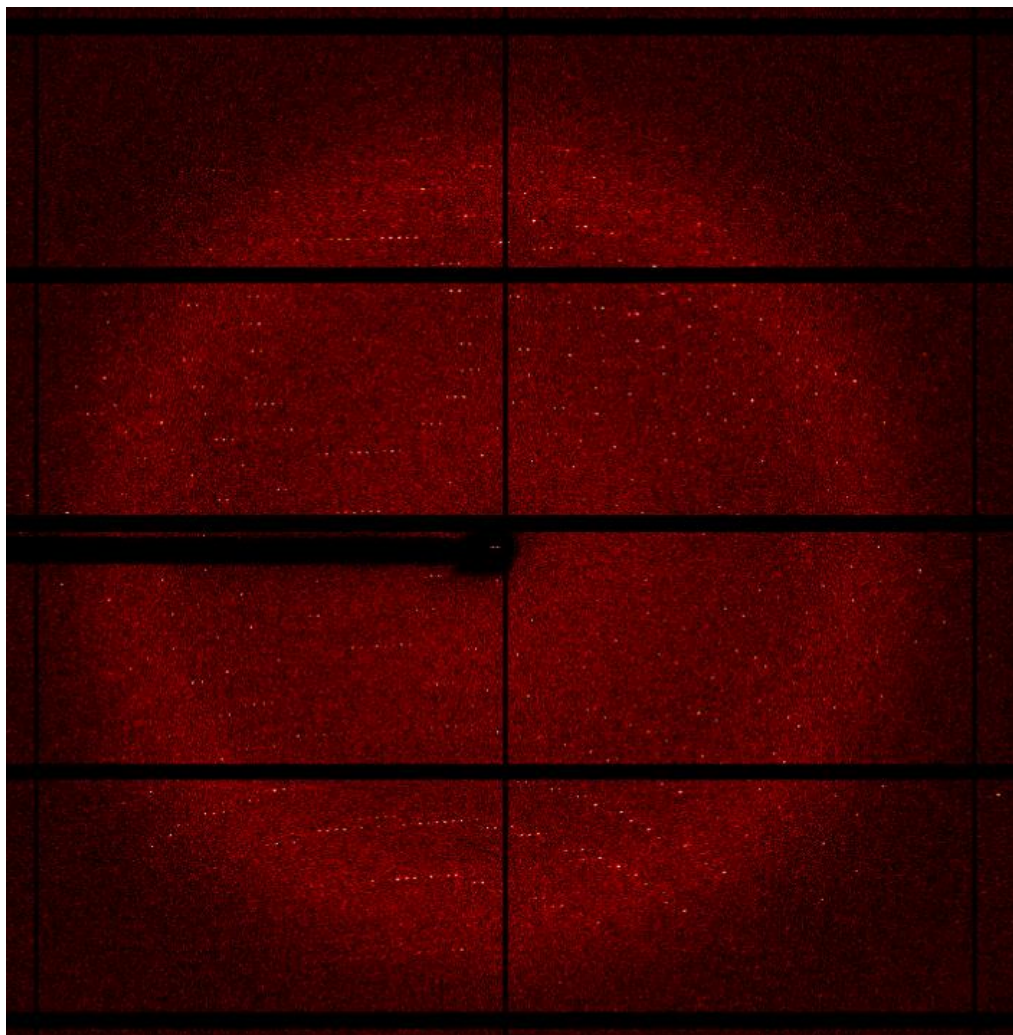
### 7.3.2. Ensayos de cristalización para la proteína ancestral AncEnHMPPK



**Figura suplementaria 4: Análogos no hidrolizables de ADP y ATP empleados para los ensayos de cristalización por gota sentada.**

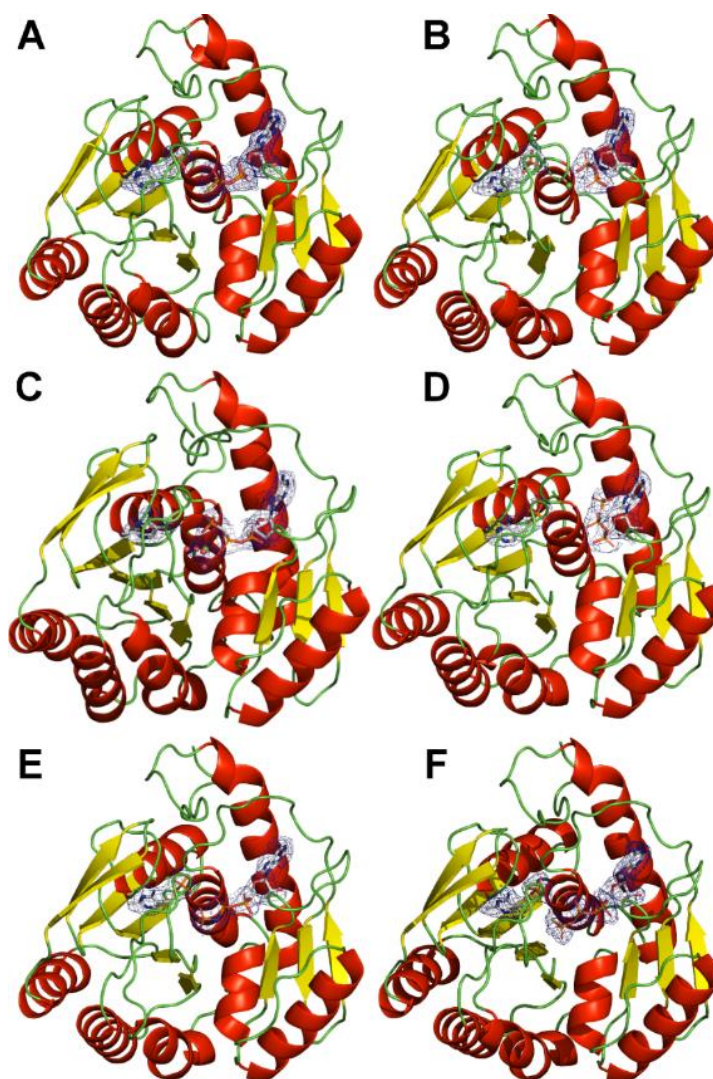
**(A)** Análogo de ADP. **(B)** Análogos de ATP. Las principales diferencias entre el análogo y la molécula de ADP y/o ATP se resaltan con un recuadro de color rojo.

**7.3.3. Patrón de difracción de rayos X representativo para los cristales de AncEnHMPPK.**



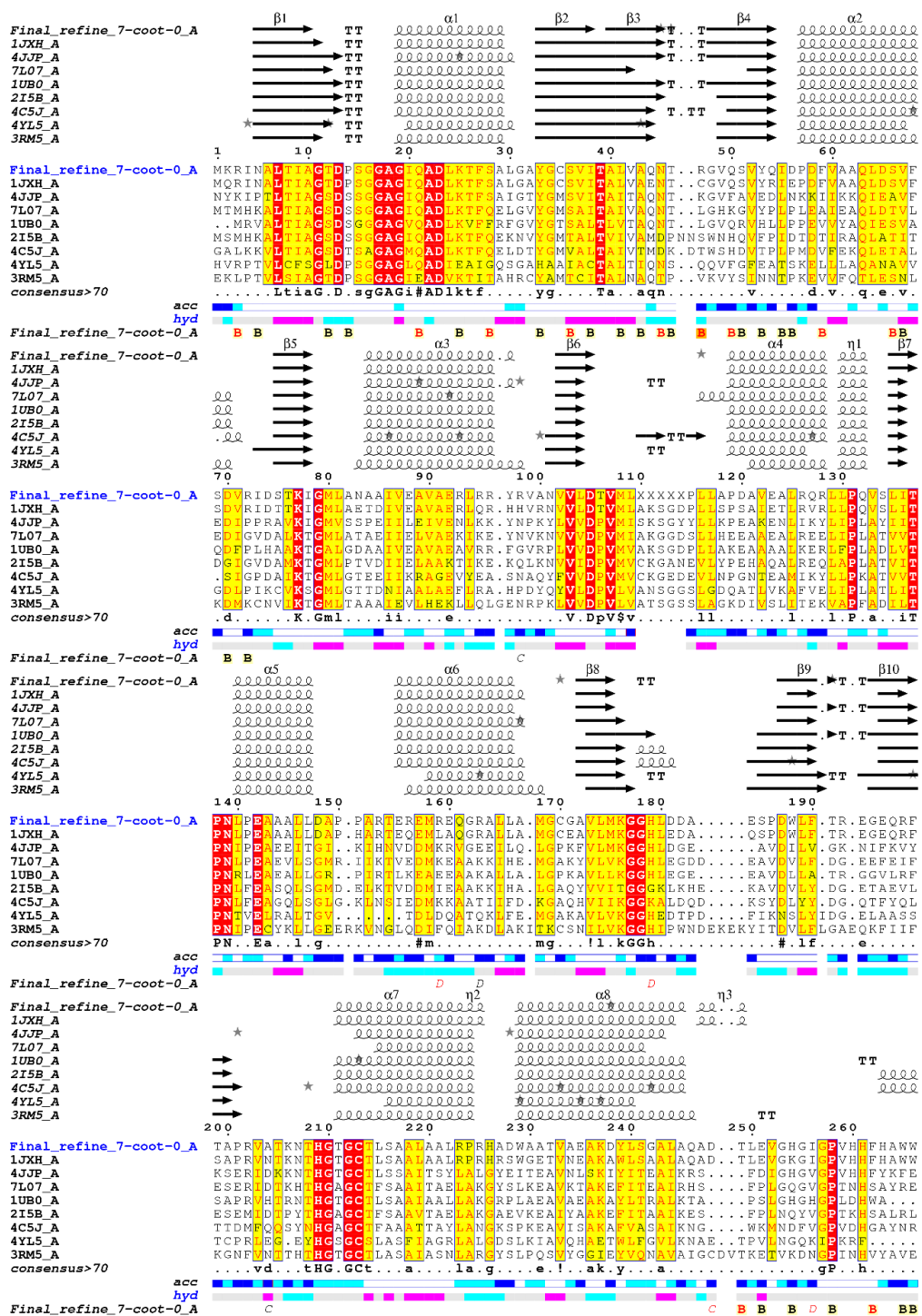
**Figura suplementaria 5: Patrón de difracción representativo de AncEnHMPPK.**

### 7.3.4. Análisis estructurales de quinasas de vitaminas dependientes de ATP de bacteria



#### Figura suplementaria 6: Estructuras cristalográficas de AncEnHMPPK.

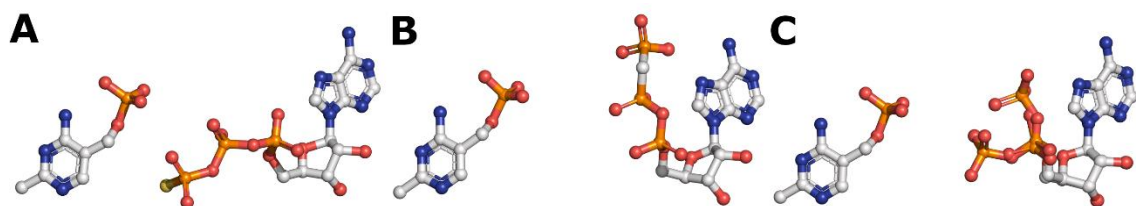
Modelos de estructuras en complejo ternario con ligandos. Se presentan los modelos tridimensionales coloreados por elementos de estructura secundaria (se presentan las  $\alpha$ -hélices en rojo, las hebras- $\beta$  en amarillo y los loops en verde). Los residuos y los ligandos del sitio activo se presentan como “stick balls” coloreados según átomos: Los átomos de carbono se coloraron en verde (proteína) y blanco (ligandos), los átomos de nitrógeno en azul, los átomos de oxígeno en rojo y los átomos de azufre en amarillo. Las líneas punteadas amarillas representan puentes de hidrógeno. **(A)** AncEnHMPPK en complejo con HMP-P y  $ADP_{\beta s}$ . **(B)** AncEnHMPPK en complejo con HMP-P y ADP. **(C)** AncEnHMPPK en complejo con HMP y Mg-AMP-PCP. **(D)** AncEnHMPPK en complejo con HMP-P y AMP-PCP. **(E)** AncEnHMPPK en complejo con HMP-P y AMP-PNP. **(F)** AncEnHMPPK en complejo con HMP-P y  $ATP_{\gamma s}$ .



**Figura suplementaria 7: Comparación de topología quinasas de vitaminas dependientes de ATP.**

Alineamiento de AncEnHMPPK con secuencias homólogas de la base de datos PDBAA, El alineamiento es coloreado de acuerdo con el grado de similitud de las secuencias y se presentan elementos de estructura secundaria calculados por DSSP. Esta figura fue construida mediante la plataforma ENDscript (Robert & Gouet, 2014).

### 7.3.5. Análisis del bolsillo de unión a sustratos de las estructuras de AncEnHMPPK



**Figura suplementaria 8: Conformaciones relevante de los grupos fosfato de los análogos de ATP en el bolsillo de unión a sustrato de AncEnHMPPK.**

Los ligandos se presentan como “stick balls” coloreados según átomos: Los átomos de carbono se coloraron en blanco, los átomos de nitrógeno en azul y los átomos de oxígeno en rojo. **(A)** Conformación Extendida. **(B)** Conformación Vertical. **(C)** Conformación atípica del fosfato  $\gamma$ .



**Tabla suplementaria 6: Distancias atómicas de residuos de la capa de coordinación de ligandos de los complejos HMP/Mg-ATP y HMP-P/MgATP.**

| Ligandos   | Grupos funcionales                        | Distancia (Å) |
|--|---|---------------|
| HMP  | Q44 – N <sub>1</sub> -N <sub>2</sub> -HMP | 3,0           |
|  | G178* – ATP-PO <sub>4α</sub>              | 3,2           |
|  | N139 – ATP-PO <sub>4β</sub>               | 3,0           |
|  | K176 – ATP-PO <sub>4β</sub>               | 2,8           |
|  | H179 – ATP-PO <sub>4β</sub>               | 2,9           |
|  | K111 – ATP-PO <sub>4γ</sub>               | 2,9           |
|  | T211* – ATP-PO <sub>4γ</sub>              | 3,3           |
|  | G212 – ATP-PO <sub>4γ</sub>               | 3,2           |
| HMP-P (ATP <sub>γS</sub> conformación extendida) | Q44 – N <sub>1</sub> -N <sub>2</sub> -HMP | 3,2           |
|  | K111* – HMP-P-met-PO <sub>4</sub>         | 2,9           |
|  | H209 – HMP-P-met-PO <sub>4</sub>          | 3,4           |
|  | G210* – HMP-P-met-PO <sub>4</sub>         | 2,9           |
|  | G178* – ATP-PO <sub>4α</sub>              | 2,9           |
|  | H179* – ATP-PO <sub>4α</sub>              | 3,5           |
|  | N139 – ATP-PO <sub>4β</sub>               | 2,8           |
|  | K176 – ATP-PO <sub>4β</sub>               | 2,8           |
|  | H179 – ATP-PO <sub>4β</sub>               | 2,4           |
|  | K111 – ATP-PO <sub>4γ</sub>               | 2,6           |
|  | T211 – ATP-PO <sub>4γ</sub>               | 3,3           |
|  | T211* – ATP-PO <sub>4γ</sub>              | 2,9           |
| G212* – ATP-PO <sub>4γ</sub>                     | 3,3                                       |               |
| HMP-P (ATP <sub>γS</sub> conformación atípica)   | Q44 – N <sub>1</sub> -N <sub>2</sub> -HMP | 3,2           |
|  | K111* – HMP-P-met-PO <sub>4</sub>         | 2,9           |
|  | H209 – HMP-P-met-PO <sub>4</sub>          | 3,4           |
|  | G210* – HMP-P-met-PO <sub>4</sub>         | 2,9           |
|  | N139 – ATP-PO <sub>4β</sub>               | 3,0           |
|  | K176 – ATP-PO <sub>4β</sub>               | 3,0           |
|  | T211 – ATP-PO <sub>4β</sub>               | 3,1           |
|  | G212* – ATP-PO <sub>4β</sub>              | 3,1           |
|  | K111 – ATP-PO <sub>4γ</sub>               | 2,4           |
|  | H179 – ATP-PO <sub>4γ</sub>               | 2,9           |

\* La interacción la realiza un grupo funcional de la cadena principal del aminoácido indicado

**Tabla suplementaria 7: Distancias atómicas de los grupos dadores y aceptores de fosforilo en AncEnHMPPK.**

| Ligandos   | Grupos funcionales                               | Distancia (Å) |
|--|--|---------------|
| HMP  | HMP-met-OH – ATP-PO <sub>4γ</sub>                | 5,2           |
|  | HMP-met-OH – Cys213-SH                           | 5,6           |
|  | ATP-PO <sub>4γ</sub> – Cys213-SH                 | 6,9           |
| HMP-P (ATP <sub>γS</sub> conformación extendida) | HMP-P-met-PO <sub>4</sub> – ATP-PO <sub>4γ</sub> | 4,3           |
|  | HMP-P-met-PO <sub>4</sub> – Cys213-SH            | 6,1           |
|  | ATP-PO <sub>4γ</sub> – Cys213-SH                 | 6,8           |
| HMP-P (ATP <sub>γS</sub> conformación atípica)   | HMP-P-met-PO <sub>4</sub> – ATP-PO <sub>4γ</sub> | 6,2           |
|  | HMP-P-met-PO <sub>4</sub> – Cys213-SH            | 5,9           |
|  | ATP-PO <sub>4γ</sub> – Cys213-SH                 | 9,4           |

## 7.4. MSA, inferencia filogenética y reconstrucción de secuencias

### ancestrales

#### 7.4.1. Colección de secuencias para construcción de MSA

**Tabla suplementaria 8: Listado de secuencias colectadas para construcción de MSA e inferencia filogenética.**

| Código de acceso | Especie                                     |
|------------------|---|
| WP_099636863.1   | Sporosarcina_sp_P7                          |
| WP_153721919.1   | Sporosarcina_sp_A4                          |
| WP_163971764.1   | Oceanobacillus_sp_YIM_98839                 |
| WP_146938140.1   | Cerasibacillus_quisquiliarum                |
| WP_164670912.1   | Virgibacillus_sp_MarseilleQ1616             |
| WP_156856029.1   | Oceanobacillus_sp_AG                        |
| WP_164214574.1   | Virgibacillus_sp_YIM_98842                  |
| WP_040954235.1   | Virgibacillus_sp_SK37                       |
| WP_094885367.1   | Virgibacillus_indicus                       |
| WP_087974488.1   | Oceanobacillus_rekensis                     |
| WP_188385290.1   | Ornithinibacillus_halotolerans              |
| WP_147668807.1   | Cerasibacillus_sp_CCCFT480                  |
| WP_115774023.1   | Oceanobacillus_arenosus                     |
| WP_114915275.1   | Oceanobacillus_sp_160                       |
| WP_040609261.1   | Salimicrobium_jeotgali                      |
| WP_179946089.1   | Vagococcus_penaei                           |
| WP_084193613.1   | Salimicrobium_flavidum                      |
| EHA32137.1       | Bacillus_subtilis_subsp_subtilis_str_SC8    |
| WP_001185462.1   | Staphylococcus                              |
| HXX36877.1       | Clostridiaceae_bacterium                    |
| NLL65064.1       | Clostridiaceae_bacterium                    |
| WP_085559446.1   | Carnobacterium_iners                        |
| TVP91026.1       | Alkalibacterium_sp                          |
| WP_091528356.1   | Alkalibacterium_subtropicum                 |
| WP_091488543.1   | Alkalibacterium_putridalgalicola            |
| WP_071847538.1   | Alkalibacterium_sp_20                       |
| WP_026866125.1   | Jeotgalicoccus                              |
| WP_106840115.1   | Brevibacillus_fortis                        |
| WP_147538907.1   | Bacillales_bacterium                        |
| KMK78188.1       | Alkalihalobacillus_pseudalcaliphilus        |
| WP_138860070.1   | Exiguobacterium_mexicanum                   |
| WP_054706457.1   | Bacillus_sp_JCM_19041                       |
| WP_143850395.1   | Bacillus_sp_P16_2019                        |
| WP_059104206.1   | Alkalihalobacillus_shacheensis              |
| WP_009711295.1   | Kroppenstedtia_eburnea                      |
| WP_186786148.1   | Paenibacillus_sp_N4                         |
| WP_159881658.1   | Paenibacillus_sp_SJY2                       |
| WP_165199230.1   | Bacillaceae_bacterium_SIJ1                  |
| WP_133581821.1   | Aureibacillus_halotolerans                  |
| WP_144704497.1   | Fictibacillus_phosphorivorans               |
| WP_073150499.1   | Seinonella_peptonophila                     |
| WP_040950426.1   | Gorillibacterium_massiliense                |
| WP_113030183.1   | Paenibacillus_contaminans                   |
| WP_044161131.1   | Virgibacillus_sp_MSP41                      |
| WP_078578304.1   | Salipaludibacillus_agaradhaerens            |
| AHD07730.1       | Paenibacillus_larvae_subsp_larvae_DSM_25430 |
| WP_191505201.1   | Candidatus_Harrysmithimonas_galli           |

|                |                                     |
|----------------|-------------------------------------|
| NLO87157.1     | Firmicutes_bacterium                |
| WP_078593432.1 | Bacillus_clarkii                    |
| WP_110521442.1 | Bacillus_lacisalsi                  |
| WP_095312538.1 | Bacillus_sp_75042                   |
| WP_101548846.1 | Bacillus                            |
| WP_117306299.1 | Bacillus_sp_V5932b                  |
| WP_187118989.1 | Bacillus_marasmi                    |
| WP_111616419.1 | Bacillus_sp_YR335                   |
| WP_071393203.1 | Bacillus_tuaregi                    |
| WP_026568148.1 | Bacillus_sp_UNC41MFS5               |
| WP_044393662.1 | Mesobacillus_subterraneus           |
| WP_007085202.1 | Neobacillus_bataviensis             |
| WP_043930142.1 | Bacillus_sp_EB01                    |
| WP_077713952.1 | Desulfotomaculum_ferrireducens      |
| WP_010278331.1 | Paenibacillus_senegalensis          |
| WP_020059715.1 | Bacillus_sp_123MFChir2              |
| WP_155477052.1 | Heliobacteriaceae                   |
| WP_131923032.1 | Hazenella_coriacea                  |
| WP_033647121.1 | Bacillus_anthraxis                  |
| GEN34504.1     | Aneurinibacillus_danicus            |
| EPR29576.1     | Geobacillus_sp_WSUCF1               |
| WP_038824112.1 | Yersinia_pseudotuberculosis         |
| WP_049597164.1 | Yersinia_nurmii                     |
| WP_061796536.1 | Serratia_ficaria                    |
| WP_034339812.1 | Deinococcus_misasensis              |
| WP_016388997.1 | Lactocaseibacillus_paracasei        |
| WP_046325780.1 | Lactobacillus_melliventris          |
| WP_038245820.1 | Virgibacillus_massiliensis          |
| WP_028784342.1 | Thalassobacillus_devorans           |
| WP_068724463.1 | Tepidibacillus_decatuensis          |
| WP_010296164.1 | Clostridium_senegalense             |
| WP_013840214.1 | Desulfotomaculum_ruminis            |
| CBX27792.1     | uncultured_Desulfobacterium_sp      |
| WP_013484000.1 | Ethanoligenens_harbinense           |
| WP_110173987.1 | Luteitalea_pratensis                |
| WP_037993393.1 | Thermoactinomyces                   |
| KUK11916.1     | Moorella_sp_60_41                   |
| KPK22132.1     | Dehalococcoidia_bacterium_SM23_28_1 |
| WP_038039052.1 | Thermorudis_peleae                  |
| WP_061087086.1 | Bacillus_coagulans                  |
| WP_002222176.1 | Neisseria_meningitidis              |
| WP_052036410.1 | Tumebacillus_flagellatus            |
| WP_044642824.1 | Risungbinella_massiliensis          |
| WP_013311666.1 | Paenibacillus_polymyxa              |
| WP_006674733.1 | Paenibacillus_dendritiformis        |
| WP_035287602.1 | Brevibacillus_massiliensis          |
| WP_036650443.1 | Paenibacillus_pini                  |
| WP_018755059.1 | Paenibacillus_terrigena             |
| WP_035315563.1 | Brochothrix_campestris              |
| WP_023176506.1 | Carnobacterium_inhibens             |
| WP_009499104.1 | Sporosarcina_newyorkensis           |
| KRG14359.1     | Virgibacillus_soli                  |
| WP_034770808.1 | Exiguobacterium_marinum             |
| WP_034798865.1 | Exiguobacterium_aurantiacum         |
| WP_012674032.1 | Sulfurihydrogenibium_azorense       |
| WP_012991289.1 | Thermocrinis_albus                  |
| WP_157560553.1 | Hydrogenivirga_sp_1285R11           |
| WP_010889431.1 | Deinococcus_radiodurans             |
| WP_041227362.1 | Deinococcus_deserti                 |

|                |  |
|----------------|--|
| WP_011531041.1 | Deinococcus_geothermalis                   |
| EYB67778.1     | Deinococcus_phoenicis                      |
| WP_015234119.1 | Deinococcus_peraridilitoris                |
| WP_018461804.1 | Thermus_oshimai                            |
| KOX90752.1     | Thermus_aquaticus                          |
| WP_014515741.1 | Thermus_sp_CCB_US3_UF1                     |
| WP_011228119.1 | Thermus_thermophilus                       |
| WP_060384599.1 | Thermus_parvatiensis                       |
| WP_018110360.1 | Thermus_igniterrae                         |
| WP_015717082.1 | Thermus_scotoductus                        |
| CFX40698.1     | Candidatus_Filomicrobium_marinum           |
| WP_023787573.1 | Hyphomicrobium_nitratorans                 |
| KWT66909.1     | Hyphomicrobium_sulfonivorans               |
| WP_013652224.1 | Polymorphum_gilvum                         |
| WP_020179112.1 | Methylopila_sp_M107                        |
| WP_056292659.1 | Afipia_sp_Root123D2                        |
| WP_056407566.1 | Methylobacterium_sp_Leaf466                |
| WP_045366486.1 | Methyloceanibacter_caenitepidi             |
| WP_017589389.1 | Nocardiosis_ganjiahuensis                  |
| WP_057791395.1 | Roseovarius_atlanticus                     |
| WP_037083007.1 | Rhizobium_vignae                           |
| WP_037064810.1 | Rhizobium_leguminosarum                    |
| WP_060834453.1 | Rhodovulum_sulfidophilum                   |
| WP_053908195.1 | Moellerella_wisconsensis                   |
| WP_035606664.1 | Edwardsiella_tarda                         |
| WP_031520779.1 | Siccibacter_colletis                       |
| WP_001188957.1 | Salmonella                                 |
| WP_000822282.1 | Escherichia_coli                           |
| WP_025760290.1 | Enterobacter_cloacae                       |
| WP_064190446.1 | Klebsiella_pneumoniae                      |
| WP_072124372.1 | Klebsiella_variicola                       |
| WP_003436492.1 | Clostridiales                              |
| TKB79584.1     | Nitrospira_sp                              |
| MRR09730.1     | bacterium                                  |
| OGP98091.1     | Deltaproteobacteria_bacterium_RBG_16_55_12 |
| HEY63192.1     | Caldilineae_bacterium                      |
| MBD3232169.1   | candidate_division_Zixibacteria_bacterium  |
| NQT34709.1     | bacterium                                  |
| MXZ12698.1     | Candidatus_Dadabacteria_bacterium          |
| WP_110956164.1 | Massilibacillus_massiliensis               |
| WP_057976662.1 | Caloramator_mitchellensis                  |
| WP_174579622.1 | Clostridia_bacterium                       |
| WP_132016623.1 | Hydrogenispora_ethanolica                  |
| WP_152805448.1 | Alkalibaculum_sp_M08DMB                    |
| WP_158741277.1 | Defluviitalea_raffinosedens                |
| WP_073007934.1 | Clostridium_amyolyticum                    |
| HBF66509.1     | Clostridium_sp                             |
| WP_114217857.1 | Firmicutes_bacterium                       |
| HAU85853.1     | Lachnospiraceae_bacterium                  |
| WP_034624936.1 | Alkalihalobacillus_okhensis                |
| WP_143848595.1 | Bacillus_sp_P16_2019                       |
| PYS02843.1     | Acidobacteria_bacterium                    |
| WP_134259386.1 | Alkalihalobacillus_lehensis                |
| WP_157407097.1 | Bacillus_anthraxis                         |

## 7.4.2. Selección de modelo evolutivo para inferencia filogenética

Tabla suplementaria 9: Valores de akaike para los 5 mejores modelos evolutivos evaluados en software Prottest.

| Modelo  | $\Delta AIC$ | AIC       | AICw | -lnL     |
|---------|--------------|-----------|------|----------|
| LG+I+G  | 0,00         | 102164,81 | 1    | 50158,91 |
| LG+G    | 133,50       | 102298,32 | 0    | 50228,55 |
| WAG+I+G | 255,44       | 102420,25 | 0    | 50286,63 |
| WAG+G   | 439,28       | 102604,09 | 0    | 50381,44 |
| VT+I+G  | 478,25       | 102643,06 | 0    | 50398,03 |
| CpREV+G | 626,66       | 102791,47 | 0    | 50417,23 |

## 7.4.3. Comandos utilizados para inferencia filogenética y reconstrucción ancestral

### 7.4.3.1. Comando utilizado en PhyML para inferencia filogenética

```
$ phymml-mpi -i alin_HMPPK_final_taxa.phy --datatype aa --bootstrap -2 --model LG -f m --pinv -e --alpha e --nclases 4 --search BEST --no_memory_check
```

### 7.4.3.2. Comando utilizado en Lazarus para la reconstrucción de secuencias ancestrales

```
$ python /home/nicolas/Lab/project-lazarus-master/lazarus.py --codeml --outdir /home/nicolas/Lab/reconstruccion-HMPPK/out --verbose --alignment /home/nicolas/Lab/reconstruccion-HMPPK/alin_HMPPK_final_taxa.fasta --tree /home/nicolas/Lab/reconstruccion-HMPPK/arbol_HMPPK2.nwk --model /home/nicolas/Lab/project-lazarus-master/paml/dat/lg.dat --branch_lengths fixed --asrv 4 --outgroup taxa107
```

### 7.4.3.3. Secuencias proteínas ancestrales

#### 7.4.3.3.1. Secuencia AncEnHMPPK

```
MKRINALTIAGTDPSSGGAGIQADLKTFSALGAYGCSVITALVAQNTRGVQSVYQIDPDFVAAQ  
LDSVFSDVRIDSTKIGMLANA AIVEAVAERLRRYRVANVVLDTVMLAKSGDPLLAPDAVEALR  
QRLLPQVSLITPNLPEAAALLDAPPARTEREMREQGRALLAMGCGAVLMKGGHLDDAESP  
WLFTREREQRFTAPRVATKNHTGTGCTLSAALAALRPRHADWAATVAEAKDYLSGALAQAD  
TLEVGHGIGPVHFFHAWW
```

#### 7.4.3.3.2. Secuencia AnchMPPK

MTMPRALTIAGSDSGGGAGIQADLKTQELDVYGMSAITAITAQNLTGVHGVYPPQVEAIEQQ  
LDAVMSDIGVDAVKTGMLFSAEIIELVAEKIKQYNIKNIVVDPVMIAKGGSSLLQEEAVQALKEQ  
LLPLATVVTPNLPEAAVLLGRKPIKTIEDMKEAARRLHDLGPKYVVIKGGHLEGDPAVDLLYDG  
EQFHQFESPRIDTKHHTGTGCTFSAAITAEALAKGKPIQEAVQTAKEFITAAIRHSLNIGQGVGP  
TNHAAYRNRKNE

#### 7.4.3.3.3. Secuencia AncC

MTMPRALTIAGSDSGGGAGIQADLKTQELDVYGMSAITAIVAQNPLGAHGVYPPQVEAIEAQ  
LDTVMEDIGVDALKTGMLFSAEIIELVAEKIKQYNIKNIVVDPVMIAKGGSELLQDEAVEALKEQ  
LIPLATVVTPNLPEASKLLGRKPIKTIEDMKEAARRLHELGAKEYVLIKGGHLEGHPAVDLLYDG  
ETFHLFESPRIDTTHHTHGAGCTYSAAITAEALAKGKPVQEAVQTAKEFITAAIRHSLNFGEGVGP  
TNHAAYRKYGEKR

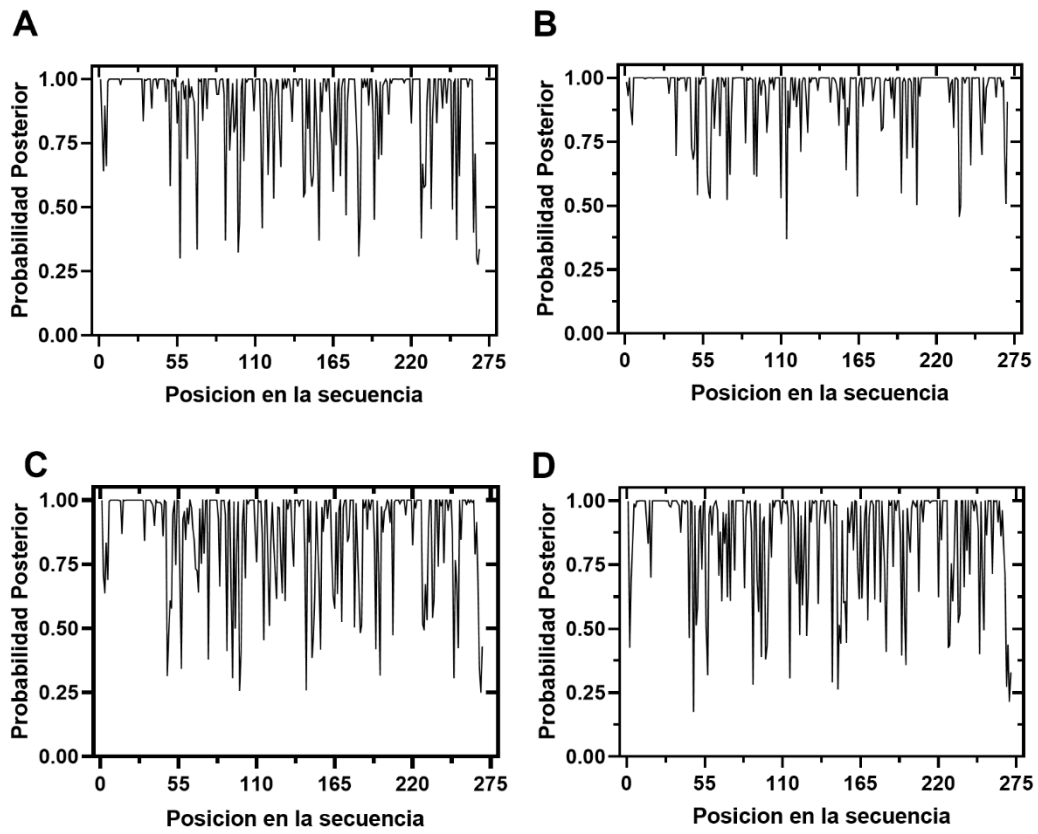
#### 7.4.3.3.4. Secuencia AncPLK/HMPPK

MTMYKALTIAGSDSSGGAGIQADLKTQELGVYGMSAITTIVSMDPKNNWAHNVFPQAVDTV  
EAQLDTILSIGVDAMKTGMLGSVEIIELAAKTIDEYKLNKVVVDPVMVCKGEDEVLPETAESL  
REVLVPRATVVTPNLFEASQLSGQGPIKTVEDMKEAARKIHELGAKEYVLIKGGSKLEHEKAVD  
LLYDGGQTFELLES DRIDTTYTHGAGCTYSAAITAEALAKGKSVKEAVQTAKEFITEAIRHSFKFN  
EFVGPNTNHAAYRKYGKG

#### 7.4.3.3.5. Secuencia AncUnk

MMIPRALTIAGSASGGGAGIQADLKTQELDVYGMSAVTAIVARNPLTGQGVYPPQPIEAIEAQ  
LDTVMEDIGVDALKTGMLFSKEIIELVAEIKQYDIKNIVVDPVMIGKMGSQLLKDEAIEALKEQL  
IPLATVITPNLPEASKLLGRREIKTVEDMKEAARRLYELGAKEYVLIKGGRLGHPAVDLLYDGE  
QFHLFESPRIDTVHTSGAGCTYSAAITAGLAKGKPVQEAVQEAKKFITAAIRHSLTFGEGVGP  
TYHAAYRKYGEKRQT

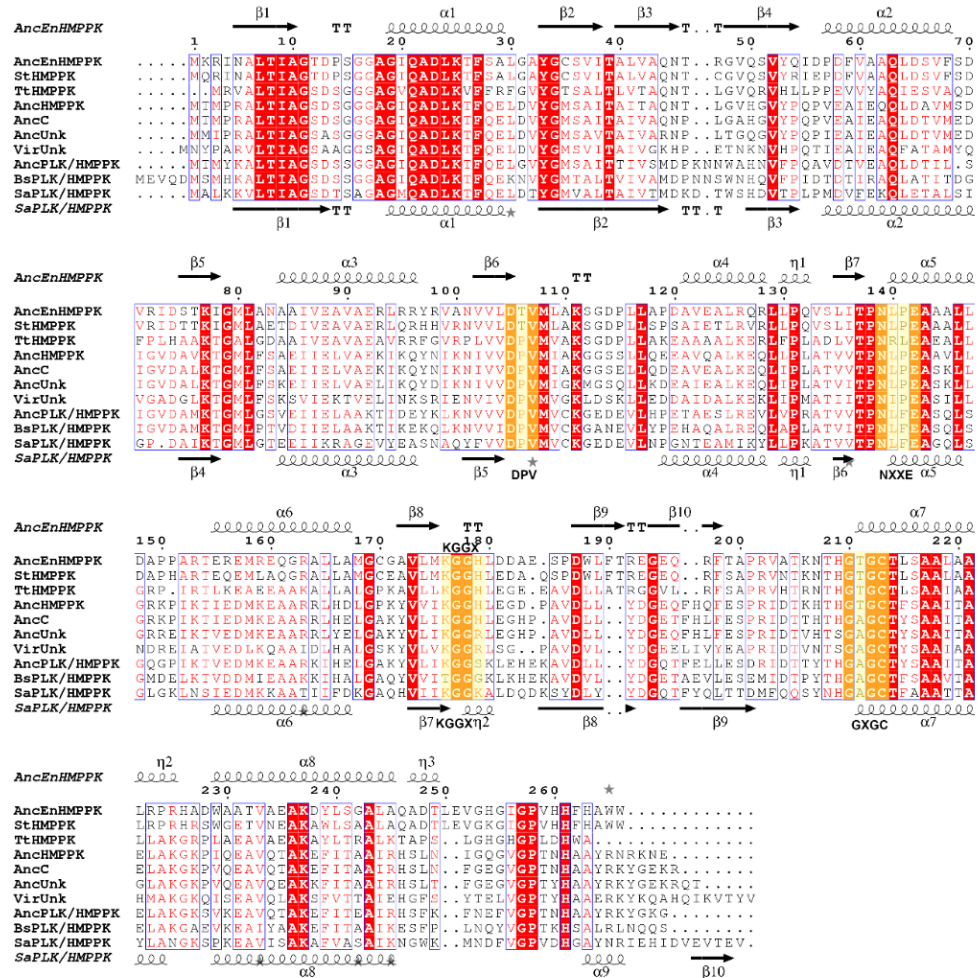
#### 7.4.3.4. Robustez reconstrucción secuencias ancestrales



**Figura suplementaria 9: Probabilidad posterior respecto posicion de secuencia para las proteínas ancestrales inferidas.**

**(A)** AncHMPPK. **(B)** AncPLK/HMPPK. **(C)** AncC. **(D)** AncUnk.

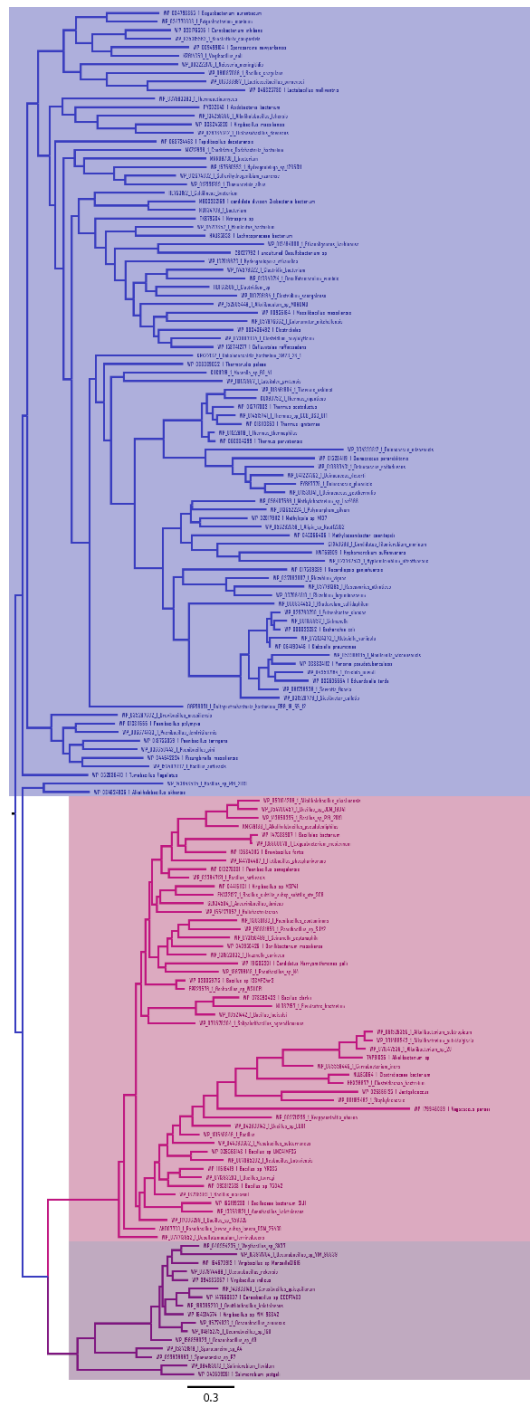
### 7.4.4. Filogenia de las quinasas de vitaminas dependientes de ATP de bacterias



**Figura suplementaria 10: Alineamiento representativo extendido quinasas de vitaminas dependientes de ATP de bacterias.**

El alineamiento se realizó con homólogos de la familia quinasas de vitaminas dependientes de ATP de bacterias, encontrándose coloreado de acuerdo con el grado de similitud de las secuencias. En la línea superior se muestran los elementos de estructura secundaria de AncEnHMPPK, mientras que en la línea inferior se muestran los elementos de estructura secundaria de la HMPPK de *Salmonella typhimurium* (PDB ID: 1JXI), calculadas con DSSP. En el alineamiento se muestran las secuencias de quinasas de vitaminas dependientes de ATP de bacteria, representativas de los grupos: 1.- ThiD-HMPPK: La HMPPK ancestral del orden Enterobacteriales (AncEnHMPPK), la HMPPK de *Salmonella typhimurium* (StHMPPK) y la HMPPK de *Thermus thermophilus* (TtHMPPK). 2.- PdxK-PLK/HMPPK: La PLK/HMPPK de *Bacillus subtilis* (BsPLK/HMPPK) y la PLK/HMPPK de *Staphylococcus aureus* (SaPLK/HMPPK). 3.- Un grupo no reportado anteriormente sin caracterización (Unk) con un representante de *Virgibacillus sp.* SK37 (VirUnk). Este alineamiento fue construido con la plataforma web ENDscript (Robert & Gouet, 2014). Se incorporaron secuencias ancestrales reconstruidas en esta tesis.





**Figura suplementaria 11: Filogenia extendida de las quinasas de vitaminas dependientes de ATP de bacterias.**

Los colores representan diferentes grupos filogenéticos, correspondiendo el negro a las ThiD-HMPPK, verde a las PdxK-PLK/HMPPK, fucsia a un grupo sin caracterización reportada. Se utilizó el modelo evolutivo LG+I+G de acuerdo con la selección de modelos por el factor de información de akaike (AIC) (Tabla suplementaria 9).