DEPARTAMENTO DE BIOQUÍMICA Y BIOLOGÍA MOLECULAR

MÁQUINAS MOLECULARES QUE SINTETIZAN ANHÍDRIDOS FOSFÓRICOS

CLARA MARCO MARÍN

UNIVERSITAT DE VALENCIA
Servei de Publicacions
2007

Aquesta Tesi Doctoral va ser presentada a Valencia el dia 19 de Febrer de 2007 davant un tribunal format per:

- D. Ignacio Fita Rodríguez
- D. Pedro Alzari
- D. Juan J. Hermoso Domínguez
- D ª. Asunción Contreras de Vera
- D. Joaquín Moreno Mariño

Va ser dirigida per:
D. Vicente Rubio Zamora

Depòsit legal:
I.S.B.N.:978-84-370-6802-2
Edita: Universitat de València
Servei de Publicacions
C/ Artes Gráficas, 13 bajo
46010 València
Spain
Telèfon: 963864115

UNIVERSITAT DE VALENCIA
Departamento de Bioquímica y Biología Molecular

INSTITUTO DE BIOMEDICINA DE VALENCIA
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

Máquinas moleculares que sintetizan anhidridos fosfóricos

Memoria presentada por
CLARA MARCO MARÍN
para optar al grado de
Doctora en Biología

Director
Prof. Vicente Rubio Zamora

Valencia, Noviembre de 2006

La autora ha disfrutado de becas FPI de la Generalitat Valenciana e I3P CSIC-Bruker España S.A. y de un contrato de Titulado Superior con cargo al proyecto BFU2004-05159. Es miembro del CIBER de Enfermedades Raras (CIBERER) del Instituto de Salud Carlos III.

El trabajo se ha financiado principalmente con las ayudas BMC2001-2182 y BFU2004-05159 del Plan Nacional de Investigación Científica y Técnica (MCT y MEC). También se ha desarrollado bajo los auspicios del Instituto de Salud Carlos III (MSC; Redes C03/08, G03/54 y FISS PI052838).

La autora agradece a los sincrotrones de Grenoble (ESRF) y Berlín (BESSY) por su apoyo en la recogida de datos y a la Unión Europea y al Ministerio de Ciencia y Tecnologia por la asistencia financiera para el uso de los mismos.

AGRADECIMIENTOS

Quiero dar la gracias a todas las personas que han contribuido a que este trabajo de tesis sea un objetivo alcanzado.

A mi director de tesis, el Profesor Vicente Rubio, al que le debo no sólo la oportunidad de trabajar en su laboratorio sino también lo mucho que he aprendido a su lado. Gracias por hacerme partícipe de tus ideas y por estar dispuesto a discutir las mias.

Gracias a todos mis compañeros del laboratorio, que han sido muchos: Andrea, Amparo, Belén, Benito, Chelo, Elena, Enea, Fernando, Jorge, Jose Luis, Juanma, Leonor, Lourdes, Matxalen, Mariano, Mari Luz, Michelle, Patricia, Rebeca, Sandra y Santi. Con cada uno de vosotros he compartido momentos diferentes, pero de todos he aprendido mucho: a trabajar en el laboratorio, a afrontar situaciones, a solucionar problemas, e incluso a conocerme mejor a mi misma. Tengo la suerte de haber hecho grandes amigos, y espero que nuestra amistad dure mucho tiempo.

A todo el personal del Instituto de Biomedicina de Valencia. A los investigadores, becarios y personal del laboratorio, muchas gracias a todos por la convivencia en estos años, por vuestra ayuda siempre que la he necesitado. Al personal de mantenimiento, administración, informática, seguridad, limpieza …, por solucionar los pequeños desastres y por facilitarnos el dia a dia.

Al Profesor Ignacio Fita, por su inestimable ayuda y apoyo, por hacer que la cristalografía sea más fácil.

Al Dr. Alberto Marina, por su disponibilidad en todo momento.

Al Dr. Javier Cervera y a su grupo, por permitirme colaborar con vosotros.

A las Dras. Nuria Verdaguer y Cristina Vega por su dedicación y paciencia en la coordinación del BAG.

A los miembros del laboratorio CRI2 del Instituto de Biologia Molecular de Barcelona por echarme un cable siempre que lo he necesitado y por su buena acogida durante mis visitas.

A todos los investigadores que he conocido durante estos años, por participar en mi formación científica y en las inquietudes de los que estamos aprendiendo.

A mis padres, mi hermana y mi marido, muchas gracias por vuestro interés, por vuestro apoyo incondicional, por compartir conmigo el entusiasmo y la euforia y por no dejarme flojear en momentos de desesperación. Gracias por motivame a estudiar y a trabajar en lo que realmente me apasiona y por apoyarme en las decisiones más importantes de mi vida. Muchas gracias.

ÍNDICE

INTRODUCCIÓN GENERAL

INTRODUCCIÓN GENERAL

1. OBJETO GENERAL DE ESTA TESIS

Este trabajo de Tesis se ocupa de enzimas que sintetizan anhidridos del ácido fosfórico. Es imposible soslayar la crucial importancia biológica de estos compuestos, de entre los que la Figura 1 selecciona unos pocos.

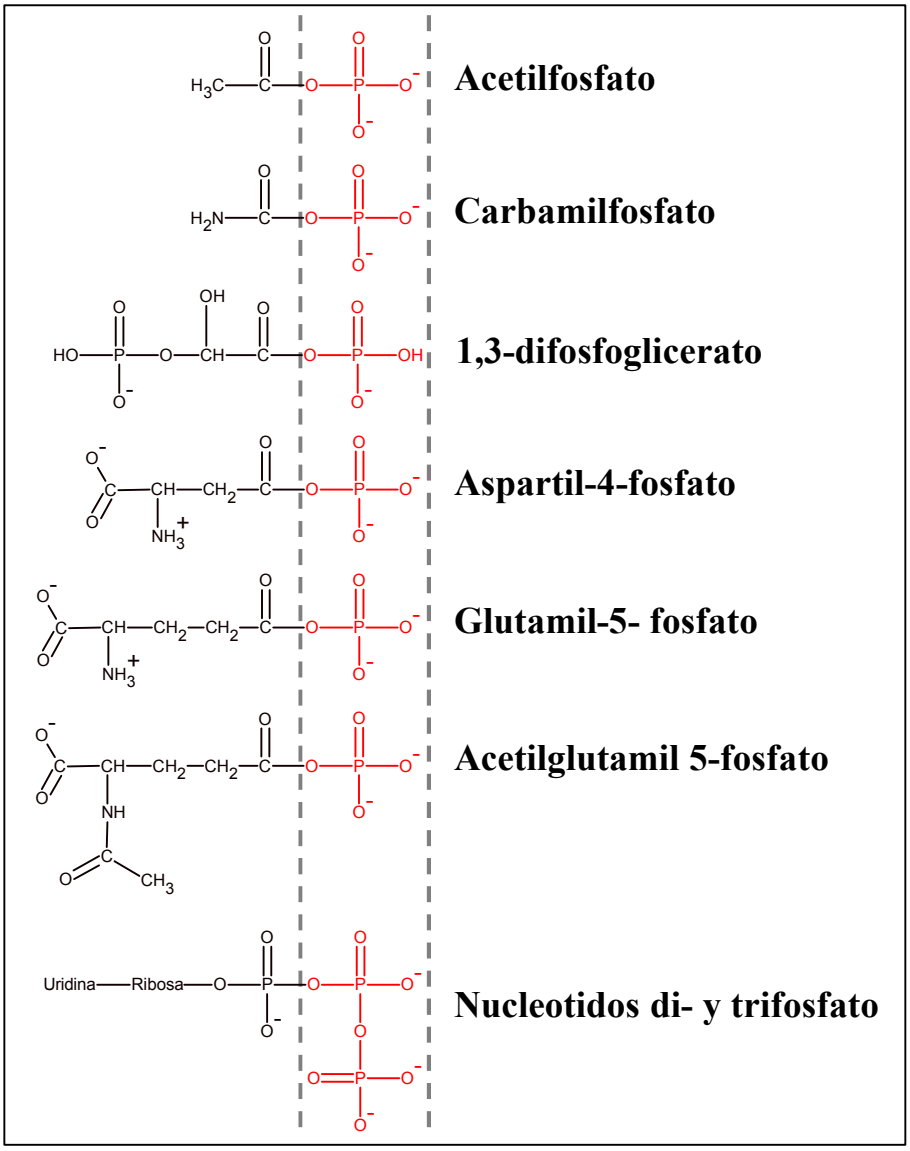


Figura 1. Anhidridos del ácido fosfórico. El fosfato unido a cada una de las moléculas por un enlace anhidrido, se indica en rojo.

Acetilfosfato y carbamilfosfato se encuentran entre los más simples y son posiblemente muy antiguos, objeto de síntesis prebiótica (Saygin, 1984) y precursores de otros compuestos en el largo periodo entre la sopa primigénia y el desarrollo de la vida, poseyendo el acetilfosfato un papel clave en el flujo global de carbono en la

biosfera (Buss et al., 2001). El 1,3-difosfoglicerato, un compuesto intermedio de la glicólisis (Michal, 1999), tiene un lugar clásico entre los anhidridos mixtos carboxílico-fosfóricos. No resaltaré por conocido el papel de los nucleótidos di y trifosfato, de entre los que destaca particularmente el ATP por su papel de moneda energética universal en los seres vivos. También el NAD, el NADP y el coenzima A cumplen funciones indispensables en las oxi-reducciones y en la transferencia biológica de acetato (Michal, 1999) y contienen anhidridos fosfóricos.

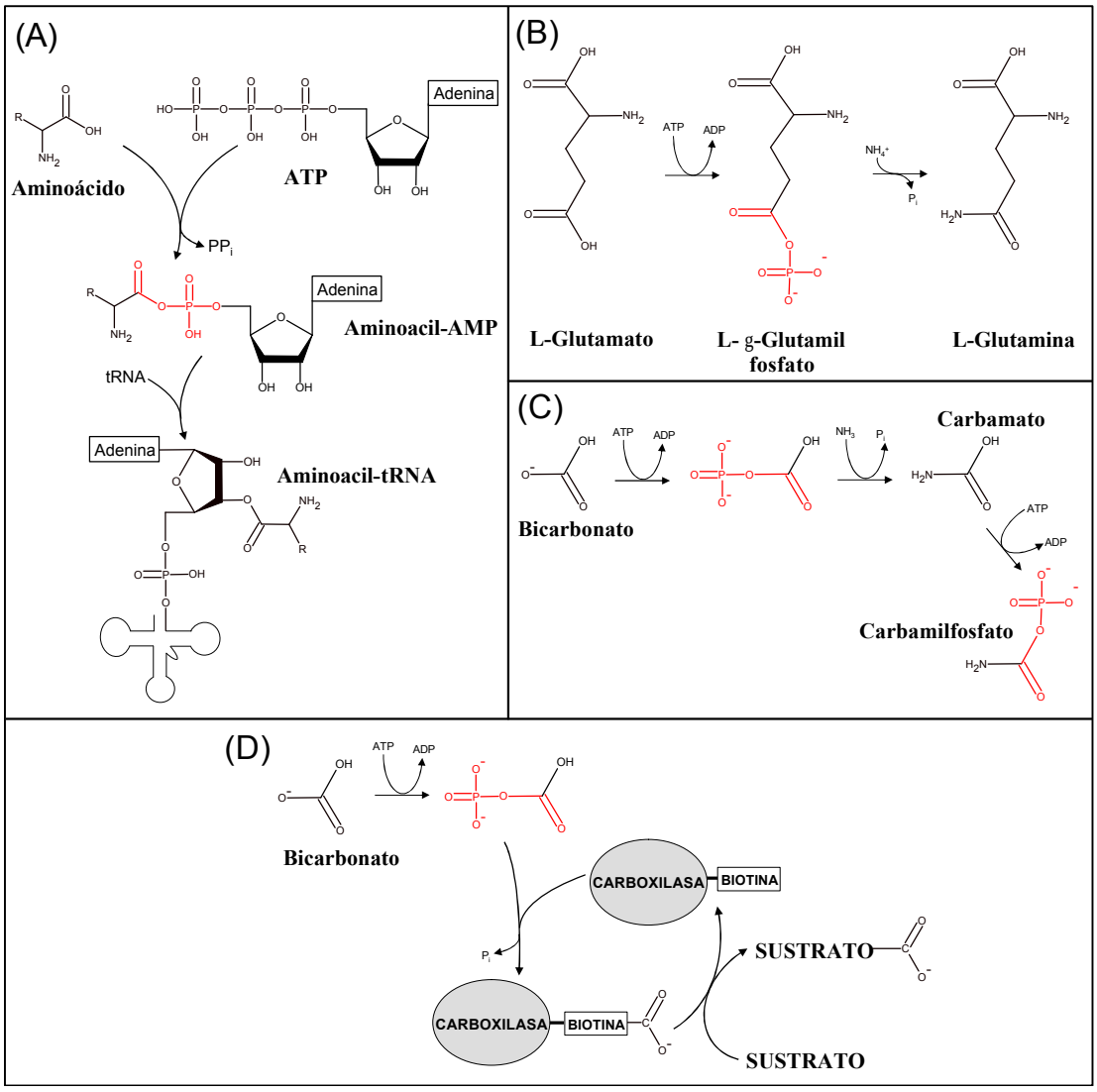


Figura 2. Reacciones catalizadas por las aminoacil-tRNA sintetasas (A), la glutamina sintetasa (B), la carbamilfosfato sintetasa (C) y las carboxilasas dependientes de biotina (D) y que implican la formación de anhidridos fosfóricos como intermedios de la reacción. En rojo se representan los grupos que participan en la formación del anhidrido fosfórico.

Las reacciones de síntesis biológica incluyen generalmente pasos en que se sintetizan anhidridos del ácido fosfórico, muchas veces como compuestos intermedios crípticos que son parte de los mecanismos reaccionales de sintetasas (ver ejemplos en la Figura 2), y otras veces como compuestos intermedios libres, producidos por un enzima

y utilizados por el siguiente (Figura 1), aunque en ocasiones, como en el caso del γ-glutamil fosfato, el papel del anhidrido es intermedio entre estas dos categorías, ya que aunque puede ser sintetizado por un enzima y utilizado por otro, es aún en ese caso posiblemente objeto de canalización entre el enzima productor y el aceptor (ver más adelante).

En todo caso, en esta tesis me ocupo de cuatro enzimas que sintetizan estos anhidridos en forma libre, en un caso (N-acetil-L-glutamato quinasa) tangencialmente, en otro (aspartoquinasa III) realizando un estudio de mutagénesis-modelizado estructural y en los otros dos (UMP quinasa arqueal y glutamato 5-quinasa bacteriana) determinando la estructura tridimensional del enzima mediante cristalografía/difracción de rayos X. Los productos de tres de estos enzimas (acetilglutamato, aspartato y glutamato quinasas) son anhidridos carboxílico-fosfóricos, y por tanto estos tres enzimas pertenecen al grupo 2.7.2. de la *Enzyme Commission*, constituido por fosfotransferasas que usan un grupo carboxilato como aceptor (Figura 3).

E.C. 2.7.2. Fosfotransferasas con un grupo carboxilo como aceptor	
2.7.2.1	Acetato quinasa
2.7.2.2	Carbamato quinasa
2.7.2.3	Fosfoglicerato quinasa
2.7.2.4	*Aspartato quinasa*
2.7.2.6	Formato quinasa
2.7.2.7	Butirato quinasa
2.7.2.8	*Acetilglutamato quinasa*
2.7.2.11	*Glutamato 5-quinasa*
2.7.2.15	Propionato quinasa

Figura 3. Enzimas del grupo 2.7.2 de la Enzyme Commission (http://www.chem.qmul.ac.uk/iubmb/enzyme/EC2/7/2/). El listado ha sido modificado para excluir las entradas transferidas a otros grupos, la glutamato 1-quinasa, la acetato quinasa difosfato y la fosfoglicerato quinasa (GTP). En cursiva se señalan los enzimas de los que me he ocupado en este trabajo.

El cuarto enzima (UMP quinasa) fabrica un anhidrido fosfórico-fosfórico (el del UDP) y por tanto pertenece al grupo de la E.C. 2.7.4., el de las fosfotransferasas que usan un fosfato como aceptor. Aunque perteneciendo este último enzima a un grupo distinto al de los otros tres, los cuatro enzimas pertenecen a una misma familia evolutiva

de seis miembros reconocidos (Figuras 4-6) que los resultados de nuestro grupo y los míos propios han demostrado comparten también un mismo patrón estructural. Esta familia es denominada por la base de datos Pfam del instituto Sanger (PF00696, http:/www.sanger.ac.uk/Software/Pfam) "Aminoácido Quinasa", denominación posiblemente no idónea, pues uno de sus enzimas no es una quinasa sino una transferasa (N-acetil-L-glutamato sintasa; transfiere el grupo acetilo del acetilCoA al grupo amino del glutamato), y dos de ellos no sintetizan aminoácidos (carbamato quinasa y UMP quinasa), si bien la carbamato quinasa es un enzima que cataliza el último paso de una ruta catabólica del aminoácido arginina.

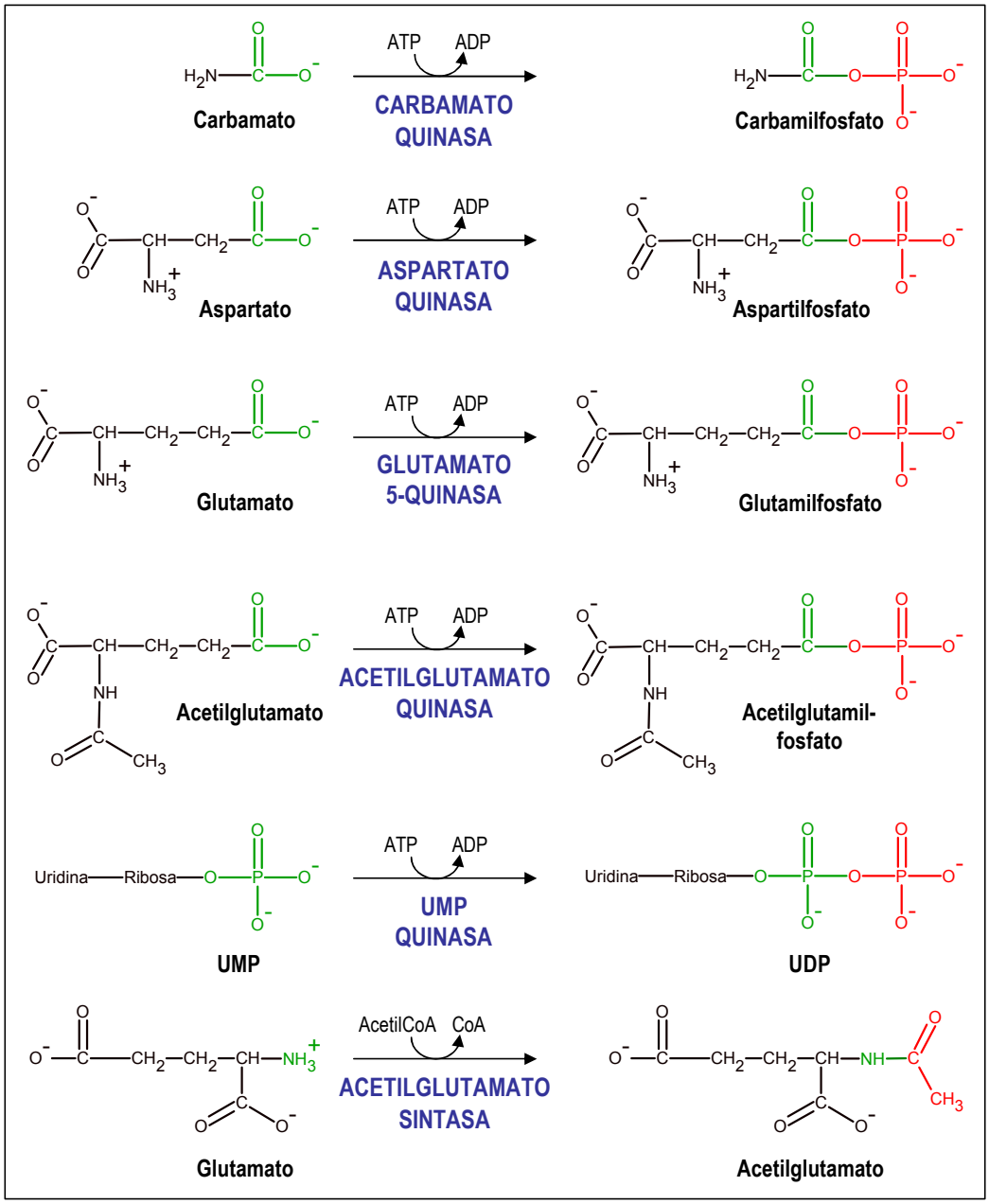


Figura 4. Enzimas de la familia aminoácido quinasa (en azul) (PF00696 http://www.sanger.ac.uk/Software/Pfam) y sus reacciones. En verde se representa el grupo reaccionante del sustrato, unido al grupo fosfato o acetato (en rojo) en el producto.

Apropiada o no, esta denominación de aminoácido quinasa será utilizada aquí, tal como hace el Instituto Sanger, para definir el conjunto de estos seis enzimas, así como para referirnos a un dominio de dichos enzimas (abreviado AAK), de unos 250 aminoácidos, que los caracteriza estructuralmente y que es responsable de la unión de los sustratos y de la catálisis por al menos aquellos de estos enzimas que tienen la característica de ser quinasas. Hay que resaltar que de estos seis enzimas, sólo dos, la glutamato 5-quinasa y la acetilglutamato sintasa, se encuentran en animales, lo que confiere a esta familia, compuesta por elementos que desempeñan funciones clave en el metabolismo de aminoácidos y de pirimidinas, importantes potencialidades de intervención práctica, bien para el desarrollo de biocidas, obvio en el caso de la UMP quinasa bacteriana (enteramente diferente, como se verá aquí, del enzima correspondiente animal, la UMP/CMP quinasa), como para el manejo biotecnológico de propiedades tales como la acumulación de lisina en semillas (limitada por la inhibición feed-back de la aspartoquinasa) o de acumulación en las plantas del osmoprotector (y por tanto protector frente al estres hídrico y a la salinidad) prolina, acumulación en la que la glutamato 5-quinasa desempeña un papel muy importante.

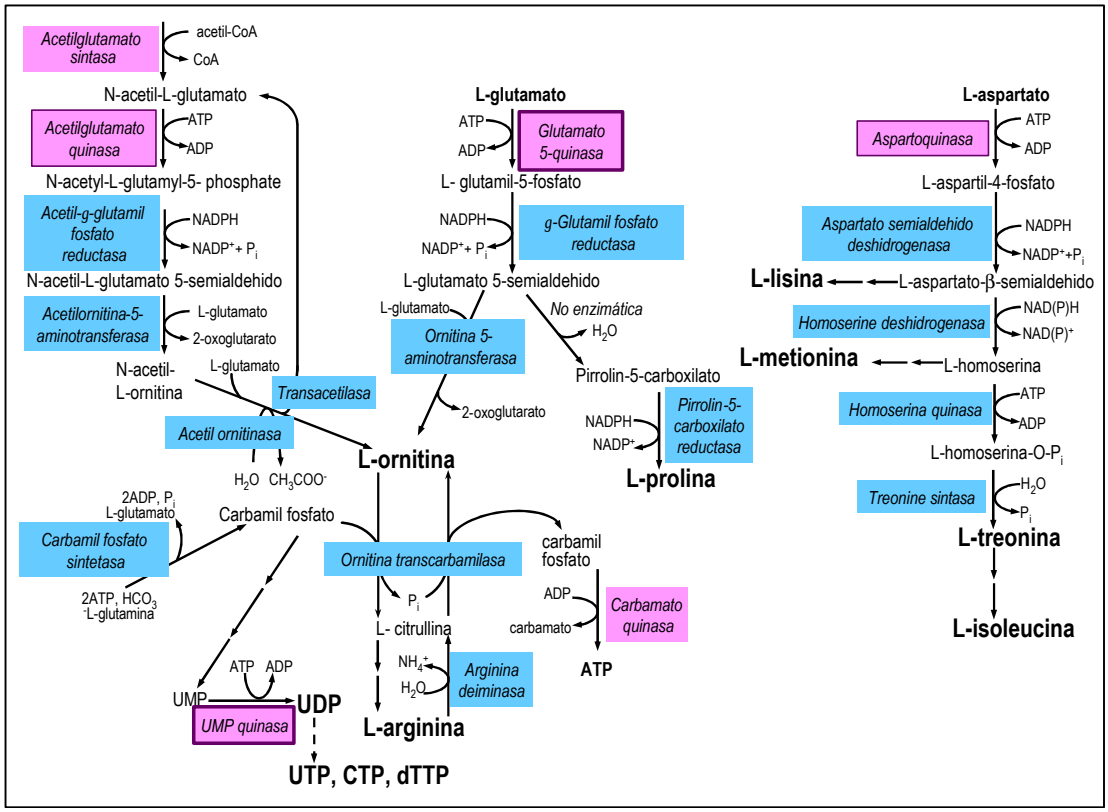


Figura 5. Rutas metabólicas en las que participan los enzimas de la familia aminoácido quinasa (sombreados en violeta). Los enzimas de los que me he ocupado en este trabajo se indican con un recuadro oscuro (la línea gruesa indica determinación de la estructura y la fina estudios de mutagénesis dirigida o de modelizado).

2. LA REGULACIÓN ES UN ASPECTO CLAVE DE LOS ENZIMAS DE LA FAMILIA AMINOÁCIDO QUINASA.

Si el dominio aminoácido quinasa es responsable de la unión de los sustratos y de la catálisis, también en algunos de los enzimas de esta familia es responsable de la regulación de la ruta en que participan. En realidad, de los seis enzimas solamente dos no tienen funciones reguladoras, la carbamato quinasa y la forma de acetilglutamato quinasa presente en los organismos que no reciclan el grupo acetilo del acetilglutamato. Es natural que los demás enzimas de la familia AAK tengan un importante papel regulador, por catalizar pasos iniciales de rutas metabólicas de síntesis (Figura 5) en que no es deseable acumular el producto final más allá de las necesidades metabólicas.

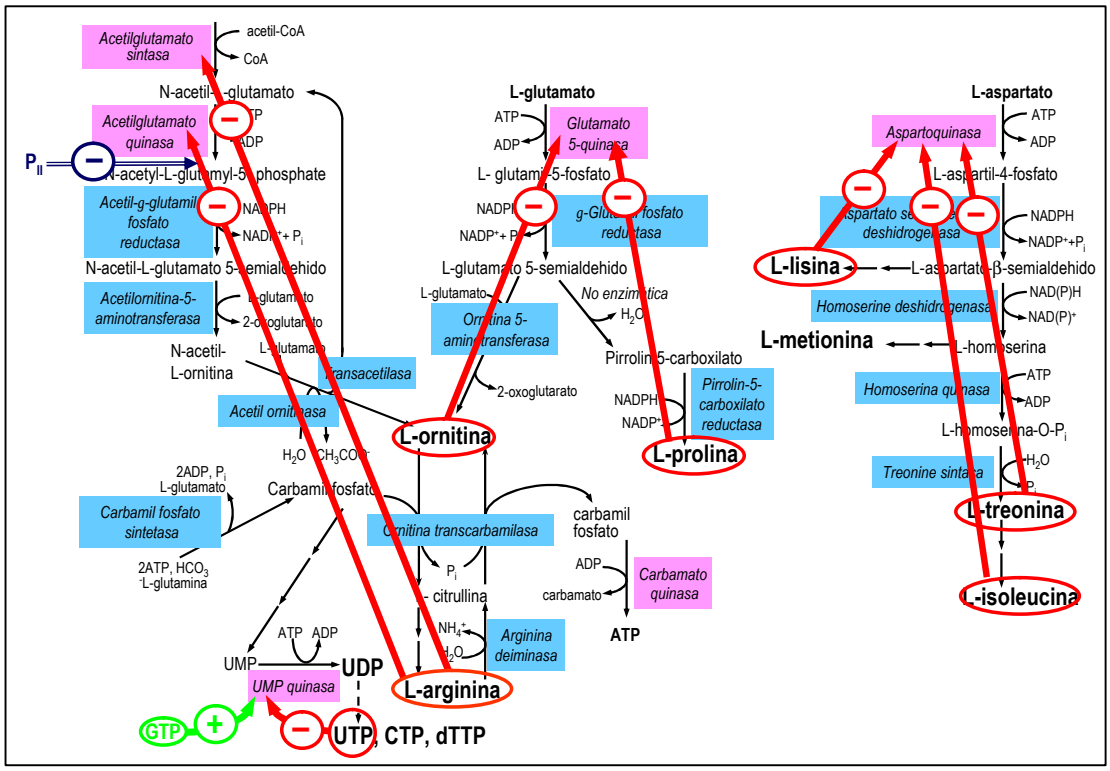


Figura 6. Regulación de los enzimas de la familia aminoácido quinasa. Las flechas rojas indican inhibición (marcada con signo negativo), y la verde activación (marcada con signo positivo). Los efectores que ejercen el control, están indicados con un círculo. El efecto de la proteina señalizadora P_{II} sobre la inhibición de la acetilglutamato quinasa por arginina, se indica con una doble flecha azul y un signo negativo.

Así, la acetilglutamato sintasa y la acetilglutamato quinasa presente en los organismos que reciclan el grupo acetilo del acetilglutamato, son inhibidas por el producto final de su ruta, el aminoácido arginina (Cunin et al., 1986; Leisinger & Haas, 1975; Haas & Leisinger, 1975; Fernandez-Murga et al., 2004) (Figura 6). Esta retroinhibición es tan limitante que, en los organismos fotosintéticos, que deben

almacenar nitrógeno en situaciones de abundancia de amonio, se ha desarrollado un mecanismo de supresión de la retroinhibición por arginina que se basa en la utilización de la proteína señalizadora P_{II}, que tiene en estos organismos como diana a la acetilglutamato quinasa (Burillo et al., 2004; Heinrich et al., 2004; Sugiyama et al., 2004; Arcondeguy et al., 2001) (Figura 6). En la ruta de síntesis de prolina, y, en animales (y quizá también en plantas) de ornitina, la glutamato 5-quinasa cataliza el primer paso (Leisinger, 1996; Aral & Kamoun, 1997; Baumgartner et al., 2000), siendo retroinhibida en unos casos por prolina y en otros por ornitina (Leisinger, 1996; Smith et al., 1985; Fujita et al., 2003; Hu et al., 1999) (Figura 6). En el caso de otro enzima de la familia, la aspartoquinasa, la Figura 5 ilustra la multiplicidad de productos derivados de la ruta en que este enzima cataliza el primer paso. Como cabría esperar de este hecho, este enzima es sujeto de una rica regulación, mediada en unos casos por treonina, en otros por lisina y en ocasiones por isoleucina o por S-adenosilmetionina, con complejos comportamientos reguladores que incluyen en ocasiones sinergias entre dos inhibidores, dependiendo del tipo y de la función biológica de la aspartoquinasa (Trufa-Bachi, 1973). Finalmente, el enzima de esta familia que forma un enlace fosfato-fosfato, la UMPK, es la puerta que permite en los microorganismos y posiblemente también en ciertos compartimientos (núcleo y mitocondria) de las plantas, la síntesis de UDP, compuesto clave para la síntesis de polisacáridos y glicoproteinas (Zrenner et al., 2006), y para la síntesis de los nucleótidos de pirimidina (Neuhard & Kelln, 1996; Michal, 1999) (Figura 5). De acuerdo con ello, se ha demostrado que la UMPK mejor estudiada, la de *Escherichia coli*, es retroinhibida por UTP (Serina et al., 1995) (Figura 6). Es interesante que también se ha descrito la activación de este enzima por GTP (Serina et al., 1995; Gagyi et al., 2003) (Figura 6), atribuyéndose esta activación a la necesidad de equilibrar los niveles de nucleótidos de pirimidina y de purina.

3. ENZIMAS MONODOMINIO Y BIDOMINIO.

Los seis enzimas de la familia aminoácido quinasa, con la excepción de algunas aspartoquinasas, son homo-oligómeros de un sólo tipo de subunidad, constituida en el caso de la carbamato quinasa, de la mayoría de las acetilglutamato quinasas, de las UMP quinasas y de algunas glutamato 5-quinasas, exclusivamente por el dominio AAK (Figura 7). Sin embargo los demás enzimas de esta familia contienen o pueden contener, además del AAK, N-terminal, un segundo dominio de menor tamaño, C-terminal,

distinto para cada enzima de la familia (Figura 7), denominado: 1) dominio transacetilasa en el caso de la acetilglutamato sintasa, porque se piensa participa en la catálisis de la transferencia del grupo acetilo; 2) dominio tipo ascomiceto en el caso de la acetilglutamato quinasa, por haberse descrito en las acetilglutamato quinasas de ascomicetos; 3) dominio PUA (Aravind et al., 1999) en el caso la mayoría de las glutamato 5-quinasas bacterianas, por su similitud con el dominio correspondiente de ciertos enzimas modificadores de RNA (Ishitani et al., 2002; Pan et al., 2003); y 4) dominio ACT (Chipman & Shaanan, 2001) en el caso de la aspartoquinasa, dominio similar al de otros enzimas en los que desempeña un papel regulador, que parece corroborarse para la aspartoquinasa con la presencia de lisina en este dominio, en la recientísima descripción de la estructura de la aspartoquinasa III de *E. coli* (Kotaka et al., 2006).

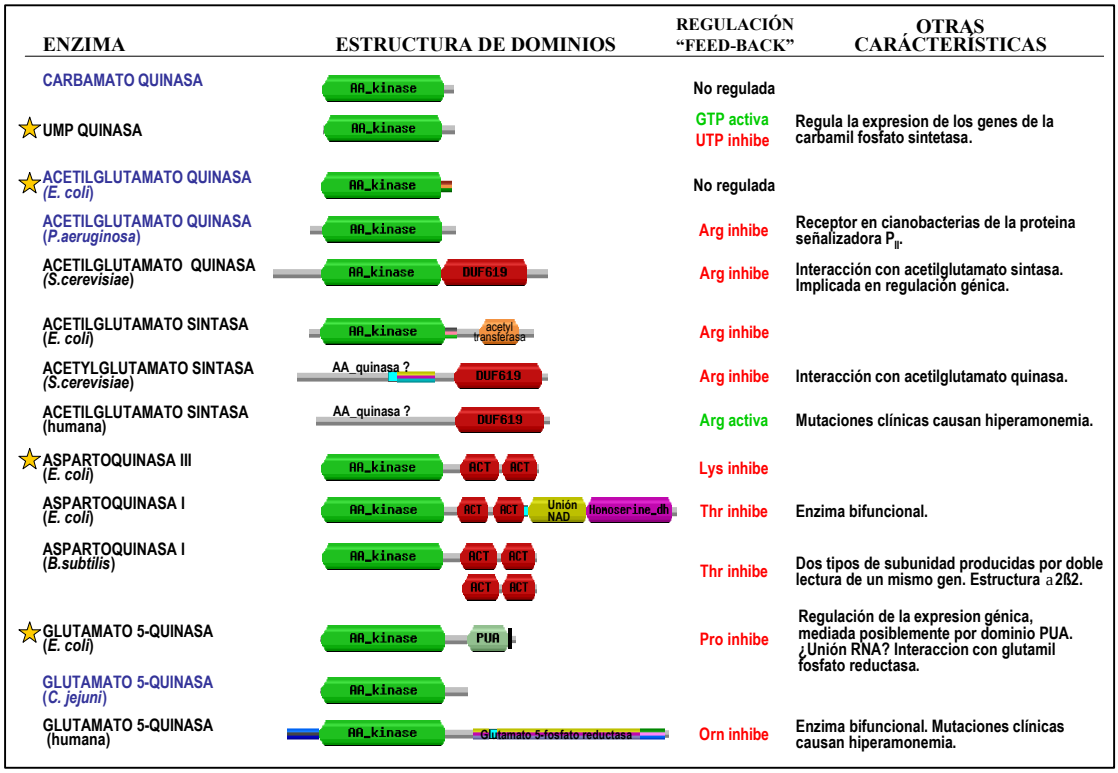


Figura 7- Esquema de la organización de dominios en distintos enzimas de la familia aminoácido quinasa, en la que se utiliza la simbologia de la base de datos de Pfam, con los dominios aminoácido quinasa representados en verde. En la primera columna se indica el organismo fuente y el nombre del enzima, coloreado en azul en los casos para los que conocemos la estructura tridimensional. Los enzimas incluidos en este trabajo se marcan con una estrella amarilla. En la tercera columna se indican los efectores para cada enzima.

La presencia de estos dominios plantea problemas funcionales que en la mayoría de casos quedan por responder. Por ejemplo, se ha descrito la unión a DNA de la acetilglutamato quinasa de levadura y su participación en la regulación de la expresión

génica en este organismo (Hall et al., 2004), y se puede especular sobre la posibilidad de que el dominio C-terminal ascomiceto de la NAGK de levadura sea responsable de la unión a DNA y de la regulación. También en levadura se ha demostrado la existencia de importantes interacciones físicas entre acetilglutamato sintasa y acetilglutamato quinasa, y es posible que este dominio C-terminal participe en la formación del metabolón formado por ambos enzimas (Abadjieva et al., 2001, Pauwels et al., 2003). Se ha descrito que la glutamato 5-quinasa de *Bacillus subtilis* tiene funciones de regulación de la expresión de genes (Ogura et al., 1994; Ogura & Tanaka, 1996), y se ha especulado que el dominio PUA sea responsable de esta función, sobre la base de un mecanismo basado en interacciones con RNA. Ya se ha dicho que en el caso de la aspartoquinasa la presencia del dominio ACT se asocia con la regulación feed-back del enzima (Kotaka et al., 2006).

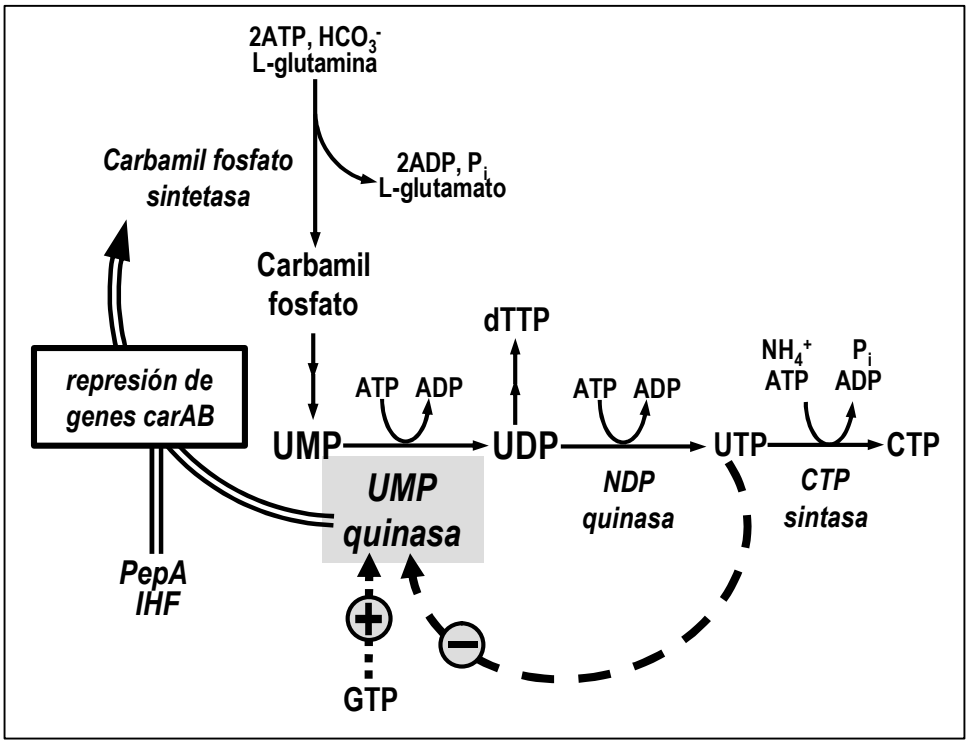


Figura 8. Ruta de biosíntesis de pirimidínas y papel de la UMP quinasa (sombreada en gris), demostrado en *Escherichia coli*. Las líneas continuas indican los distintos pasos de la ruta. La línea discontinua y la línea punteada indican la inhibición feedback por UTP y la activación por GTP respectivamente. La doble linea indica la participación de la UMP quinasa en el mecanismo de represión de los genes *carAB*, que codifican para la carbamilfosfato sintetasa, proceso en el que también participan las proteínas PepA e IHF.

Sin embargo, parece claro que en algunos enzimas de esta familia, regulación feed-back y regulación génica, cuando tienen lugar, no dependen de la existencia de un dominio extra. Tal es el caso de la regulación feed-back de la acetilglutamato quinasa y de la acetilglutamato sintasa por arginina, que sólo depende en el caso del primer

enzima, y que parece depender en el caso del segundo, de la existencia de una corta extensión N-terminal y de ciertos elementos específicos de secuencia, dentro del dominio AAK (Fernandez-Murga et al., 2004; Ramón-Maiques et al., 2006). Igualmente, la regulación feed-back de la UMP quinasa, así como la regulación negativa por altos niveles de pirimidinas, de la expresión de los genes para carbamil fosfato sintetasa en *Escherichia coli*, mediada por la UMP quinasa en combinación con las proteinas IHF y PepA (Kholti et al., 1998) (Figura 8), sólo requieren del dominio AAK de la UMP quinasa, ya que éste es el único dominio de este enzima. Por tanto, el dominio AAK tiene potencialidades multifuncionales: no sólo une los sustratos y cataliza la reacción, sino que también puede mediar la regulación feed-back e incluso la regulación de la expresión génica, mediante mecanismos que estaban en completa oscuridad al comienzo de este trabajo de tesis.

Un caso particular de presencia de dos dominio en la cadena polipeptídica se da en el caso del polipéptido bifuncional que incluye la glutamato 5-quinasa y la γ-glutamil fosfato reductasa en los organismos eucarióticos (Fujita et al., 2003; Strizhov et al., 1997; Hu et al., 1999). En este caso la quinasa carece de dominio PUA y es N-terminal, y la reductasa es C-terminal. Déficits en el ser humano de este complejo bienzimático se asocian con hiperamoniemia clínica por déficit de ornitina, aminoácido esencial para el funcionamiento del ciclo de la urea (Baumgartner et al., 2000). Precisamente en este complejo humano, denominado pirrolin 5-carboxilato sintetasa, se han identificado dos formas alternativas de "splicing" de las que una de ellas difiere en sólo dos residuos del dominio AAK de la otra forma de "splicing" (Hu et al., 1999). De estas dos formas, sólo la más corta es inhibida por ornitina, lo que posiblemente indica que la inhibición feed-back por ornitina es una función exclusiva del dominio AAK, e involucra la región donde se encuentran los dos residuos que pueden estar presentes o ausentes. En todo caso, este complejo bienzimático parece implicado en la canalización del γ-glutamil 5-fosfato, que es inestable y debe ser proporcionado a la reductasa si hay que evitar su ciclación a oxoprolina (Figura 9). Es más, el problema de la inestabilidad del γ-glutamil fosfato debe obligar también a que incluso las glutamato 5-quinasas monofuncionales se asocien con la reductasa también monofuncional, si debe evitarse la destrucción del γ-glutamil fosfato antes de su utilización por la reductasa. Resultados experimentales del laboratorio de Meister (Seddon et al., 1989) con los enzimas de *E. coli*, demuestran que la mayoría del γ-glutamil fosfato formado por la glutamato 5-quinasa se cicla espontáneamente a oxoprolina incluso en presencia de hidroxilamina en ausencia de la reductasa, pero no en su presencia incluso sin NADPH (el reductor utilizado por la

reductasa). Estos resultados son fuertemente sugerentes de la asociación entre ambos enzimas, al menos durante la catálisis.

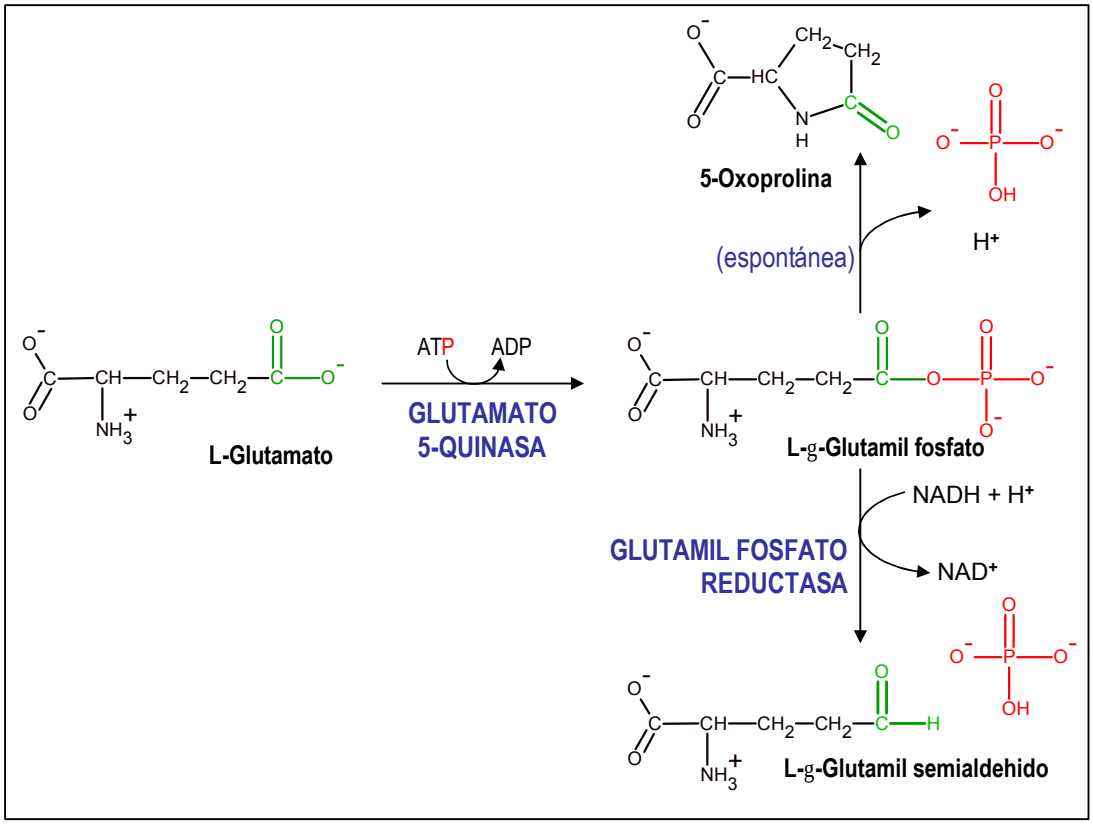


Figura 9. Síntesis de L-?-glutamil fosfato catalizada por la glutamato 5-quinasa. En rojo se representa el fosfato transferido desde et ATP al grupo carboxilato (en verde) del glutamato, que es liberado tanto si de manera espontánea el L-?-glutamil fosfato cicla, como si es transformado en L-?-glutamil semialdehido, reacción catalizada por la glutamilfosfato reductasa.

4. ELEMENTOS DE LOS QUE CONSTA ESTA TESIS

Esta tesis, configurada como memoria por publicaciones, incluye cuatro trabajos multiautor.

El primero es un estudio de la estructura de la aspartoquinasa III de *Escherichia coli*, la aspartoquinasa más simple de este organismo, implicada en la síntesis de lisina e inhibida por este aminoácido (Cassan et al., 1986). No se trata de una determinación experimental de la estructura, sino del modelizado de la estructura del dominio aminoácido quinasa de este enzima, a partir de la estructura de la acetilglutamato quinasa de *Escherichia coli*, enzima cuya estructura se ha determinado a alta resolución en nuestro laboratorio, en forma de complejo con sus dos sustratos o con análogos de

los sustratos (Ramón-Maiques et al., 2002; Gil-Ortiz et al., 2003). Un estudio de modelizado es, en un laboratorio cristalográfico, el último recurso cuando resulta imposible determinar la estructura de una proteína por procedimientos experimentales directos (principalmente cristalografía de rayos X). Este fue nuestro caso para la aspartoquinasa. Aunque no se refleje en este trabajo de tesis, he clonado y expresado el gen de la aspartoquinasa de *Mycobacterium tuberculosis* y he intentado su cristalización. También cloné y expresé la aspartoquinasa III de *Eschericha coli* e intenté cristalizarla, pero los resultados fueron infructuosos a pesar de la gran cantidad de esfuerzos invertidos. Como alternativa intenté entonces, conjuntamente con el Dr. Santiago Ramón Maiques, modelizar el dominio aminoácido quinasa de este enzima a partir de la estructura del mismo dominio de la acetilglutamato quinasa. Como el modelizado requiere un alineamiento fiel de la proteína molde y de la que es objeto de modelización, utilicé mutagénesis dirigida del molde sobre residuos estimados a partir de la estructura como cruciales para la unión de los sustratos y para la catálisis. Las mutaciones abordadas fueron siempre leves (por ejemplo, sustitución de lisina por arginina o de aspartato por glutamato), para tratar de evitar que las mutaciones causaran una inactivación completa del enzima, permitiendo así el análisis cinético de los resultados obtenidos. El paso siguiente fue tratar de inducir mutaciones equivalentes en la aspartoquinasa, para corroborar que los efectos eran similares a los observados con la acetilglutamato quinasa y por tanto que los residuos alineados se corresponden en uno y otro enzima. Una precaución adicional fue predecir los elementos de estructura secundaria en ambos enzimas. En el caso de la acetilglutamato quinasa esta predicción fue esencialmente correcta, a juzgar por la comparación de los elementos predichos con los observados en la estructura experimental de este enzima. Por tanto, pensamos que la prediccción sería también correcta para la aspartoquinasa, y el alineamiento de ambos enzimas se mejoró mediante la utilización, en las zonas dudosas, de la superposición de los elementos equivalentes de estructura secundaria de ambos enzimas. Finalmente se procedió a modelizar la aspartoquinasa y a interpretar a partir del modelo los resultados obtenidos en ensayos de mutagénesis realizados con otra aspartoquinasa de una fuente diferente (Kobashi et al., 1999). Como complemento de la predicción de la estructura del dominio AAK de la aspartoquinasa se utilizaron los datos resultantes de la deleción del dominio ACT, datos proporcionados por el trabajo experimental de Dña. Sandra Tavárez, coautora de este trabajo. El resultado obtenido puede ahora compararse con los resultados experimentales: otro grupo ha sido más afortunado que nosotros en la determinación, casi en el momento del depósito de este trabajo de tesis, de la estructura

de la aspartoquinasa III de *E. coli* (Kotaka et al., 2006). En el texto de dicha reciente publicación, así como en la Discusión final de esta memoria de tesis se comparan los resultados de nuestro modelo con la estructura experimental.

El segundo enzima objeto de esta Tesis ha sido la UMP quinasa de *Pyrococcus furiosus*, sujeto de dos de los cuatro manuscritos que componen esta memoria. Uno de estos manuscritos tiene carácter preparatorio: se trata de la descripción de la cristalización y estudios de difracción preliminares del enzima. En este manuscrito Juan Manuel Escamilla fue responsable de los aspectos preparativos, y yo de los cristalográficos. Finalmente, en un segundo trabajo describo la estructura cristalográfica de este enzima, a partir de estudios de difracción de rayos X, siendo yo la responsable principal de este trabajo, aunque en él ha participado también el Dr. Fernando Gil Ortiz, quien ha ayudado a superar alguna de las dificultades mayores del trabajo cristalográfico.

Otro enzima cuya determinación estructural es objeto de la presente tesis es la glutamato 5-quinasa de *Eschericha coli*. Este trabajo ha sido fruto de una colaboración con los Drs. Isabel Pérez-Arellano y Javier Cervera, del Centro de Investigación Príncipe Felipe. La historia de mi relación con este enzima es relativamente compleja: el gen había sido clonado, y el enzima expresado y purificado por la Dra. Pérez-Arellano, quien, en colaboración con nuestro grupo a través del Dr. Fernando Gil, había iniciado la producción de cristales (Pérez-Arellano et al., 2004). La producción de estos fue muy laboriosa: fueron precisos varios meses de espera para que en alguna de las gotas se desarrollaran cristales apropiados para estudios de difracción, cristales que fueron poco reproducibles, lo que en la práctica bloqueó todos los intentos de obtención de las fases. Me incorporé al proyecto cuando el Dr. Fernando Gil lo tuvo que abandonar y se depositó la estructura de la glutamato 5-quinasa de *Campylobacter jejuni*, un enzima que, a diferencia del de *E. coli*, carece de dominio PUA. Usando dicha estructura depositada (PDB 2AKO) me ha sido posible obtener las fases cristalográficas por reemplazo molecular, y determinar la estructura de la glutamato 5-quinasa de *E. coli* en dos formas cristalinas diferentes, ambas obtenidas en presencia de glutamato y de ADP. La responsabilidad básica de esta determinación estructural, y el protagonismo principal han sido míos en este proyecto.

5. ¿POR QUÉ ME HE OCUPADO DE ESTOS ENZIMAS?

Además de las implicaciones prácticas apuntadas más arriba, razones científicas, históricas y estratégicas han determinado el contenido de esta tesis. Haré a continuación un tratamiento conjunto de todas ellas.

El grupo en cuyo seno he desarrollado mi trabajo había publicado en 1999, antes de mi incorporación al laboratorio, la estuctura de la carbamato quinasa (Marina et al., 1999), un enzima poco estudiado previamente, que manifestó un plegamiento característico y nuevo que llevó al grupo a buscar otras proteínas que lo presentaran.

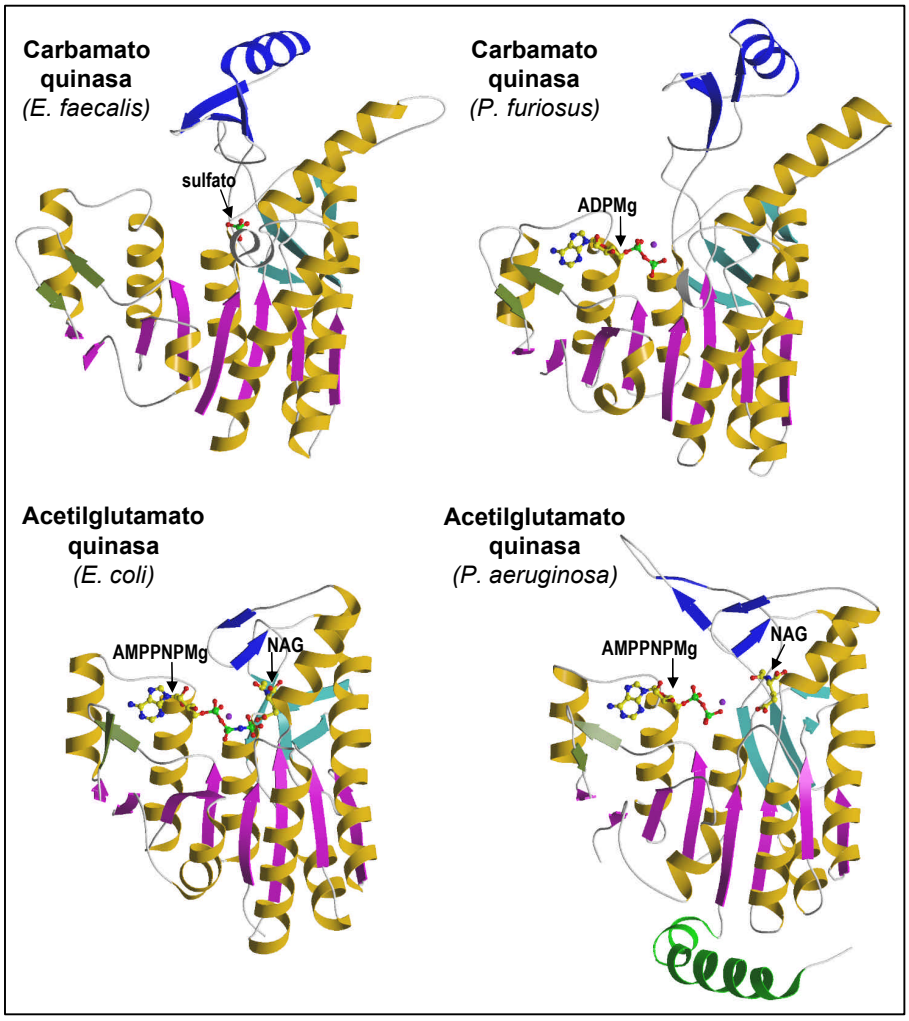


Figura 10. Representación en modelo de cintas de las subunidades de la carbamato quinasa de *Enterococcus faecalis* con un ion sulfato (PDB 1B7B), de la carbamato quinasa de *P. furiosus* con ADPMg (PDB 1E19), de la acetilglutamato quinasa de *E. coli* con AMPPNPMg y N-acetil-L-glutamato (NAG) (PDB 1GS5) y de la acetilglutamato quinasa de *Pseudomonas aeruginosa* con ADP y NAG (PDB 2BUF). Los elementos de la hoja ß principal estan coloreados en magenta y las a hélices en amarillo a excepción de la hélice N-terminal de la acetilglutamato quinasa de *T. maritima* que está en verde. Los elementos que forman la hoja ß secundaria están en cyan, los de la horquilla ß comun a todos los enzimas en verde y los elementos particulares de las acetilglutamato quinasas o carbamato quinasas, en azul oscuro.

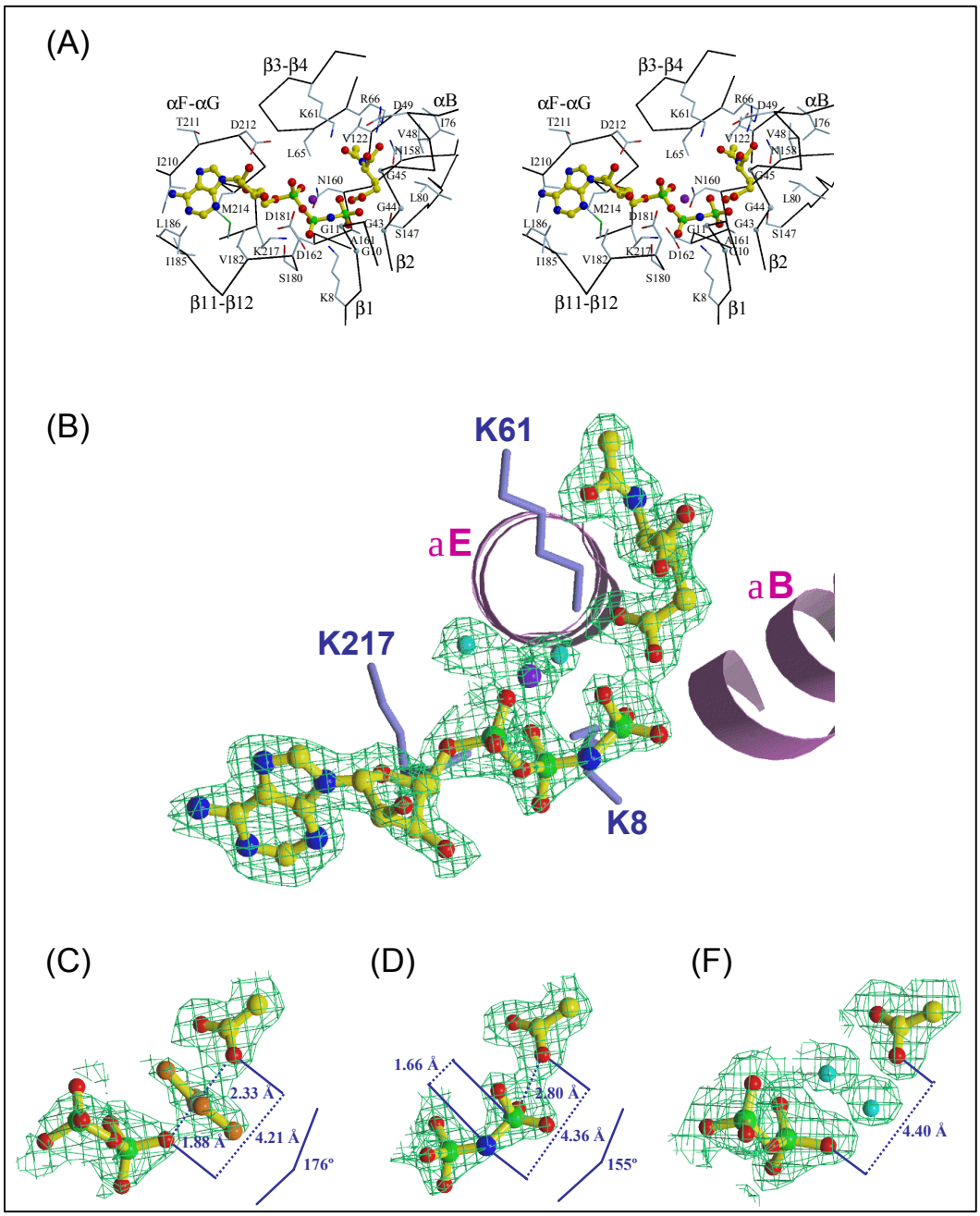
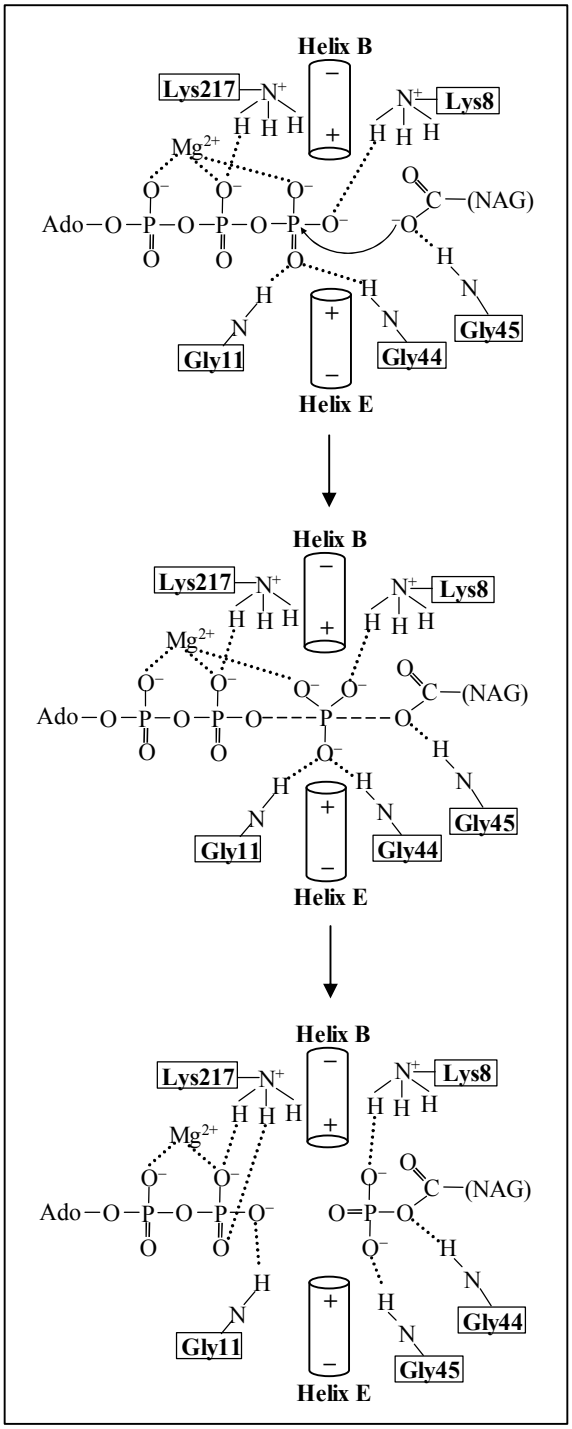


Figura 11. Unión de los sustratos y catálisis en la N-acetil-L-glutamato quinasa (NAGK), el paradigma de la familia aminoácido quinasa. (A) Vista estereoscópica del centro activo de NAGK en la que los elementos del centro activo se representan como trazos entre carbonos alfa, y representando las cadenas laterales que participan en la unión de los sustratos, AMPPNPMg y N-acetil-L-glutamato (NAG), representados en colores. (B) Mapa de densidad electrónica $2F_{obs}$-F_{calc} para el MgAMPPNP y el NAG unidos a acetilglutamato quinasa de *E. coli* (se incluyen dos moléculas de agua, en azul claro). Se muestran las hélices B y E y las cadenas laterales de las lisinas 8, 61 y 217. (C-F) Mapa $2F_{obs}$-F_{calc} y modelo de los elementos implicados en la transferencia del grupo fosforilo en los complejos de NAGK de *E. coli* con MgADP-AlF_4^--NAG (C), MgAMPPNP-NAG (D) y MgADP-NAG (F). En (C) dos moleculas de agua localizadas entre ambos sustratos se representan en azul claro. El PDB correspondiente a los paneles A, B y D es 1GS5, al panel C, 1OH9 y al panel F, 10HA. (A y B) tomadas de Ramon-Maiques *et al.* (2002) *Structure 10, 329-342* y (C y D), de Gil-Ortiz *et al. (*2003) *J. Mol Biol. 331, 231-244*, con permiso de los autores.

Su búsqueda dio frutos, primero con lo que parecía ser una carbamil fosfato sintetasa arqueal pero que luego el grupo mismo demostró era realmente una carbamato quinasa hipertermofílica (Uriarte et al., 1999; Ramón-Maiques et al., 2000), que unía el nucleótido con alta afinidad, proveyendo así la primera visión del centro activo uniendo al menos uno de sus sustratos (o productos, dependiendo de la dirección en que se considere la reacción, que es reversible), el MgADP (Ramón-Maiques et al., 2000).



Poco despues, en 2002, cuando me incorporé yo al laboratorio, el grupo describió la estructura de la acetilglutamato quinasa de *Eschericha coli* (Ramón-Maiques et al., 2002), corroborando que este enzima, que no es sujeto de regulación alostérica, presenta esencialmente el mismo plegamiento que la carbamato quinasa, sirviendo además esta estructura, determinada a alta resolución, para caracterizar plenamente el centro activo y los diferentes pasos del proceso catalítico (Ramón-Maiques et al., 2002; Gil-Ortiz et al., 2003).

Figura 12. Esquema ilustrativo de la transferencia del grupo fosforilo en la reacción catalizada por la acetilglutamato quinasa, señalando la implicación de los residuos Gly11, Gly44, Gly45, Lys8 y Lys217 y de los dipolos de las a hélices B y E. La linea curva indica el ataque nucleofílico del O del grupo ?-COO^- del acetilglutamato sobre el átomo ?-P del ATP. Tomada de de Gil-Ortiz *et al.* (2003) *J. Mol Biol. 331, 231-244*, con permiso de los autores.

Puede decirse que con el muy detallado estudio de la acetilglutamato quinasa se cerraba la primera fase de la caracterización estructural de la familia aminoácido quinasa, fase que había permitido identificar los elementos básicos: el plegamiento de la cadena polipeptídica (Figura 10), la union de sustratos y la catálisis (Figura 11 y12), y también lo que parecía ser una organización homodimérica monótona (Figura 13) que mis resultados actuales, junto con los recién publicados con la aspartoquinasa (Mas-Droux et al., 2006; Kotaka et al., 2006), han desmentido como universal para los enzimas de esta familia.

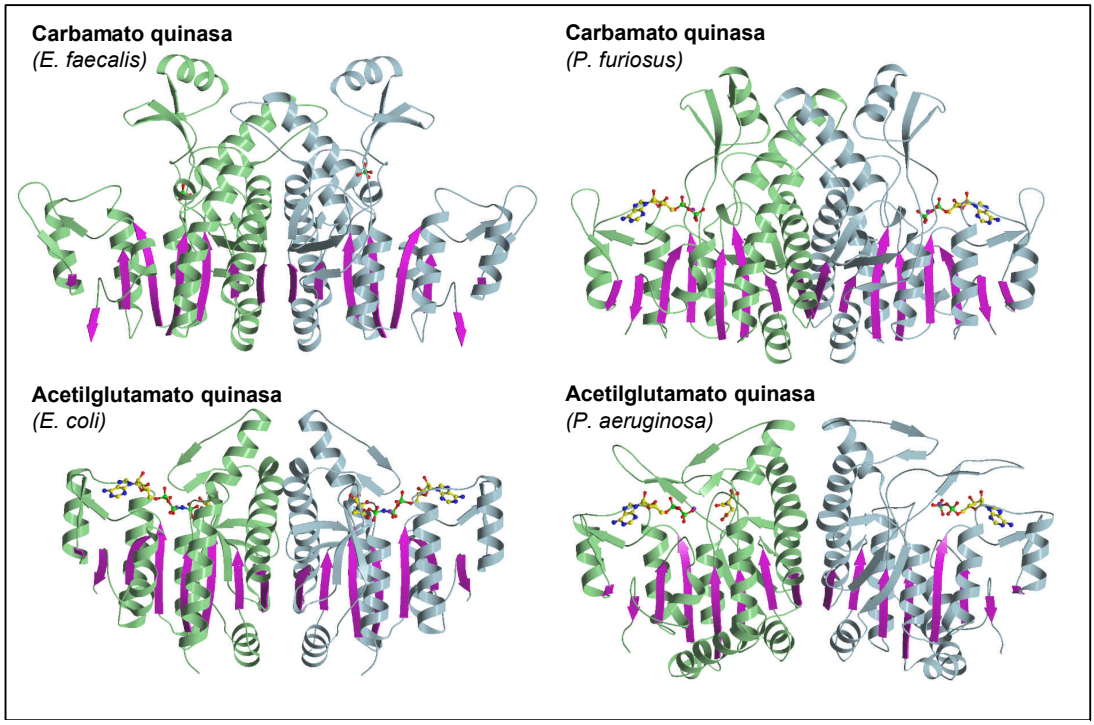


Figura 13. Representación en modelo de cintas de los dímeros de la carbamato quinasa de *Enterococcus faecalis* con un ion sulfato (PDB 1B7B), de la carbamato quinasa de *P. furiosus* con ADPMg (PDB 1E19), de la acetilglutamato quinasa de *E.coli* con AMPPNPMg y N-acetil-L-glutamato (PDB 1GS5) y de la acetilglutamato quinasa de *Pseudomonas aeruginosa* con ADP y N-acetil-L-glutamato (PDB 2BUF). Las subunidades se colorean en verde y azul, con los elementos de la hoja ß principal en magenta.

La etapa siguiente del estudio de la familia aminoácido quinasa, suponía asentar sobre bases estructurales la complejidad presente en muchos de sus miembros, complejidad que incluye tanto la regulación feed-back como la integración de dominios diferentes y la generación de arquitecturas complejas, incluyendo la formación de oligómeros superiores, en ocasiones ya detectados antes mediante estudios bioquímicos. Los diferentes miembros del laboratorio nos lanzamos a la determinación de la estructura de la aspartoquinasa, de la acetilglutamato quinasa inhibida por la arginina, de la UMP quinasa, y, en colaboración con el grupo del Dr. Cervera, de la glutamato 5-quinasa, además de lanzarnos también al estudio de la acetato quinasa, enzima que

sabíamos diferente y para el que nuestros esfuerzos no nos permitieron llegar primero a la determinación de la estructura de este enzima. Los resultados recogidos en esta tesis dan cuenta fehaciente de aquellos esfuerzos que han dado lugar a la publicación de trabajos, pero no de aquellos que no han tenido dicha conclusión final (acetato quinasa, muchos estudios con aspartoquinasa). En conjunto, el balance global de las estructuras determinadas por el laboratorio en esta fase parece razonablemente positivo: se han determinado las estructuras de dos acetilglutamatato quinasas inhibidas por arginina (por otros miembros del grupo; Ramón-Maiques et al., 2006) (Figura 14), de UMP quinasa (esta tesis, publicado en 2005) y de glutamato 5-quinasa (manuscrito en proceso de consideración editorial en el momento de la defensa de esta tesis). Los resultados han permitido la caracterización de arquitecturas diversas a partir de un único plegamiento de subunidad, han delineado en todos los casos estudiados la unión del sustrato que se fosforila, dando cuenta de la diversidad debida a la variedad de sustratos, han permitido integrar el domino PUA en la estructura de la glutamato 5-quinasa, y han permitido caracterizar plenamente la modulación alostérica por arginina en acetilglutamato quinasa, y sentar las bases para entender la regulación alostérica de UMP quinasa y glutamato 5-quinasa, así como la regulación génica por UMP quinasa.

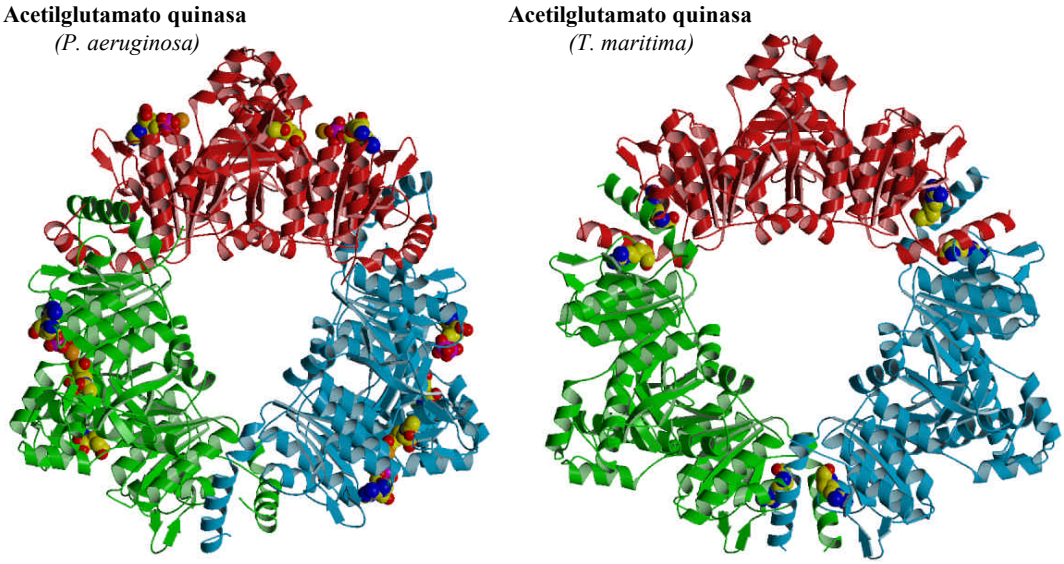


Figura 14. Representación en modelo de cintas de los hexameros de la N-acetil-L-glutamato quinasa de *Pseudomonas aerugiosa* en presencia de ADP y acetilglutamato (PDB 2BUF) y de *Thermotoga maritima* en presencia de arginina (PDB 2BTY). Cada dímero se representa en un color diferente y los ligandos se representan en modelo de bolas. Tomado de Ramón-Maiques et al., (2006) *J. Mol. Biol*. 356, 695-713.

Este trabajo de tesis abre las puertas a una nueva etapa de la investigación en esta familia, la referente a la formación de complejos supramoleculares, ámbito en el que el laboratorio está plenamente inmerso en el momento en que escribo esta tesis. Estoy

convencida de que los resultados presentados aquí van a ser cruciales para guiar el entendimiento de los complejos y de sus implicaciones funcionales. Y confío también en que, al menos en el caso de la UMP quinasa, mis hallazgos sirvan de base para el desarrollo de nuevos antibióticos.

OBJETIVOS

OBJETIVOS

El objetivo último de mi trabajo ha sido profundizar en el conocimiento de los enzimas de la familia aminoácido quinasa, intentando en particular sustentar sobre bases estructurales aspectos funcionales clave de estos enzimas que determinan buena parte de sus propiedades catalíticas y reguladoras, que implican a otros dominios además del dominio aminoácido quinasa, que resultan en arquitecturas más complejas que las inicialmente descritas para los primeros enzimas caracterizados de esta familia, que incluso resultan en asociaciones entre distintas proteínas, o que determinan en algunos casos la implicación de miembros de esta familia en procesos de regulación génica.

La consecución con éxito de este objetivo general y de los objetivos específicos que se indican a continuación puede ser de interés para el desarrollo de compuestos de interés antimicrobiano o herbicida, habida cuenta del carácter estrictamente no animal de la acetilglutamato quinasa, la aspartoquinasa y la UMP quinasa. La estructura de la glutamato 5-quinasa debe ayudar a entender las hiperamoniemias congénitas humanas debidas a fallos en el complejo pirrolin 5-carboxilato sintetasa.

Más específicamente, me he ocupado aquí de las siguientes cuestiones:

1. Corroborar mediante mutagénesis dirigida el papel en la catálisis y en la unión de los sustratos, asignado sobre la base de la estructura, a determinados residuos de la acetilglutamato quinasa de *Escherichia coli*.

2. Caracterizar el papel en la catálisis y en la unión de los sustratos de determinados residuos de la aspartoquinasa III de *Escherichia coli*, guiando de este modo los alineamientos de la secuencia de aminoácidos de esta aspartoquinasa con la secuencia de la acetilglutamato quinasa de *Escherichia coli*.

3. Construir sobre la base de este alineamiento un modelo de la estructura tridimensional del dominio aminoácido quinasa de la aspartoquinasa III de *Escherichia coli*, con la finalidad de tratar de entender la unión de los sustratos, la catálisis, la regulación, y en lo posible la arquitectura de este enzima, y, por extensión, de otras aspartoquinasas.

4. Determinar mediante técnicas cristalográficas de difracción de rayos X la estructura tridimensional de la UMP quinasa de la arquea *Pyrococcus furiosus*, sóla y asociada a sus sustratos y efectores, con la finalidad de entender las diferencias entre UMP quinasas eucarióticas y bacterianas/arqueales, y de justificar en la mayor medida posible las propiedades catalíticas, de regulación, arquitectónicas del hexámero, y reguladoras génicas de las UMP quinasas bacterianas.

5. Determinar mediante técnicas cristalográficas de difracción de rayos X la estructura tridimensional de la glutamato 5-quinasa de *Eschericha coli*, con el objeto de entender la integración del dominio PUA C-terminal presente en este enzima en la arquitectura global de este miembro de la familia aminoácido quinasa, y también para intentar caracterizar sobre bases estructurales la arquitectura oligomérica, especificidad de sustratos y regulación de este enzima, así como su posible asociación con el enzima siguiente de su ruta metabólica.

Los manuscritos de los que se compone esta memoria de tesis son los siguientes:

Capítulo 1

Título: Site-directed mutagenesis of *Escherichia coli* acetylglutamate kinase and aspartokinase III probes the catalytic and substrate-binding mechanisms of these amino acid kinase family enzymes and allows three-dimensional modelling of aspartokinase.

Autores: Clara Marco-Marín, Santiago Ramón-Maiques, Sandra Tavárez y Vicente Rubio.

Referencia : *Journal of Molecular Biology* (2003) **334**, 459-476.

Capítulo 2

Título: First-time crystallization and preliminary X-ray crystallographic analysis of a bacterial-archaeal type UMP kinase, a key enzyme in microbial pyrimidine biosynthesis.

Autores: Clara Marco-Marín, Juan Manuel Escamilla-Honrubia y Vicente Rubio.

Referencia : *Biochim. Biophys. Acta.* (2005) **1747**, 271-275.

Capítulo 3

Título: The crystal structure of *Pyrococcus furiosus* UMP kinase provides insight into catalysis and regulation in microbial pyrimidine nucleotide biosynthesis.

Autores: Clara Marco-Marín, Fernando Gil-Ortiz y Vicente Rubio.

Referencia : *Journal of Molecular Biology* (2005) **352**, 438-454.

Capítulo 4

Título: A novel two-domain architecture within the amino acid kinase enzyme family revealed by the crystal structure of *Escherichia coli* glutamate 5-kinase.

Autores: Clara Marco-Marín, Fernando Gil-Ortiz, Isabel Pérez-Arellano, Javier Cervera, Ignacio Fita y Vicente Rubio.

Referencia : enviado a *Journal of Molecular Biology*.

Capítulo 1

Site-directed mutagenesis of *Escherichia coli* acetylglutamate kinase and aspartokinase III probes the catalytic and substrate-binding mechanisms of these amino acid kinase family enzymes and allows three-dimensional modelling of aspartokinase.

Trabajo publicado en

Journal of Molecular Biology **(2003) 334, pags. 459-476**

doi:10.1016/j.jmb.2003.09.038 J. Mol. Biol. (2003) 334, 459–476

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Site-directed Mutagenesis of *Escherichia coli* Acetylglutamate Kinase and Aspartokinase III Probes the Catalytic and Substrate-binding Mechanisms of these Amino Acid Kinase Family Enzymes and Allows Three-dimensional Modelling of Aspartokinase

Clara Marco-Marín†, Santiago Ramón-Maiques†, Sandra Tavárez and Vicente Rubio*

Department of Genomics and Proteomics, Instituto de Biomedicina de Valencia Consejo Superior de Investigaciones Científicas (IBV-CSIC), C/Jaime Roig 11 46010 Valencia, Spain

We test, using site-directed mutagenesis, predictions based on the X-ray structure of *N*-acetyl-L-glutamate kinase (NAGK), the paradigm of the amino acid kinase protein family, about the roles of specific residues on substrate binding and catalysis. The mutations K8R and D162E decreased $V^{[sustrate]=\infty}$ 100-fold and 1000-fold, respectively, in agreement with the predictions that K8 catalyzes phosphoryl transfer and D162 organizes the catalytic groups. R66K and N158Q increased selectively K_m^{Asp} three to four orders of magnitude, in agreement with the binding of R66 and N158 to the C^{α} substituents of NAG. Mutagenesis in parallel of aspartokinase III (AKIII phosphorylates aspartate instead of acetylglutamate), another important amino acid kinase family member of unknown 3-D structure, identified in AKIII two residues, K8 and D202, that appear to play roles similar to those of K8 and D162 of NAGK, and supports the involvement of E119 and R198, similarly to R66 and N158 of NAGK, in the binding of the amino acid substrate, apparently interacting, respectively, with the α-NH_3^+ and α-COO^- of aspartate. These results and an improved alignment of the NAGK and AKIII sequences have guided us into 3-D modelling of the amino acid kinase domain of AKIII using NAGK as template. The model has good stereochemistry and validation parameters. It provides insight into substrate binding and catalysis, agreeing with mutagenesis results with another aspartokinase that were not considered when building the model.

AKIII is homodimeric and is inhibited by lysine. Lysine may bind to a regulatory region that is C-terminal to the amino acid kinase domain. We make a C-terminally truncated AKIII (AKIIIt) and show that the C-region is involved in intersubunit interactions, since AKIIIt is found to be monomeric. Further, it is inactive, as demanded if dimer formation is essential for activity. Models for AKIII architecture are proposed that account for these findings.

Keywords: acetylglutamate kinase; lysine-sensitive aspartokinase; amino acid kinase; site-directed mutagenesis; structure modelling

*Corresponding author

† C.M.-M. and S.R.-M. contributed equally to this work.

Abbreviations used: ACT, acronym for a recently characterized domain; the, initials stand for aspartate kinase, chorismate mutase and Tyr A, from proteins known to host this domain; AK, aspartokinase; AKIII, lysine-sensitive aspartokinase; AKIIIt, aspartokinase III truncated at codon 298; IPTG, isopropyl-β-D-thiogalactoside; NAG, *N*-acetyl-L-glutamate; NAGK, *N*-acetyl-L-glutamate kinase; PDB, Protein Data Bank.

E-mail address of the corresponding author:
rubio@ibv.csic.es

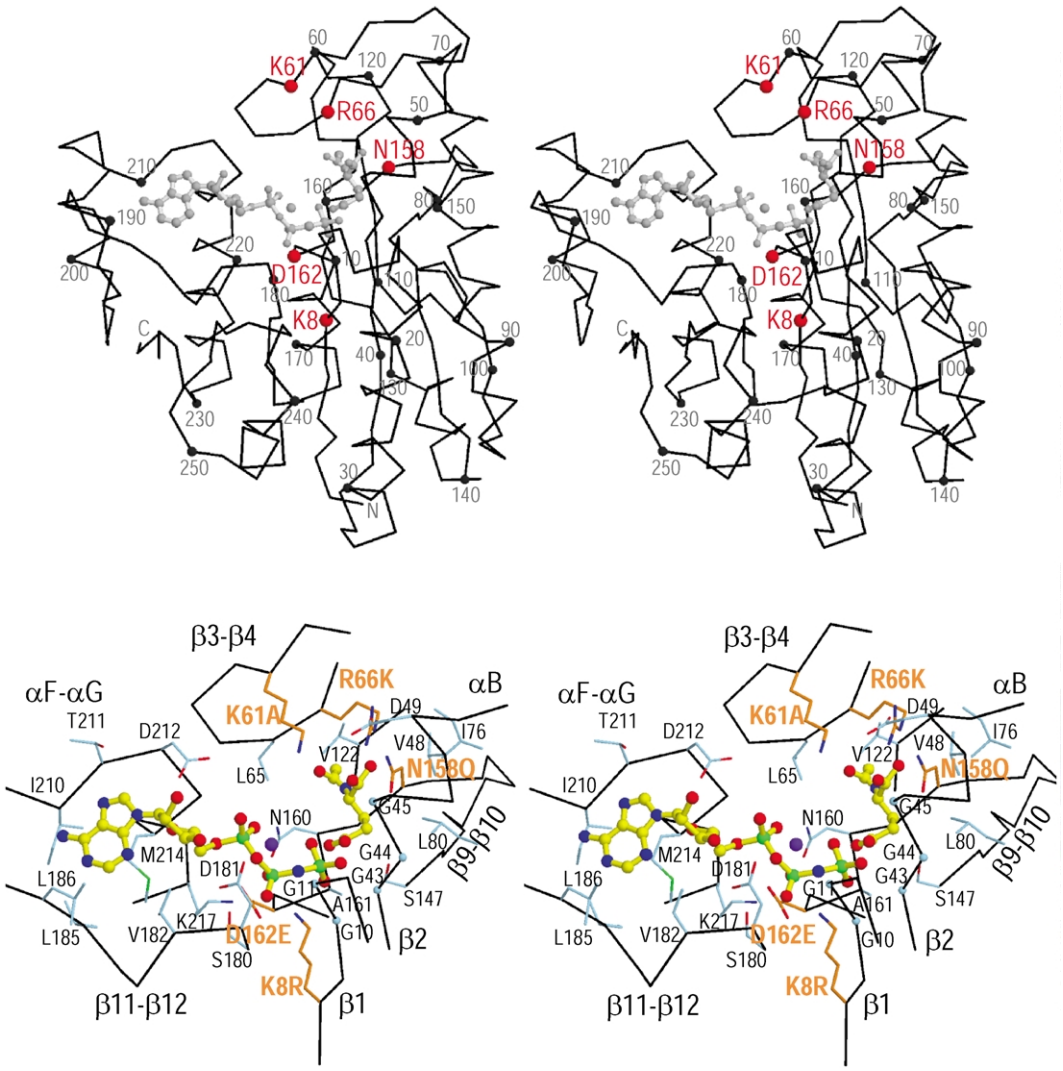


Figure 1. Stereo views of the localization in the *E. coli* NAGK structure (PDB accession number 1GS5) of the residues mutated here. Top panel: position of the mutated residues (in red) in the main chain of the enzyme subunit. Bottom panel: detail of the active center to highlight the interactions of the side-chains of the mutated residues (in orange) with the bound substrates.

Introduction

The amino acid kinase enzyme family (Pfam PF00696†)[1] groups on the basis of sequence homology the enzymes carbamate kinase, *N*-acetyl-L-glutamate kinase (NAGK), 5-glutamyl kinase, and the major domain of the enzyme aspartokinase. All these enzymes catalyze essentially the same type of reaction, the reversible transfer of the γ-phosphoryl group of ATP to a carboxylate or carbamate group, resulting in the synthesis of an acylphosphate anhydride. Two additional enzymes catalyzing other types of reactions, the bacterial UMP kinase and the bacterial acetylglutamate synthetase, are recognized as members of the family, although only the former of these enzymes uses ATP and makes an anhydride, in this case a phosphate–phosphate anhydride.

The structures of two enzymes of this enzyme family, carbamate kinase and NAGK, have been determined recently at atomic resolution using X-ray crystallography.[2–4] The two enzymes are homodimers and share the same fold, consisting of an eight element, mainly parallel, central open β sheet that is sandwiched between α helices and that continues across the intersubunit surface to form with the corresponding sheet of the other subunit a sheet of 16 elements that spans the entire homodimer. One independent active center is formed in each subunit by several loops and by two α helices that emerge from the C-edge of the sheet. Although studies with a hyperthermophilic carbamate kinase revealed the mode of binding of MgADP,[3] the complete active center of an enzyme of this family has been characterized only for the NAGK of *Escherichia coli* thanks to the

† http://www.sanger.ac.uk/cgi-bin/Pfam

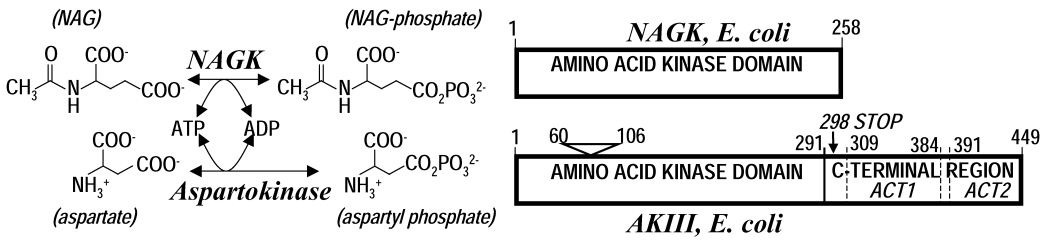


Figure 2. Reactions catalyzed by NAGK and AKIII, and schematic bar representation of the corresponding enzyme polypeptides, with indication of the different enzyme domains. Figures denote residue number. The AKIII sequence spanning residues 60–106 has no correspondence in the NAGK sequence and is schematized as an insertion (triangle above AKIII). The residue at which the AKIII chain has been truncated to delete its C-region is indicated with an arrow.

determination of the structure of the ternary complex of this enzyme with the substrate *N*-acetyl-L-glutamate (NAG) and with the inert MgATP analog MgAMPPNP[4] (Figure 1) and, more recently, with NAG, MgADP and the transition-state planar analog AlF_4.[5] A number of predictions concerning the roles of different enzyme residues in substrate binding and catalysis have been formulated[4,5] on the basis of the structure of these complexes, but these predictions have not been put to test until now. We report here the effects on the kinetic parameters of the enzyme of mutations affecting five residues, three of them proposed to be involved in the catalytic process, Lys8, Lys61 and Asp162, and two residues proposed to be involved in binding NAG and in conferring the enzyme specificity for this substrate, Arg66 and Asn158.[4,5] Four of the mutations tested, K8R, R66K, N158Q and D162E, have been chosen to be mild changes to avoid complete enzyme inactivation and to allow meaningful kinetic analysis, with the purpose of revealing experimentally the specific role of each mutated residue. The fifth change, K61A, was intended to test whether the positive charge on the side-chain of Lys61 plays a catalytic role, as suggested by some structural data.[4,6]

The present results have been exploited to try to obtain information on aspartokinase (AK), another enzyme of the amino acid kinase family (see Pfam group PF00696 and Ramón-Maiques *et al.*[4]) for which structural information is lacking. AK catalyzes the same reaction as NAGK (Figure 2), except for the use of aspartate rather than NAG as the phosphorylatable substrate,[7] and is endowed with important biotechnological potential, given its key controlling role in the biosynthesis of the essential amino acids lysine and isoleucine,[8,9] with its feed-back inhibition by lysine apparently limiting the content of this essential amino acid in cereals.[9] In addition, *Mycobacterium tuberculosis* deficient in AK fails to synthesize its cell wall properly and becomes more susceptible to antituberculous agents,[10] rendering AK a potential target in the fight against tuberculosis. Therefore, it would appear important to gather more information on AK, including information about its 3-D structure. Unhappily, although crystals of AKIII of *E. coli* (the form of AK involved in lysine synthesis in this microorganism[8]) have been generated,[11] the crystallographic structure of an AK has not been determined. We reasoned that if we could ascribe to individual AK residues specific roles on substrate binding and catalysis, this information would be of much use toward understanding the function of AK, and it might guide efforts to construct a structural model of the amino acid kinase domain of this enzyme. We have chosen for this study AKIII of *E. coli* because it exhibits a relatively simple quaternary structure[12] compared to other AKs,[13–15] being homodimeric as NAGK,[4]

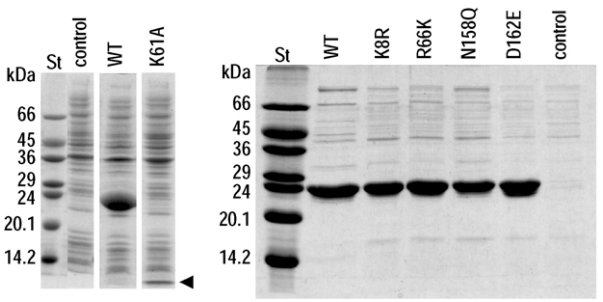


Figure 3. Expression and purification of wild-type (WT) and mutant forms of NAGK, monitored by SDS-PAGE (15% (w/v) polyacrylamide gels). Left panel: whole-cell extracts of *E. coli* BL21(DE3) transformed with either the empty parental plasmid (control), or with the plasmid pNAGK24, which encodes the wild-type enzyme (WT), or with plasmid pNAGK61A, which encodes the K61A mutant form of NAGK (K61A). Approximately equivalent numbers of *E. coli* cells collected after three hours induction with IPTG were boiled for five minutes in electrophoresis sample solution containing SDS, and were subjected to SDS-PAGE and staining with Coomassie brilliant blue.[40] A large NAGK band is observed in the WT track but not in the K61A track. In the latter, a small extra band near the front is marked with an arrowhead. Right panel: analysis of the purified wild-type and mutant enzyme forms, as indicated. Control denotes material purified from extracts of cells transformed with the parental empty pET plasmid. All lanes contain equivalent amounts of protein, except the control lane, in which the amount of protein is only 10%, to try to approximate the amount of host cell contamination in the purified enzyme samples. St, molecular mass markers (Dalton Mark VII-L, Sigma).

and because it has been overexpressed successfully in active form in *E. coli*.[11] We have used the results of the present study on NAGK mutations to set a reference frame for judging the effects of mutations that we introduce now into the AKIII of *E. coli*,[7,8] at five residues that we propose, on the basis of sequence alignment with NAGK, to play roles in catalysis and in the binding of the substrates similar to those of their counterparts in NAGK. For the same reason as in the mutagenesis of NAGK, mild amino acid changes have been introduced in the mutant AKIII forms.

The results of this site-directed mutagenesis study have provided experimental justification and clues for attempting structural modelling of the amino acid kinase domain of AKIII on the basis of the structure of NAGK. We build and present here this structural model, which accounts for the present site-directed mutagenesis results, and for the results of prior mutagenesis work carried out on the heterooctameric ($\alpha_4\beta_4$) aspartokinase of *Thermus flavus*.[16]

In addition to the amino acid kinase domain, the AKIII polypeptide (and the α and β polypeptides of other types of AK) contains a C-terminal region (residues 292–449 of AKIII) that includes a 71 residue (residues 309–384) small domain belonging to the ACT domain family (Pfam accession number PF01842)[17] and another similar domain (residues 391–449) that may be, as in the AKII from *B. subtilis* (see Pfam PF01842 alignment), another ACT domain (Figure 2). ACT domains are regulatory ligand-binding domains having a characteristic βαββαβ structure that have been found in a wide variety of regulated proteins.[17] Aspartokinases are feed-back inhibited by the final amino acid products of the biosynthetic routes to which they belong and, in particular, AKIII is inhibited by the final product of its pathway, lysine.[7] We test experimentally here the role of the C-terminal region of *E. coli* AKIII by deleting this region, determining the consequences of this deletion, and using the observations as the basis to propose models for the architecture of dimeric AKIII.

Results

Preparation of site-directed mutant forms of NAGK

Figure 1 localizes in the NAGK chain fold and relative to the bound substrates the five NAGK residues, Lys8, Lys61, Arg66, Asn158 and Asp162, that have been mutated here (mutations introduced, K8R, K61A, R66K, N158Q and D162E). The structure of NAGK[4,5] (Figure 1, lower panel) allows us to propose the involvement of K8 in abstracting a negative charge from the bipyramidal phosphorus transition state; of D162 in the organization of the catalytic groups (mainly the positive ξ-NH_3^+ of K8 and K217); and of R66 and N158 in the binding of the substituents of the C^{α} of NAG. The structural data for both NAGK[4] and carbamate kinases[2,3] suggested the involvement of the positive charge on the side-chain of K61 on the catalysis of phosphoryl group transfer, although more recent structural results do not favor such possibility.[5]

Upon IPTG induction of *E. coli* transformed with the appropriate pNAGK24-derived plasmids,[18] the wild-type enzyme and the enzyme forms carrying the individual mutations K8R, R66K, N158Q and D162E were expressed in high yield, in soluble form, by BL21(DE3) *E. coli* cells transformed with the corresponding plasmids (shown in Figure 3, left panel, for wild-type NAGK), and were purified

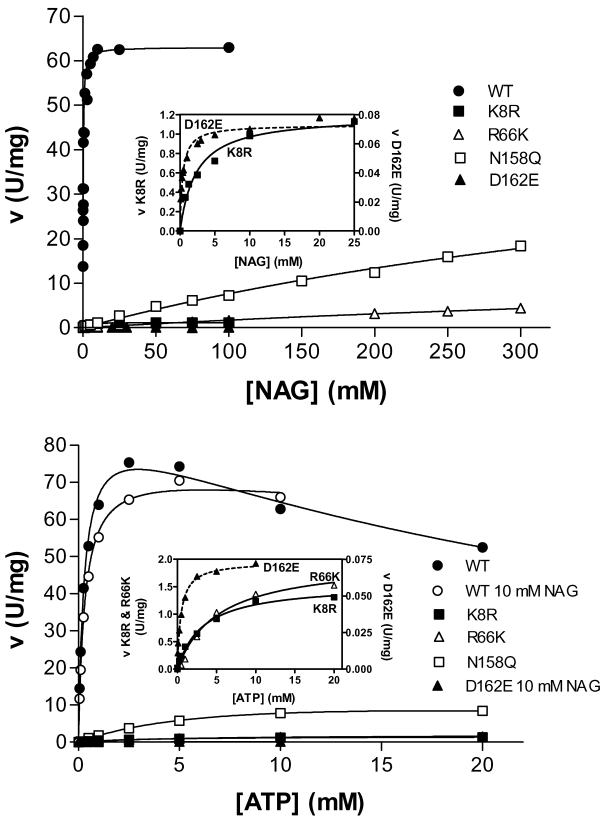


Figure 4. Enzyme activity of wild-type and mutant forms of NAGK as a function of the concentrations of (top panel) *N*-acetyl-L-glutamate (NAG) or (bottom panel) ATP. The larger graphs plot the results for all the enzyme forms on the same scale, whereas the insets zoom on the forms with the lower activities, and give the activity for D162E on the widest scale (right-hand *Y*-axis). Plots for variable ATP at both 0.1 M and 10 mM NAG are given for wild-type NAGK, to allow comparison with the D162E mutant, which has been assayed at 10 mM fixed NAG, and with the other mutants, which have been assayed at 0.1 M NAG. For the wild-type enzyme and for the N158Q mutant, the lines fit to the data for variable ATP correspond to hyperbolic kinetics combined with inhibition by excess substrate, using the kinetic constants given in Table 1. In all other cases, the lines are hyperbolae defined by the kinetic constants given in Table 1 for each enzyme form.

to essential homogeneity (Figure 3, right panel) by a simple three-step procedure (see Materials and Methods). In contrast, the K61A mutation did not yield enzyme (monitored by SDS-PAGE; Figure 3, left panel), although a low-molecular mass protein band was observed in the gels (arrowhead, Figure 3, left panel), suggesting strongly that the K61A mutation causes misfolding and rapid degradation of the mutant protein.

Functional consequences of the NAGK mutations

In crude cell extracts, the hydroxylamine-coupled colorimetric assay of NAGK activity[19] (an assay based on the conversion of NAG phosphate to NAG hydroxamate, which has limited linearity and sensitivity) revealed for the K8R, R66K and N158K mutants substantial although diminished activity (relative to wild-type), but not for the D162E mutant, despite the presence of a large amount of this mutant enzyme in the corresponding cell extracts (data not shown). The same assay using purified enzymes failed to demonstrate substantial activity with the D162E mutant (detection limit, 1% of wild-type) and confirmed that the N158Q, K8R and R66K mutants are active, although $<15\%$ as active as the wild-type enzyme (data not shown). As expected, all the active enzymes produced equivalent amounts (within experimental error) of NAG phosphate and ADP. (ADP was measured in the hydroxylamine-containing mixture by taking acid-quenched samples to a pyruvate kinase–lactate dehydrogenase ADP assay.[20]) A more sensitive continuous assay that monitors at 340 nm the NAG-dependent ADP production and that is well suited for kinetic constant determinations was used in further enzyme activity measurements (see Materials and Methods). With this assay, the D162E mutant was shown to exhibit approximately 0.1% of the activity of the wild-type enzyme. The possibility that this low activity is due to contamination by the endogenous NAGK of the host *E. coli* cells was excluded by showing, using SDS-PAGE (followed by elution of Coomassie brilliant blue-stained material with 25% (v/v) pyridine and colorimetric quantification at 605 nm,[21]) that the purified D162E mutant enzyme exhibited only 6% contamination by other *E. coli* proteins, and by demonstrating in appropriate assays that this amount of contaminating protein could not account for the observed NAGK activity. In fact, no endogenous NAGK activity was detected in the protein fraction isolated when the entire NAGK purification procedure was applied to extracts of control cells transformed with an empty pET plasmid carrying no insert, possibly because the expression of the chromosomal *E. coli* arginine-synthesizing genes is repressed by the rich LB medium used for growing the cells.[22]

Plots of NAGK activity as a function of the concentration of NAG (Figure 4, top) or ATP (Figure 4, bottom) are hyperbolic, although for ATP there is substrate inhibition with the wild-type enzyme. Apparent K_m values for wild-type NAGK (Table 1) are in the range of 0.2–0.3 mM for both NAG and ATP, with little change in the K_m^{ATP} when the concentration of NAG was decreased from 0.1 M to 0.01 M. At the substrate concentration ranges tested, the activity of the mutant enzymes was considerably smaller than for the wild-type enzyme, decreasing in the order N158Q > R66K > K8R > D162E.

For the mutants N158Q and R66K, enzyme activity grew nearly linearly with concentration of NAG, rendering the accurate estimation of the kinetic constants for NAG in these mutants very difficult to achieve. Our best estimates are (Table 1), for N158Q, an unchanged $V^{[NAG]=\infty}$ and a 3000-fold increase in the K_m^{NAG}, and for R66K either a 4500-fold increase in the K_m^{NAG} and a four-fold decrease in $V^{[NAG]=\infty}$ or conservation of $V^{[NAG]=\infty}$ and a nearly 20000-fold increase in the K_m^{NAG}. When the varied substrate was ATP, more modest increases (20–30-fold) in the K_m^{ATP} and low

Table 1. Influence of NAGK mutations on the kinetic constants of this enzyme

	Varied substrate									
	N-Acetyl-L-glutamate				MgATP					
	K_m^{NAG}		$V^{[NAG]=\infty}$		[NAG]	K_m^{ATP}		$V^{[ATP]=\infty}$		K_i^{ATP}
Mutant	(mM)	(K_m^{mut}/K_m^{wt})	(units/mg)	(V^{mut}/V^{wt})	(mM)	(mM)	(K_m^{mut}/K_m^{wt})	(units/mg)	(V^{mut}/V^{wt})	(mM)
WT	0.20	1	63.0	1	100	0.29	1	87.8	1	30.4
					10	0.31	1	74.8	1	122[a]
R66K	898[b] (4000)[c]	4490 (20,000)[c]	17.2[b] (63)[c]	0.27 (1)[c]	100	5.2	18	2.0	0.023	–
N158Q	600	3000	53.6	0.85	100	7.9	27	15.8	0.18	41.5[b]
K8R	2.3	11.5	1.19	0.019	100	3.3	11	1.6	0.018	–
D162E	0.37	1.9	0.073	0.0012	10	0.46	1.5[d]	0.073	0.0010[d]	–

The values of the kinetic constants are those for the curves shown in Figure 4. The K_i^{ATP} values are those for inhibition by excess nucleotide (see Materials and Methods and Figure 4). Standard errors did not exceed 20% of the values given for the constants, except when K_m^{NAG} or K_i^{ATP} values were very high; in these cases standard errors were approximately ([a]) 60%, ([b]) 40% of the value of the constant.

[c] Alternative estimate fitting also the data.

[d] Value relative to the results obtained at 10 mM NAG with the wild-type enzyme.

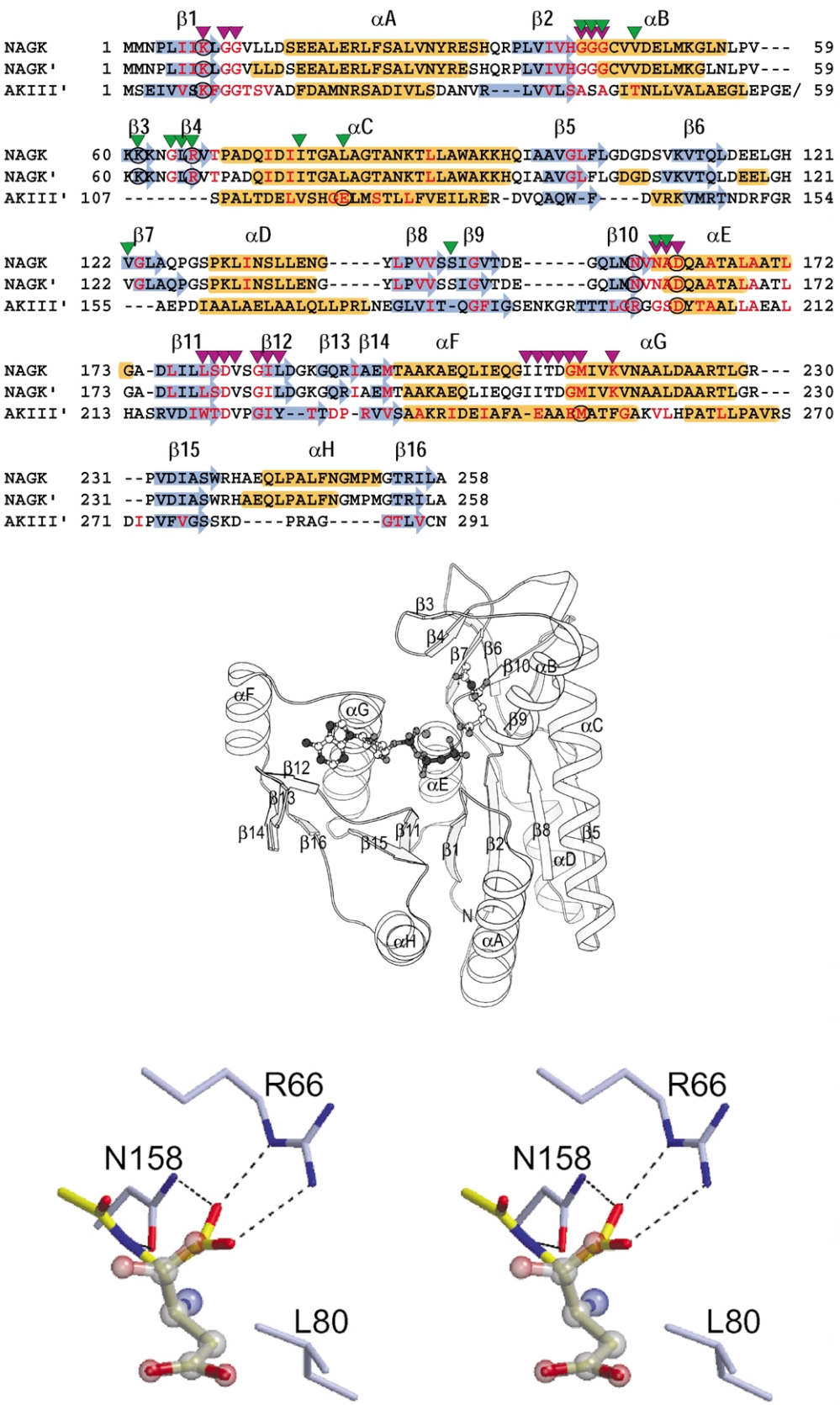


Figure 5. Top panel, experimental (NAGK) or predicted (NAGK′ and AKIII′) secondary structure-guided alignment of the amino acid sequences of *E. coli* acetylglutamate kinase and of the amino acid kinase domain of *E. coli* aspartokinase III. The secondary structure elements are superimposed on the sequences by shadowing in blue (β strands) or orange (α helices), and are designated above the alignment according to the observed NAGK 3-D structure (shown in ribbon representation in the middle panel). Residues conserved or replaced conservatively in at least 27 of 28 NAGK sequences or in at least 24 of 25 AK sequences are colored red. Residues mutated here are encircled. Residues that in

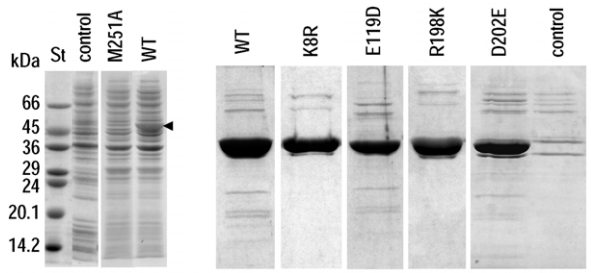


Figure 6. Expression and purification of wild-type (WT) and mutant forms of NAGK, monitored by SDS-PAGE (12% (w/v) polyacrylamide gels). Left-hand panel: whole-cell extracts of *E. coli* BL21(DE3) transformed with either the empty parental plasmid (control), or with the plasmids pAKM251A or pAKIIIwt, which encode the M251A mutant or the wild-type (WT) enzyme forms. Approximately equivalent numbers of *E. coli* cells collected after three hours induction with IPTG were boiled for five minutes in electrophoresis sample solution containing SDS, and were subjected to SDS-PAGE and staining with Coomassie brilliant blue.[40] A large AKIII band (indicated with an arrowhead) is observed in the WT track but not in the M251A track. St, molecular mass markers (Dalton Mark VII-L, Sigma). Right-hand panel: analysis of the purified wild-type and mutant enzyme forms, as indicated. Control denotes material purified from extracts of cells transformed with the parental empty pET plasmid. All lanes contain equivalent amounts of protein, except the control lane, in which the amount of protein is only 10%, to try to approximate the amount of host cell contamination in the purified enzyme samples.

$V^{[ATP]=\infty}$ values were observed with these two mutants. The low $V^{[ATP]=\infty}$ values are due to the use of a sub-optimal fixed NAG concentration (0.1 M) relative to the K_m^{NAG}. Overall, the N158Q and R66K mutations appear to have a strong selective negative effect on the apparent affinity of the enzyme for NAG, judged from the increases in K_m^{NAG} values.

With the K8R and D162E mutations, the dominant effects are 50-fold and 1000-fold decreases, respectively, on the apparent V_{max} values, with only 11-fold (K8R) or <2-fold (D162E) increases in the apparent K_m values for the substrates. Thus, despite the mildness of these mutations, they decrease strongly the catalytic efficiency of the enzyme.

Selection of aspartokinase III mutations and cloning and isolation of mutant forms of the enzyme

To identify in aspartokinase III (AKIII) residues potentially playing roles similar to those of the residues mutated in NAGK, we have aligned (Figure 5, top) the sequence of the amino acid kinase domain of AKIII (residues 1–291) with that of *E. coli* NAGK, on the basis of a previous alignment[4] and with assistance of secondary structure prediction for AKIII (see the legend to Figure 5), so that the predicted elements were aligned with the experimentally determined structure of NAGK. Control prediction of the secondary structure of NAGK was in excellent agreement with the experimentally determined structure (Figure 5). The alignment obtained in this way exhibits 18% identity and 35% identity + similarity between the AKIII and NAGK sequences. From this alignment, K8 and D202 of AKIII, two invariant residues, correspond to K8 and D162 of NAGK and, as these NAGK residues, they have been mutated in AKIII to arginine (K8R) and glutamate (D202E), respectively.

The residues corresponding in AKIII to R66 and N158 of NAGK have been identified tentatively by assuming that NAG and aspartate bind similarly in NAGK and AKIII, respectively, as shown in Figure 5, lower panel. Thus, residues corresponding to L80 and N158 of NAGK would be located properly for interaction with the α-NH_3^+ and α-COO^- groups of aspartate (Figure 5, lower panel). In AKIII, E119 and R198 are invariant residues, they have the appropriately charged side-groups for interaction with the α-NH_3^+ and α-COO^- groups, and align, respectively, with L80 and N158 of NAGK (Figure 5, top). Therefore, we have tested the effects of introducing into AKIII the mutations E119D and R198K. A further modification introduced into AKIII affects the conserved residue M251, which has been replaced by alanine (M251A) in an attempt to probe the involvement of M251 in the binding of the adenine ring of ATP, since the corresponding residue in the aligned NAGK sequence is a conserved methionine residue (M214) that has been shown by the NAGK structure[4] to be the residue on whose side-chain the adenine ring lies flat (Figure 1, lower panel).

PCR cloning into a pET plasmid of the *E. coli* gene for AKIII, and the transformation with the resulting plasmid, pAKIIIwt, of BL21(DE3) cells, allowed, upon induction with IPTG, to overexpress the wild-type enzyme (Figure 6, left), which was purified in a matter of hours by a method involving sonication, precipitation in ammonium sulphate and FPLC ion-exchange chromatography. The resulting enzyme is essentially homogeneous, as demonstrated by SDS-PAGE (Figure 6, right), and exhibits good enzyme activity. All mutant

NAGK decrease their accessibility with the binding of MgAMPPNP or NAG[4] are marked with triangles colored magenta or green, respectively. Bottom panel, stereo view of NAG bound to *E. coli* NAGK, showing (broken lines) the interactions with some NAGK residues (in blue). Superimposed on the NAG, in semi-transparent representation, aspartate is shown in the conformation in which we propose it to be bound to AKIII. The N and O atoms of NAG or aspartate are shown as red and blue spheres, respectively.

forms of AKIII except the M251A mutant were overexpressed in soluble form and were purified in the same way to a similarly high level of purity (Figure 6, right). However, the M251A mutation resulted in the complete loss of AKIII expression (Figure 6, left), possibly because of misfolding and rapid degradation of the mutant enzyme.

Functional consequences of the different AKIII mutations

The well recognized lack of linearity of the colorimetric hydroxylamine assay of AK activity[23] rendered this assay rather unsuitable for kinetic determinations and, therefore, we have assayed AKIII activity by monitoring spectrophotometrically ADP production with pyruvate kinase and lactate dehydrogenase, as described.[24] Nevertheless, hydroxylamine assays on the various purified enzyme forms revealed activity, although the activity was lower with all the mutants than with the wild-type enzyme, particularly in the case of the K8R and D202E mutants. A control preparation generated by carrying out the entire purification procedure on an extract of cells transformed with the empty parental pET plasmid yielded no substantial aspartate-dependent enzyme activity, either with the hydroxylamine assay or with the ADP-monitoring assay.

The dependency of the activity on the concentration of aspartate (Figure 7, top) conforms in all cases to hyperbolic kinetics, although the very low apparent affinity for aspartate of the E119D mutant causes the activity of this mutant to increase virtually linearly with the concentration of aspartate over the entire concentration range tested. In the case of the dependency of the activity on the concentration of ATP (Figure 7, lower panel) the kinetic pattern is also hyperbolic, although the wild-type enzyme, R198K and E119D mutants, exhibited inhibition at high concentrations of ATP. For variable concentrations of both NAG and ATP the mutations K8R and D202E reduced the activity at infinite substrate drastically, to approximately 2–3% of wild-type (Table 2), as expected for

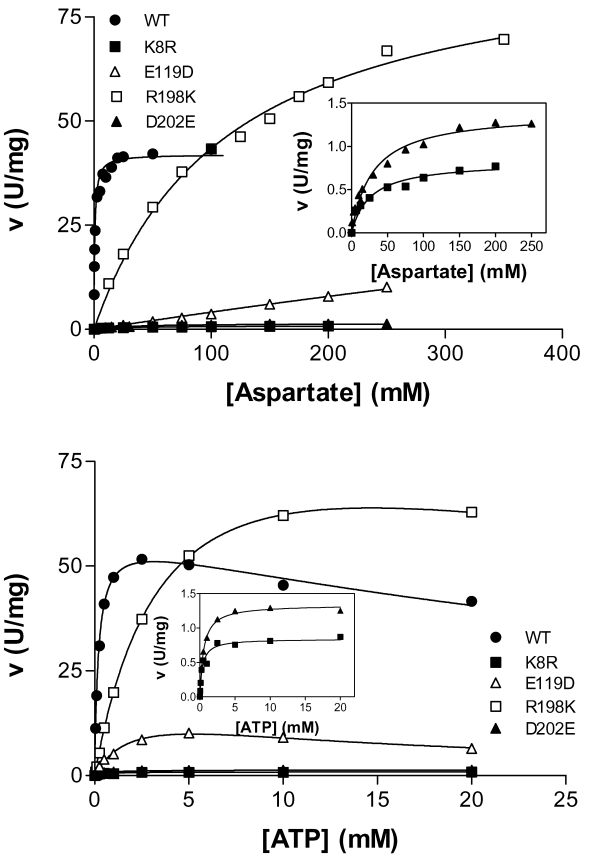


Figure 7. Enzyme activity of wild-type and mutant forms of AKIII as a function of the concentrations of (top panel) L-aspartate or (bottom panel) ATP. The larger graphs plot the results for all the enzyme forms on the same scale, whereas the insets zoom on the forms with the lower activities. For wild-type AKIII and for the R198K and E119D mutants, the lines fit to the data for variable ATP correspond to hyperbolic kinetics combined with inhibition by excess substrate, using the kinetic constants given in Table 2. In all other cases the lines are hyperbolae defined by the kinetic constants given in Table 2 for each enzyme form.

Table 2. Influence of AKIII mutations on the kinetic constants of the enzyme

	Varied substrate								
	Aspartate				ATP				
	K_m^{Asp}		$V^{[Asp]=\infty}$		K_m^{ATP}		$V^{[ATP]=\infty}$		K_i^{ATP}
Mutant	(mM)	(K_m^{mut}/K_m^{wt})	(units/mg)	(V^{mut}/V^{wt})	(mM)	(K_m^{mut}/K_m^{wt})	(units/mg)	(V^{mut}/V^{wt})	(mM)
WT	0.69	1	42	1	0.20	1	58	1	49
E119D	2152	3119	93[a]	2.2[a]	2.3	12	19	0.33	12
R198K	114	165	93	2.2	3.9	20	98	1.7	55
K8R	26	38	0.83	0.020	0.36	1.8	0.84	0.015	–
D202E	29	42	1.4	0.033	0.52	2.6	1.34	0.023	–

The values of the kinetic constants are those for the curves shown in Figure 7. The K_i^{ATP} values are those for inhibition by excess nucleotide (see Materials and Methods and Figure 7). Standard errors did not exceed 20% of the values given for the constants.
[a] The $V^{[Asp]=\infty}$ has been assumed to be 93 units/mg, the value estimated for the R198K mutant.

residues involved in catalysis of phosphoryl group transfer, as is the case with the corresponding residues in the alignment with NAGK, K8 and D162. Nevertheless, the two mutations increased the K_m^{Asp} value by approximately 40-fold, and therefore they might contribute, possibly not crucially, to the binding of this substrate. A more important effect on the K_m^{Asp}, amounting to a nearly 200-fold increase, was caused by the mutation R198K, affecting the residue proposed above to be involved on the binding of the α-COO^- group of aspartate. This effect was highly selective for aspartate, since the R198K mutation increased much less (approximately 20-fold) the K_m^{ATP}, supporting the selective implication of R198 on aspartate binding. A surprising finding with this mutant was the observation that the velocity at infinite aspartate or ATP was approximately two-fold higher than for the wild-type enzyme.

Finally, the E119D mutation had a drastic, selective effect on K_m^{Asp}, which was increased very much in this mutant. In fact, the nearly linear dependency of the activity on the concentration of aspartate even up to 0.25 M of this substrate rendered impossible the accurate quantitative determination of the kinetic constants for aspartate. It was impossible to fit a hyperbola to these data if $V^{[Asp]=\infty}$ was assumed to be the same or lower than for the wild-type enzyme, but the fitting became possible if $V^{[Asp]=\infty}$ was assumed to be the same as for the R198K mutant, approximately twice that for the wild-type enzyme. In this fitting, K_m^{Asp} is approximately 3000-fold higher than for wild-type. This large increase in K_m^{Asp}, and the selectivity of this effect for aspartate (K_m^{ATP} was increased by only 12-fold) supports the specific involvement of E119 in aspartate binding, as expected if it interacts with the α-NH_3^+ group of aspartate, as proposed here. With this mutant, the velocity at infinite ATP is reduced to approximately one-third of that of wild-type, but this is due to the use of a low (relative to K_m^{Asp}) fixed concentration of aspartate.

An additional point examined here is the influence of the mutations on the feed-back inhibition of AKIII by lysine. As shown in Figure 8, lysine inhibited the wild-type and the E119D and R198K mutants similarly. In agreement with earlier findings,[25,26] the inhibition was sigmoidal, yielding a Hill number N close to 2. $I_{0.5}$ values for the wild-type and E119D and R198K mutants were of the order of 0.3 mM lysine, whereas with the K8R and D202E mutants, in which the catalytic efficiency of the enzyme is decreased, the $I_{0.5}$ values were increased 10–25-fold, although the degree of inhibition attained at infinite lysine, the sigmoidal character of the plot, and the Hill value of approximately 2, remained unchanged (Figure 8).

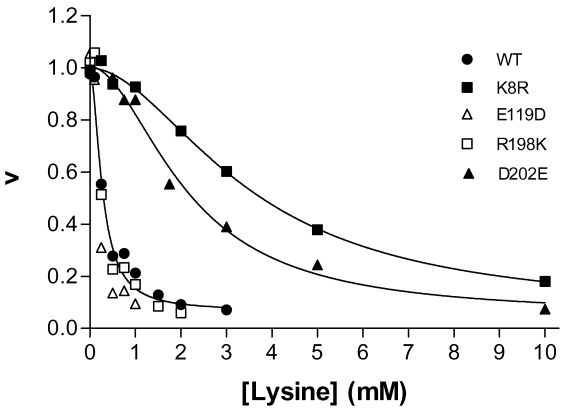


Figure 8. Influence of the concentration of lysine on the activity of wild-type (WT) or mutant forms of AKIII. Velocities (v) are given as fractions of the velocity in the absence of lysine. The lines drawn are sigmoids given by the equation:

$$v = 1 - ([\mathrm{Lys}]^N/(K + [\mathrm{Lys}]^N))$$

for an N value of 1.9 and K values of 0.1 (WT), 4.2 (D202E) and 10.5 (K8R) mM, yielding $I_{0.5}$ values of 0.27 (WT), 2.2 (D202E) and 3.5 (K8R) mM. Results for the forms with the mutations E119D and R198K approximate those for the wild-type enzyme, and correspondingly no sigmoid has been drawn for these two mutants.

Modelling the 3-D structure of the amino acid kinase domain of AKIII

The alignment of the entire length of the NAGK polypeptide with residues 1–291 of AKIII (with the exception of residues 60–106, see below), the prediction in AKIII of a number of secondary structure elements that conform well with the disposition and number of most such elements found in the NAGK structure, and the obvious similar roles in catalysis and in the binding of the phosphorylatable substrate of the residues that have been mutated here in parallel in the two enzymes, support the feasibility of using the NAGK structure for deriving a reliable 3-D model of the amino acid kinase domain of AKIII. Residues 60–106 of the AKIII sequence are excluded from alignment because many aspartokinases lack this sequence stretch,[16] which is not represented in NAGK. The alignment (Figure 5, top) indicates that among the secondary structure elements of NAGK, only the last α-helix (αH) and the β3-β4 hairpin of NAGK may not be present in AKIII.

We have used the alignment shown in Figure 5 and the crystal structure of NAGK complexed with MgAMPPNP and acetylglutamate[4] to model the 3-D structure of the amino acid kinase domain of AKIII (residues 1–291, excepting residues 56–106) with the MODELLER[27] program (see Materials and Methods). A highly similar model produced when the structure of the *Pyrococcus furiosus* carbamate kinase[3] complexed with MgADP was used as an additional template is not discussed here. The model has been optimized by refinement of loop regions and manual tracing of side-chains in the active center to maximize ligand contacts. No severe disallowed atomic contact was detected with WHATIF,[28] and PROCHECK[29]

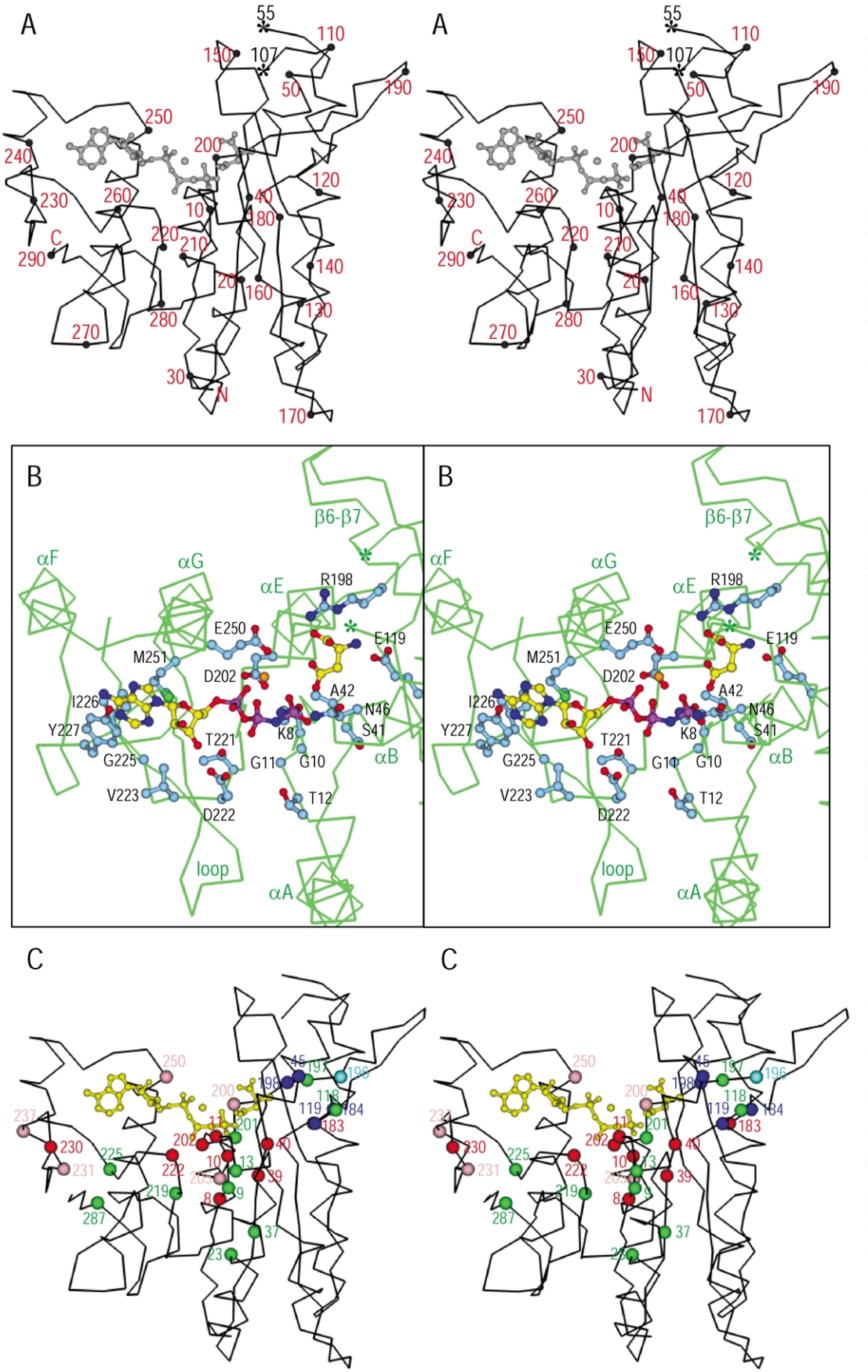


Figure 9. Stereoview of the modelled structure (C^{α} trace) of the AKIII subunit with bound MgAMPPNP and aspartate (A), detailing the proposed active center (B), and the localization of prior[16] and present mutations introduced

revealed essentially good stereochemistry, with 87% and 13% of the residues in the most favored and additional allowed regions, respectively, and with no residue in the generously allowed or disallowed regions of the Ramachandran plot. Energy profiles obtained with ProsaII[30] have revealed a lower energy value for the C-terminal lobe (residues 214–291) than for the N-terminal lobe (residues 1–213), as is observed with the NAGK subunit (in NAGK, the N-lobe and C-lobes host the sites for NAG and for the ADP moiety of ATP, respectively,[4]) indicating that the C-terminal lobe is, by itself, more stable than the lobe that may be expected by similarity with NAGK to be involved in dimer formation. Superposition of the modelled AKIII domain with NAGK gives a root-mean-square deviation (rmsd) value of 1.54 Å for 209 C^{α} atoms of the NAGK polypeptide, a value that is comparable to that obtained when the backbones of the *P. furiosus* carbamate kinase and NAGK are superimposed (rmsd 2.17 Å for 141 C^{α} atoms[4]). Overall, these results suggest strongly that the model is reasonable, possibly representing a good approximation to the genuine structure of the amino acid kinase domain of the AKIII subunit.

According to the model (Figure 9A), the amino acid kinase domain of AKIII exhibits the central β-sheet with all its eight strands having similar lengths as in NAGK, and being sandwiched between the front and back layers of α-helices, as in NAGK and carbamate kinase,[2–4] except for the lack of the more C-terminal α helix, αH, which is replaced in AKIII by a short sequence with no assigned prediction of secondary structure, and which we have modelled as a loop interposed between the two final β strands. As in NAGK and carbamate kinase, helix C is the longest and, together with helix B, protrudes out from the C-edge of the main sheet.

In addition to this basic scaffold, the loops and hairpins that in NAGK form the sites for the substrates are recognized in the aligned sequence of AKIII, allowing their modelling, except in the case of the β3-β4 hairpin of NAGK, which has no recognizable counterpart in the AKIII sequence and has not been modelled. The loop that hosts the β6-β7 hairpin is seven residues shorter in AKIII than in NAGK and has been modelled preserving the position and β-strands interactions of the β hairpin. In NAGK, this loop contributes to the wall of the NAG-binding pocket in the region that is farther away from the NAG γ-carboxylate group[4] and, therefore, this loop might be expected to be shorter in AK than in NAGK, given the smaller size of the pocket needed to accommodate aspartate. The loop that hosts the β9-β10 hairpin is in AKIII five residues longer than in NAGK, due to the insertion of a short sequence between the two β-strands. Neither this enlargement nor the shortening of the other loop have important structural consequences for the overall fold and can be accommodated easily, since the changes occur in the surface of the subunit. The loop that hosts the β12-β13 hairpin and that in NAGK contours the adenosine moiety of the bound nucleotide (Figure 1, bottom panel) is well conserved in AKIII and can be modelled easily. The identification of the loop linking helices F and G, which in NAGK contours on the other side the site for the adenosine moiety of the nucleotide, is helped in AKIII by the alignment in helix G of the conserved methionine that in NAGK provides a hydrophobic surface on which the adenine ring lies.

Consequences of the deletion of the C-terminal region of aspartokinase III

In addition to the amino acid kinase domain, AKIII has a C-terminal region of 158 amino acid residues that is believed to be involved in lysine inhibition[31] and where sequence comparisons afford recognition of two ACT domains (see Introduction and Figure 2). We have examined the consequences of truncating the AKIII polypeptide after the end of the amino acid kinase domain, by replacing codon 298 of the cloned gene by a stop codon (L298X). The truncated AKIII enzyme was overexpressed in soluble form and in good yield

at conserved sites in the amino acid kinase domain of aspartokinases (C). A, The residue at every tenth position is marked with a sphere and is numbered in red. Residues 56–106 of AKIII have not been modelled (highlighted with asterisks and black residue numbers at the chain breaks). MgAMPPNP and aspartate are colored gray, and are shown in ball-and-stick representation. In B the C^{α} trace is given, important structural elements and side-chains are identified, and these side-chains (in light blue) and the substrates (in yellow) are illustrated in plausible conformations, in-ball-and stick representation. N, O, Mg and P atoms are colored deep blue, red, orange and magenta, respectively. The asterisks again mark breaks in the peptidic chain model. C, The stereo view representation includes the bound substrates (in yellow, as ball and stick models) and the mutated residues are numbered and marked with spheres, with the color-coding of the spheres reflecting the consequences of the mutations: green, little or no effect; red, important (at least eightfold) and pink, substantial (between three- and eightfold) decrease in k_{cat}; deep blue, large (>20-fold) and light blue, substantial (eightfold) selective increase in the K_m^{Asp}. When several mutations of a certain residue were tested, the result shown is that for the mutation having the more important effect. The changes studied directly in *E. coli* AKIII have been given preference over mutations in equivalent residues in *T. flavus* aspartokinase.[16] The equivalence of residue numbering in *T. flavus* AK with numbering in *E. coli* AKIII is the following for the residues that were mutated (*T. flavus*/*E. coli*):[16] 7/8, 8/9, 9/10, 10/11, 12/13, 23/23, 39/37, 41/39, 42/40, 47/45, 73/118, 74/119, 135/183, 136/184, 148/196, 149/197, 150/198, 152/200, 153/201, 154/202, 157/205, 171/219, 174/222, 177/225, 182/230, 183/231, 189/237, 202/250 and 238/287.

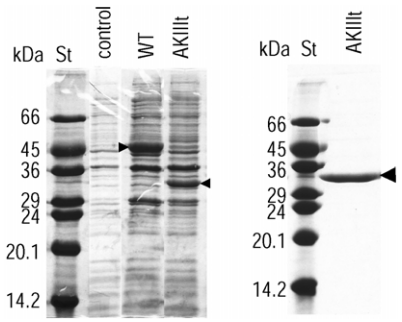


Figure 10. Expression and purification of the amino acid kinase domain of AKIII (AKIIIt), prepared by introducing a stop codon at position 298 of the cloned AKIII gene. The results of SDS-PAGE[42] (15% (w/v) polyacrylamide gels) revealed by staining with Coomassie brilliant blue are shown. Left-hand panel: whole-cell extracts of *E. coli* BL21(DE3) transformed with either the empty parental plasmid (control), or with the plasmids pAKIIIwt or pAKIIIt, encoding, respectively, the wild-type (WT) and the truncated (AKIIIt) forms of AKIII. Approximately equivalent numbers of *E. coli* cells collected after three hours induction with IPTG were boiled for five minutes in electrophoresis sample solution containing SDS, and were subjected to SDS-PAGE and staining with Coomassie brilliant blue.[42] The bands corresponding to the wild-type and the truncated polypeptide are indicated with arrowheads. Right-hand panel: isolated truncated enzyme form after purification. In both panels, St denotes molecular mass markers (Dalton Mark VII-L, Sigma) having the indicated molecular masses (in kDa).

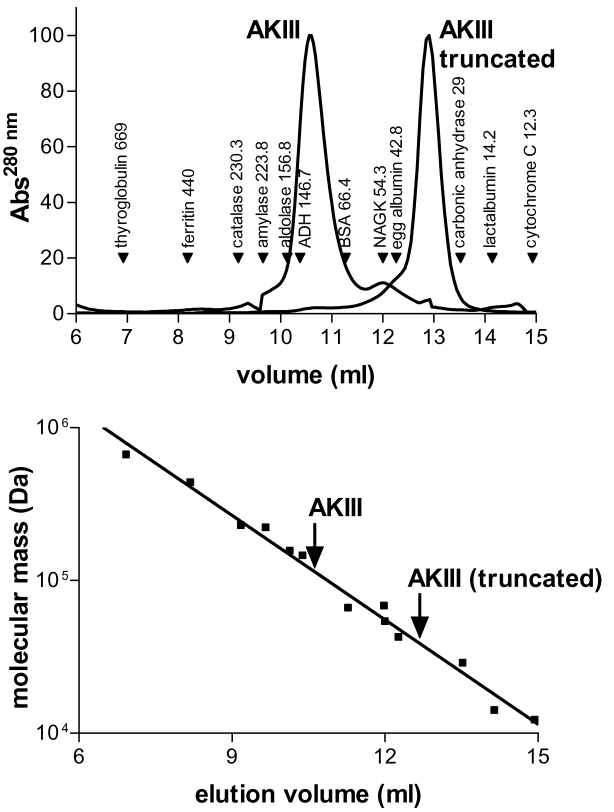


Figure 11. Estimation of the native molecular mass of the wild-type and truncated AKIII forms by size-exclusion chromatography. Top: elution profiles of the two enzyme forms (monitored at 280 nm). The arrowheads signal the elution volume of the indicated protein standards, indicating the molecular mass of the protein (in kDa), following its name. Bottom: calibration curve used to determine the mass of the wild-type and truncated AKIII, with the elution position of these indicated by the arrows.

in *E. coli*, and was purified to homogeneity (Figure 10) following the same purification procedure as that used for the wild-type enzyme (although its elution from the HiTrap Q column occurred at a lower concentration of salt than for the wild-type enzyme; see Materials and Methods). Hydroxylamine-based and ADP release-based aspartokinase activity assays revealed that the truncated enzyme, either in the crude initial cell extracts or after purification, was inactive even at relatively high concentrations of the substrates (16 mM ATP and 45 mM aspartate). The lack of activity was not due to wrong assignment of the end of the amino acid kinase domain, since longer truncated enzyme forms prepared by introducing stop codons at positions 336 or 387 were inactive (data not shown).

We have proposed for the amino acid kinase enzymes carbamate kinase[2,3] and acetylglutamate kinase[4] that dimer formation is important, since the intersubunit interface provides a scaffold on which to anchor the binding sites and the catalytic machinery. Therefore the isolated monomers of these enzymes would be expected to be inactive. We reasoned that dissociation of AKIII could result in inactivation, and therefore we investigated by size-exclusion chromatography the oligomeric state of the truncated enzyme. Whereas wild-type AKIII was eluted at a position corresponding to a molecular mass of 110 kDa (Figure 11), and thus was dimeric (expected molecular mass of the dimer, 97 kDa), the elution position of the truncated mutant corresponded to the monomer (experimental molecular mass, 39 kDa; sequence-deduced molecular mass of the truncated monomer, 32 kDa). These results reveal that the C-terminal region of AKIII having no homologous counterpart in NAGK is crucially involved in intersubunit contacts, and thus that AKIII may differ from NAGK or carbamate kinase in the surfaces involved in dimer formation.

Discussion

NAGK mutations confirm functional predictions based on the NAGK structure

The main consequences of the mutations tested are highly consistent with the structure-based postulated roles of the mutated residues. The increases in K_m^{NAG} values caused by the N158Q and R66K mutations suggest strongly that these

mutations hamper NAG binding, in accordance with the proposal[4] that both the amide group of N158 and the guanidinium group of R66 make two hydrogen bonds with the acetamido group and α-carboxylate group of NAG. The decrease in the catalytic efficiency of NAGK caused by the K8R and D162E mutations highlights the exquisite sensitivity of the reaction to the exact localization of the positive charge on residue 8, used in the abstraction of negative charge from the transition state, and the crucial role of the interactions of the negatively charged side-chain of D162 with the ζ-amino groups of Lys8 and Lys217,[4] stressing the importance of D162 as an organizer of the catalytic site.[5] However, it may appear surprising that the K61A mutation should result in lack of NAGK accumulation possibly because of misfolding and rapid degradation, since Lys61 is outside the enzyme core (Figure 1, bottom). However, Lys61 is a part of the β3-β4 hairpin that connects helices B and C, which are involved in dimer formation,[4] and a change in this residue might contribute to misplace these helices, hampering dimer formation and perhaps triggering degradation. In addition, by forming an ion pair with Asp212,[5] Lys61 may bridge the loops of the N and C-terminal lobes of the enzyme, possibly preventing disorder and potential proteolytic cleavage of these loops.

Significance of the AKIII mutations

Although with some quantitative differences, the overall picture for AKIII parallels that obtained for NAGK: the AKIII mutations K8R and D202E, as the K8R and D162E mutations of NAGK, affect mainly the catalysis of the reaction, whereas the mutations E119D and R198K of AKIII, as the NAGK mutations R66K and N158Q, appear to affect primordially the binding of the phosphorylatable substrate, given the increase in K_m value. This supports a parallelism between the active centers of the two enzymes, including the localization of the residues that in AKIII conform the site for aspartate, as might be expected from the similarity of the reactions catalyzed (Figure 2) and from the identification of these enzymes as members of the same protein family (the amino acid kinase family). Additionally, the similarity of the effects of the mutations of the two enzymes provides further support for the correctness of the sequence alignment (at least in the regions corresponding to the residues that have been mutated) on which modelling has been based.

The observation with the R198K (and possibly also with the E119D) mutant of increased velocity at infinite ATP and aspartate compared with wild-type AKIII indicates that the intrinsic catalytic efficiency of the enzyme can be increased. Comparison of the specifity constants (V_{max}/K_m; data not shown) indicates that evolution appears to have optimized the specificity constant rather than V_{max}, trading velocity for specificity, as shown by the nearly two orders of magnitude higher specificity constant for the wild-type than for the more active R198K mutant.

It is surprising that the mutations of amino acid kinase domain residues, K8R and D202E, increase 10–25-fold the $I_{0.5}$ value for lysine, since previous studies have shown that mutations in the C-terminal region of AKIII and, more specifically, within the canonical ACT domain, abolish the inhibition by lysine.[31] The lysine-binding site might be formed in AKIII at an interface between the ACT and amino acid kinase domains, and Arg8 and Asp202 might be at or near this interface, although this would appear unlikely if Arg8 and Asp202 are localized in the same way as the homologous residues of NAGK. Alternatively, AKIII might exist in two conformations and lysine might bind to only one of these, which would be inactive, whereas the K8R and D202E mutations would favor the other conformation, explaining why these mutations decrease the apparent affinity for lysine. Perhaps the low catalytic efficiency of the enzyme carrying any of these mutations is associated with alteration of the partitioning between the different enzyme forms in the catalytic cycle, with predominance of an enzyme form charged with both substrates and having decreased affinity for lysine.

Substrate binding and catalysis in AKIII

The modelled loops of the C-lobe delineate plausibly the site for the adenosine moiety of ATP similar to the corresponding site in NAGK. After all, the nucleotide binds nearly identically in NAGK[4] and carbamate kinase,[3] despite the existence of a low level of sequence identity between these two enzymes. If this represents a characteristic way of nucleotide binding that is common to all the amino acid kinase enzymes, ATP would bind in AKIII (Figure 9A and B) extended along the C-edge of the main β sheet with the adenosine group pointing away and the γ-phosphate group pointing towards the N-lobe, with the adenine ring sitting flat over the side-chain of M251 (M214 in NAGK) and being encircled by the β11-β12 (NAGK secondary structure denomination) and αF-αG connections. The impact of the mutations affecting D202 and K8 indicate that these residues, as in NAGK, are involved in the catalysis of phosphoryl group transfer and therefore are near the ATP γ-phosphate group, at the boundary between the two lobes of the amino acid kinase domain. The modelled β1-αA and β2-αB connections share with NAGK and with carbamate kinase the presence of small residues. In NAGK[4,5] (and apparently also in carbamate kinase[2,3]) these small residues allow unhampered hydrogen bonding of the main chain to the nucleotide γ-phosphate group and the attacking carboxylate group, and this may be their role in AKIII. Our present success in identifying residues whose mutation drastically affects the K_m^{Asp} using as working hypothesis that

aspartate binds in AKIII as NAG in NAGK supports the correctness of this hypothesis. If this were the case, the β-carboxylate group of the substrate aspartate would sit in AKIII between the β2-αB and the β10-αE connections, at the opening of the aspartate site, and the site would be formed between the β10-αE connection, the initial turns of αB, the second turn of αC and β7 (Figure 9A and B). The results of our mutagenesis studies render plausible the proposal that the α-NH_3^+ of aspartate interacts with the side-chain of E119 (which would emerge from αC), and that the α-COO^- interacts with the side-chain of R198 (which would emerge from β10) (Figure 9B). It is unfortunate that residues 56–106 cannot be modelled, because this sequence may cap the site for aspartate group, closing the pocket from above and leaving only one narrow opening at the site of binding of the β-COO^-, as observed for NAG in NAGK.[4]

Catalysis by NAGK involves[5] negative charge abstraction from the transition state by the positive charges on Lys8 and Lys217 and on the N-ends of helices B and E, and by donor hydrogen bonding by the β1-αA and β2-αB connections. In the modelled AKIII structure (Figure 9B) the same elements can be recognized, except the residue corresponding to Lys217 of NAGK, which is missing. Perhaps the corresponding positive charge is provided by the sequence (residues 60–106) that has been omitted from the alignment, given the proximity of this sequence to the proposed active center and given the presence in this sequence of several positively charged residues.[16] Even in the aspartokinases that lack this extra sequence there is constantly one arginine residue that might play that role, localized in the linker between the regions corresponding to the modelled helices B and C.[16]

Kobashi *et al.*[16] studied, using a relatively insensitive enzyme activity assay utilizing low aspartate concentrations, the effects of mutations (mainly to alanine) affecting the invariant or the highly conserved residues of the region of the aspartokinase from *Thermus flavus* (an $\alpha_4\beta_4$ oligomer) corresponding to the amino acid kinase domain.[16] As our present results, these results highlighted the involvement of this domain in the catalysis of the reaction. The caveats in the assay used possibly account for the negligible activity or low k_{cat} reported[16] for mutations affecting E74 and R150, the residues corresponding to E119 and R198 of *E. coli* AKIII, two residues whose mutations are shown by our present studies to increase very importantly the K_m^{Asp}. Nevertheless, the findings with *T. flavus* generally agree with the predictions of our model. All the highly conserved residues were mutated (colored spheres, Figure 9C; numbering for *E. coli* AKIII sequence; equivalence to *T. flavus* sequence given in the Figure legend) and, as in carbamate kinase[2,3] and NAGK,[4] these generally map within or near the active center proposed here. Mutation of residues that in our model do not interact with the bound substrates had little or no effect (Figure 9C, green spheres), whereas mutations at residues (colored red) proposed here to be involved in the active site abolish or greatly reduce the activity. Mutation at residues that in our model are close but involved less directly in the nucleotide site reduced significantly but not extremely the k_{cat} (colored pink). All the residues that when mutated increased drastically (deep blue) or substantially (light blue) the K_m^{Asp} map in

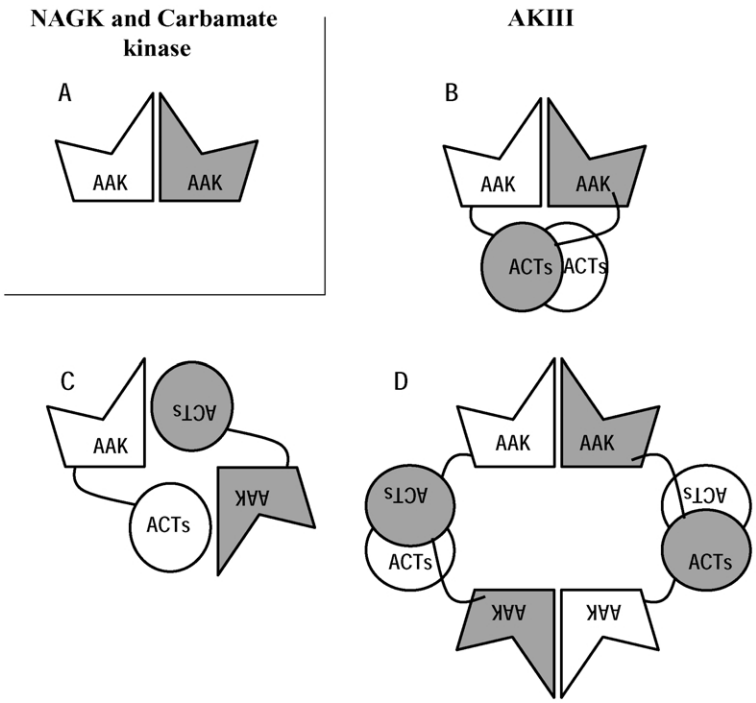


Figure 12. Schematic representations of the NAGK and carbamate kinase homodimer (A), and of potential organizations (B–D) of the amino acid kinase (AAK) and double ACT (ACTs) domain in the AKIII dimer and tetramer.

the site or very near the site predicted in the model for aspartate. Thus, these results, which were not taken into consideration when the model was built, provide strong additional validation for the present model.

AKIII quaternary structure

Since AKIII is a homodimer,[12] it would be conceivable that the amino acid kinase domains of the two enzyme subunits would associate as in the NAGK and carbamate kinase dimers (Figure 12A).[2–4] In fact, Verify3D[32] yielded scores of 0.21 and 0.27 for the isolated modelled subunit and for the modelled dimer, respectively (corresponding values for the NAGK structure, 0.54 and 0.56), suggesting the proper identification of the dimerization surface in each amino acid kinase domain. However, since the deletion of the C-terminal region containing the two ACT domains abolishes dimer formation, the C-terminal region appears to be involved centrally in inter-subunit interactions, as is customary for ACT domains,[17,33–35] and either the amino acid kinase domains of the two AKIII subunits do not associate mutually as in NAGK or carbamate kinase, or if they associate (Figure 12B) the association is too weak to allow stable dimer formation. The low stability would contrast with the remarkable stability of the carbamate kinase and NAGK dimers. A satisfactory explanation of the dissociating effect of the deletion of the C-terminal region would be provided by the AKIII architecture schematized in Figure 12C, in which the canonical interaction surface of the amino acid kinase domain of one subunit binds to the C-terminal region of the other subunit. Drawbacks of this potential architecture are its large deviation from the architectures of NAGK and carbamate kinase, and the difficulties that it poses for nucleation of more complex quaternary structures such as those observed with the aspartokinases of the oligomeric $\alpha_2\beta_2$ or $\alpha_4\beta_4$ types[13–16] or even that of AKIII under conditions of high concentrations of salt and lysine, when homotetramers can be formed.[12,36] In contrast, the architecture schematized in Figure 12B may easily explain tetramer formation by changing the interacting "sticky" double ACT domains, as schematized in Figure 12D. Examples of enzymes having ACT domains at the C terminus and conforming to the two structural models represented in Figure 12B and C are provided by the structures of threonine deaminase[33] and phenylalanine hydroxylase,[34] respectively. Indeed, the structure of a homotetrameric enzyme molecule having ACT domains and conforming to the model schematized in Figure 12D has been determined.[35] Further work with aspartokinases, including the determination of the 3-D structure of at least AKIII, is needed to clarify fully the architecture of these enzymes and their feed-back inhibition.

Materials and Methods

Site-directed mutagenesis of *E. coli* acetylglutamate kinase and aspartokinase III

Standard protocols were employed for electrotransformation of *E. coli*, and for isolation, digestion and ligation of DNA. To construct plasmid pAKIIIwt, carrying the *E. coli* gene *lysC*,[8] which encodes AKIII, the *lysC* gene was PCR-amplified with a high-fidelity DNA polymerase (Pwo, from Roche Diagnostics), using as template genomic DNA from *E. coli* BL21(DE3) (from Novagen), and the direct and reverse primers $^{5'}$CACGAGGTACAT ATGTCTGAAATTGTTGTCTCC$^{3'}$ and $^{5'}$CTTCCAGGCC ATGGAGTATTTACTCAAAC$^{3'}$, which introduce *Nde*I and *Nco*I restriction sites just upstream of the initiator ATG and downstream of the stop codon, respectively. The PCR product, digested with *Nde*I and *Nco*I, was ligated into the corresponding sites of the expression vector pET22b (Novagen) using bacteriophage T4 DNA ligase (USB), transforming *E. coli* DH5α cells (from Clontech) with the product of the ligation. The resulting pAKIIIwt plasmid was isolated and was shown by restriction analysis and DNA sequencing to encompass the full-length NAGK gene having the correct sequence.

Site-directed mutagenesis was performed with the QuickChange™ site-directed mutagenesis kit (from Stratagene), using as templates the pET-derived plasmids pNAGK24[18] (a pET-15b-derived plasmid carrying the *E. coli* gene for NAGK) and pAKIIIwt, respectively. When possible, the complementary plasmid pairs used in the mutagenesis experiments carried, in addition to the relevant mutation, a new restriction site to facilitate identification of the mutated clones. The oligonucleotide pairs used for mutating NAGK were: K8R mutation, $^{5'}$CCATTAATTATCCGTCTGGGCGGCGTAC TGCTGGATAGTG$^{3'}$ and $^{5'}$CCAGCAGTACGCCGCCCA GACGGATAATTAATGGATTC$^{3'}$; K61A mutation, $^{5'}$CTG CCGGTGAAAGCGAAAAACGGCCTGCGGGTG$^{3'}$ and $^{5'}$CGCAGGCCGTTTTTCGCTTTCACCGGCAGATTCA GC$^{3'}$; R66K mutation, $^{5'}$CTGCCGGTGAAAAAGAAAA ACGGTTTAAAGGTGACGCCTGCTGATCAG$^{3'}$ and $^{5'}$CAGCAGGCGTCACCTTTAAACCGTTTTTCTTTTTC ACCGGCAGATTCAG$^{3'}$; N158Q mutation, $^{5'}$GGGCAA CTGATGCAGGTCAACGCCGACCAGGCGGCAACG$^{3'}$ and $^{5'}$CGCCTGGTCGGCGTTGACCTGCATCAGTTGCC CTTCG$^{3'}$; D162E mutation, $^{5'}$GATGAACGTCAATGCT GAGCAGGCGGCAACGGCGC$^{3'}$ and $^{5'}$CCGTTGCCGC CTGCTCAGCATTGACGTTCATCAGTTGCC$^{3'}$. For mutating AKIII, the oligonucleotide pairs were, for the K8R mutation, $^{5'}$GAAATTGTTGTCTCCCGTTTTGGCG GTACCAGCGTAGC$^{3'}$ and $^{5'}$CGCTGGTACCGCCAAA ACGGGAGACAACAATTTCAGAC$^{3'}$; E119D mutation, $^{5'}$CTGGTCAGCCATGGCGATCTGATGTCGACCCTGC TG$^{3'}$ and $^{5'}$CAGGGTCGACATCAGATCGCCATGGCT GACCAGCTC$^{3'}$; R198K mutation, $^{5'}$CGTACAACGAC CTTAGGCAAGGGAGGCAGCGATTATACGGCAGCC$^{3'}$ and $^{5'}$GCCGTATAATCGCTGCCTCCCTTGCCTAAGGT CGTTGTACGACC$^{3'}$; D202E mutation, $^{5'}$CGCTTGGC CGTGGTGGATCCGAATATACGGCAGCCTTGCTGG$^{3'}$ and $^{5'}$GGCTGCCGTATATTCGGATCCACCACGGCCA AGCGTCGTTGTACG$^{3'}$; M251A mutation, $^{5'}$CGTTTG CCGAAGCTGCAGAGGCGGCAACTTTTGGTGC$^{3'}$ and $^{5'}$CCAAAAGTTGCCGCCTCTGCAGCTTCGGCAAACG CG$^{3'}$; L298X mutation (X is a stop codon), $^{5'}$GA AAATCCGCCGTAGTTCCGAGCTCTGGCGCTTCGTC G$^{3'}$ and $^{5'}$CGAAGCGCCAGAGCTCGGAACTACGGCG GATTTTCAG$^{3'}$. The mutant plasmids were identified by

site restriction analysis and were confirmed to carry the desired mutation by DNA sequencing.

Expression and purification of the wild-type and mutant proteins

BL21(DE3) *E. coli* cells were transformed by electroporation with the appropriate plasmid and were grown at 37 °C and made to express the desired protein by three hours induction with 1 mM IPTG as reported for NAGK.[18] Wild-type and mutant NAGK forms were purified as reported,[18] except for the use of a HiTrap Q HP 1 ml column mounted on an ÄKTA FPLC instrument (both from Amersham Biosciences) to speed the ion-exchange chromatography step, for the omission of the final dye column chromatography step, and for the utilization of a 50 mM sodium phosphate (pH 7) buffer containing 0.2 mM dithiothreitol throughout the entire purification procedure. For purification at 4 °C of wild-type and mutant forms of AKIII, the procedure was the same as for NAGK (sonication, centrifugation, two ammonium sulphate cuts and HiTrap Q HP chromatography) with the following modifications. (1) The initial suspending buffer (2 ml/g of cell pellet) for sonication was 10 mM potassium phosphate (pH 6.8) containing 0.1 mM KCl, 2 mM $MgCl_2$, 1 mM L-lysine monohydrochloride, 1 mM L-threonine, 1 mM Na_2EDTA and 0.1 mM dithiothreitol. (2) Ammonium sulphate cuts were at 20% and 40% saturation of the salt, collecting the precipitate from the 40% cut. (3) The buffer used for HiTrap Q HP chromatography step was 20 mM potassium phosphate (pH 6.8), 30 mM β-mercaptoethanol, 0.5 mM L-lysine, 0.1 mM Na_2EDTA, 0.1 mM $MgCl_2$, and the elution of the protein was effected with a 20 ml linear gradient between 0 M and 0.5 M KCl, in the same buffer. Wild-type and mutated forms of AKIII were eluted as a single peak at approximately 0.35 M KCl, except in the case of the truncated L298X mutant, which was eluted at 0.25 M KCl.

Enzyme activity assays and kinetic analysis

Hydroxylamine-based assays for both NAGK[19] and AKIII[23] were used for detecting the enzymes during purification and for initial evaluation of the activity of the purified preparations. However, for kinetic analysis and for most other assays, the more sensitive continuous spectrophotometric assays were used in which the production of ADP triggered by the addition of NAG (NAGK assay) or L-aspartate (AKIII assay[24]) was monitored at 340 nm, at 25 °C, using pyruvate kinase/lactate dehydrogenase coupling. In the coupled NAGK activity assay, the reaction mixture contained 20 mM sodium phosphate (pH 7), 10 mM β-mercaptoethanol, 0.3 mM NADH, 1.5 mM phosphoenolpyruvate, 50 mM KCl, 0.02 mg/ml of pyruvate kinase (from rabbit muscle; Roche Diagnostics) and 0.01 mg/ml of lactate dehydrogenase (from hog muscle; Boehringer Mannheim) and, except where indicated, 20 mM $MgCl_2$, 10 mM Na_2ATP (neutralized with NaOH) and 0.1 M NAG (neutralized with KOH). When ATP was varied, the concentration of $MgCl_2$ was made to exceed, by 10 mM, the concentration of the nucleotide. The coupled assay mixture for AKIII activity contained 20 mM potassium phosphate (pH 7.2), 10 mM β-mercaptoethanol, 0.6 mM NADH, 1.5 mM phosphoenolpyruvate, 0.15 M KCl, 0.02 mg/ml of pyruvate kinase, 0.01 mg/ml of lactate dehydrogenase and, except where indicated, 30 mM magnesium acetate, 16 mM Na_2ATP (neutralized with NaOH) and 0.2 M L-aspartate (neutralized with KOH). When ATP was varied, the concentration of magnesium acetate was made to exceed, by 10 mM, the concentration of the nucleotide. One enzyme unit is the amount of enzyme producing 1 μmol of ADP per minute at 25 °C.

The program GraphPadPrism was used for curve-fitting and estimation of the values of kinetic constants. Initial velocity results obtained at variable concentrations of substrate were fit either to hyperbolic kinetics, or to hyperbolic kinetics with inhibition by excess substrate:[37]

$$v = (V_{max}[S])/([S] + K_m + ([S]^2/K_i))$$

Plots of initial velocity *versus* variable lysine concentrations were fit to sigmoid inhibition kinetics, according to the expression:

$$v = 1 - ([Lys]^N/(K + [Lys]^N))$$

Construction of a 3-D model of the amino acid kinase domain of aspartokinase III

A model of the 3-D structure of the amino acid kinase domain of AKIII (residues 1–291, excluding the region from residues 56–106) was built with MODELLER 6v2[27] using as template the *E. coli*NAGK dimer complexed with MgAMPPNP and NAG[4] (PDB accession code 1GS5), based on the alignment between AKIII and NAGK sequences shown in Figure 4. The location of the gaps in the alignment was optimized iteratively by several rounds of model building and aligning of dubious regions to minimize progressively the distance between consecutive C^α atoms in the model. The ligands MgAMPPNP and aspartate were placed initially in the AKIII model in positions equivalent to those of the NAGK ligands. The model with the lowest value of MODELLER target function[27] was subjected to further optimization, by refinement of loops and manual tracing of side-chains in the active center. Sequence alignment restraints were removed at this stage for modelling the αD-β8 connection and the loop containing the β9-β10 hairpin, since these regions are longer in AKIII than in NAGK. In the final stages of modelling, program O[38] was used to trace manually some difficult parts of the model, using the library of commonly observed side-chain conformations of this program. Program O was used for superposition of the experimental NAGK and modelled AKIII structures. PROCHECK[29] and WHATIF[28] were used to evaluate the quality of the stereochemistry and of the packing and contacts, respectively. Quality indexes used were the average score from Verify3D[32] profiles, and the *Z*-score given by PROSAII[30] upon calculation of the 3-D mean force for the model.

Other methods

Size-exclusion chromatography was carried out on a Superdex TM 200 HR 10/30, 24 ml column (Amersham Biosciences) mounted on a ÄKTA FPLC system run at 0.25 ml/minute. The buffer used throughout was 50 mM Tris–HCl (pH 7.5) containing 0.15 M NaCl. Samples were applied in volumes of <0.2 ml. Elution of the injected proteins was monitored continuously by measuring the absorbance of the effluent at 280 nm. Semilogarithmic plots of the molecular mass of the different calibration standards *versus* the volumes at

which they were eluted were used for mass determination of unknowns.

When required, protein solutions were concentrated by centrifugal ultrafiltration using Microsep 10K devices from Gelman. Protein was assayed by the method of Bradford[39] using a commercial reagent from Bio-Rad, and bovine serum albumin as a standard. SDS-PAGE was carried out according to Laemmli.[40] When quantitative determinations in the gel were needed, Coomassie brilliant blue was eluted from SDS-PAGE bands by gentle rocking of the excised bands in 25% (v/v) pyridine, followed by determination of the color at 605 nm.[21]

Secondary structure prediction was calculated using programs 3D-PSSM†[41] and 123D‡.[42] The program GraphPadPrism was used for curve-fitting and statistical evaluation.

Illustrations

Figures 1, 5 (middle and bottom panels) and 9 were prepared using MOLSCRIPT[43] and Raster3D.[44]

Protein Data Bank accession number

Coordinates for the modelled AKIII enzyme described here have been deposited in the Protein Data Bank with accession code 1OHI.

Acknowledgements

This work was supported by grants Rayos X of Fundación Ramón Areces, GV01-259 of the Generalitat Valenciana and BMC2001-2182 of the Spanish Ministry of Science and Technology (MCYT). C.M.-M. and S.T. are predoctoral fellows of the Generalitat Valenciana and of the Spanish Ministry of Science and Technology (MCYT).

References

1. Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R. *et al.* (2002). The Pfam protein families database. *Nucl. Acids Res.* **30**, 276–280.
2. Marina, A., Alzari, P. M., Bravo, J., Uriarte, M., Barcelona, B., Fita, I. & Rubio, V. (1999). Carbamate kinase: new structural machinery for making carbamoyl phosphate, the common precursor of pyrimidines and arginine. *Protein Sci.* **8**, 934–940.
3. Ramón-Maiques, S., Marina, A., Uriarte, M., Fita, I. & Rubio, V. (2000). The 1.5 Å resolution crystal structure of the carbamate kinase-like carbamoyl phosphate synthetase from the hyperthermophilic archaeon *Pyrococcus furiosus*, bound to ADP, confirms that this thermostable enzyme is a carbamate kinase, and provides insight into substrate binding and stability in carbamate kinases. *J. Mol. Biol.* **299**, 463–476.
4. Ramón-Maiques, S., Marina, A., Gil-Ortiz, F., Fita, I. & Rubio, V. (2002). Structure of acetylglutamate kinase, a key enzyme for arginine biosynthesis and a prototype for the amino acid kinase family, during catalysis. *Structure*, **10**, 329–342.
5. Gil-Ortiz, F., Ramón-Maiques, S., Fita, I. & Rubio, V. (2003). The course of phosphorus in the reaction of *N*-acetyl-L-glutamate kinase, determined from the structures of crystalline complexes, including a complex with an AlF_4^- transition state mimic. *J. Mol. Biol.* **331**, 231–244.
6. Ramon-Maiques, S., Britton, H. G. & Rubio, V. (2002). Molecular physiology of phosphoryl group transfer from carbamoyl phosphate by a hyperthermophilic enzyme at low temperature. *Biochemistry*, **41**, 3916–3924.
7. Truffa-Bachi, P. (1973). Microbial aspartokinases. In *The Enzymes* (Boyer, P., ed.), pp. 509–553, Academic Press, New York.
8. Cassan, M., Parsot, C., Cohen, G. N. & Patte, J. C. (1986). Nucleotide sequence of lysC gene encoding the lysine-sensitive aspartokinase III of *Escherichia coli* K12. Evolutionary pathway leading to three isofunctional enzymes. *J. Biol. Chem.* **261**, 1052–1057.
9. Azevedo, R. A. & Lea, P. J. (2001). Lysine metabolism in higher plants. *Amino Acids*, **20**, 261–279.
10. Rapaport, E., Levina, A., Metelev, V. & Zamecnik, P. C. (1996). Antimycobacterial activities of antisense oligodeoxynucleotide phosphorothioates in drug-resistant strains. *Proc. Natl Acad. Sci. USA*, **93**, 709–713.
11. Blanco, J. & Viola, R. E. (2002). Purification, crystallization and preliminary X-ray analysis of aspartokinase III from *Escherichia coli*. *Acta Crystallog. sect. D*, **58**, 352–354.
12. Richaud, C., Mazat, J. P., Gros, C. & Patte, J. C. (1973). Subunit structure of aspartokinase III of *Escherichia coli* K12. *Eur. J. Biochem.* **40**, 619–629.
13. Moir, D. & Paulus, H. (1977). Properties and subunit structure of aspartokinase II from *Bacillus subtilis* VB217. *J. Biol. Chem.* **252**, 4648–4651.
14. Sritharan, V., Wheeler, P. R. & Ratledge, C. (1989). Metabolism of aspartate in *Mycobacterium smegmatis*. *Eur. J. Biochem.* **180**, 587–593.
15. Kalinowski, J., Cremer, J., Bachmann, B., Eggeling, L., Sahm, H. & Puhler, A. (1991). Genetic and biochemical analysis of the aspartokinase from *Corynebacterium glutamicum*. *Mol. Microbiol.* **5**, 1197–1204.
16. Kobashi, N., Nishiyama, M. & Tanokura, M. (1999). Kinetic and mutation analyses of aspartate kinase from *Thermus flavus*. *J. Biosci. Bioeng.* **87**, 739–745.
17. Chipman, D. M. & Shaanan, B. (2001). The ACT domain family. *Curr. Opin. Struct. Biol.* **11**, 694–700.
18. Gil, F., Ramón-Maiques, S., Marina, A., Fita, I. & Rubio, V. (1999). *N*-Acetyl-L-glutamate kinase from *Escherichia coli*: cloning of the gene, purification and crystallization of the recombinant enzyme and preliminary X-ray analysis of the free and ligand-bound forms. *Acta Crystallog. sect. D*, **55**, 1350–1352.
19. Haas, D. & Leisinger, T. (1975). *N*-acetylglutamate 5-phosphotransferase of *Pseudomonas aeruginosa*. Purification and ligand-directed association–dissociation. *Eur. J. Biochem.* **52**, 365–375.
20. Jaworek, D. & Welsch, J. (1986). Adenosine 5′-diphosphate and adenosine 5′-monophosphate. UV-Method. In *Methods of Enzymatic Analysis* (Bergmeyer, H. U., ed.), 3rd edit., vol. 7, pp. 365–370, VCH Verlagsgesellschaft, Weinheim, Germany.
21. Hames, B. D. (1981). An introduction to polyacrylamide gel electrophoresis. In *Gel Electrophoresis of Proteins: A Practical Approach* (Hames, B. D. &

† http://www.sbg.bio.ic.ac.uk/servers/3dpssm
‡ http://123D.ncifcrf.gov

Rickwood, D., eds), vol. 46, pp. 46, IRL Press, Oxford, Washington DC.
22. Cunin, R., Glansdorff, N., Pierard, A. & Stalon, V. (1986). Biosynthesis and metabolism of arginine in bacteria. *Microbiol. Rev.* **50**, 314–352.
23. Stadtman, E. R., Cohen, G. N., LeBras, G. & Robichon-Szulmajster, H. (1961). Feed-back inhibition and repression of aspartokinase activity in *Escherichia coli* and *Saccharomyces cerevisiae*. *J. Biol. Chem.* **236**, 2033–2038.
24. Wampler, D. E. & Westhead, E. W. (1968). Two aspartokinases from *Escherichia coli*. Nature of the inhibition and molecular changes accompanying reversible inactivation. *Biochemistry*, **7**, 1661–1671.
25. Patte, J. C., Loviny, T. & Cohen, G. N. (1965). Co-operative inhibitory effects of L-lysine with other amino acids on an aspartokinase from *Escherichia coli*. *Biochim. Biophys. Acta*, **99**, 523–530.
26. Richaud, C., Mazat, J. P., Felenbok, B. & Patte, J. C. (1974). The role of lysine and leucine binding on the catalytical and structural properties of aspartokinase III of *Escherichia coli* K 12. *Eur. J. Biochem.* **48**, 147–156.
27. Sali, A. & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815.
28. Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* **8**, 52–56.
29. Laskowsky, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
30. Sippl, M. J. (1993). Recognition of errors in three-dimensional structures of proteins. *Proteins: Struct Funct. Genet.* **17**, 355–362.
31. Ogawa-Miyata, Y., Kojima, H. & Sano, K. (2001). Mutation analysis of the feedback inhibition site of aspartokinase III of *Escherichia coli* K-12 and its use in L-threonine production. *Biosci. Biotechnol. Biochem.* **65**, 1149–1154.
32. Eisenberg, D., Luthy, R. & Bowie, J. U. (1997). VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol.* **277**, 396–404.
33. Gallagher, D. T., Gilliland, G. L., Xiao, G., Zondlo, J., Fisher, K. E., Chinchilla, D. & Eisenstein, E. (1998). Structure and control of pyridoxal phosphate dependent allosteric threonine deaminase. *Structure*, **6**, 465–475.
34. Kobe, B., Jennings, I. G., House, C. M., Michell, B. J., Goodwill, K. E., Santarsiero, B. *et al.* (1999). Structural basis of autoregulation of phenylalanine hydroxylase. *Nature Struct. Biol.* **6**, 442–448.
35. Schuller, D. J., Grant, G. A. & Banaszak, L. J. (1995). The allosteric ligand site in the Vmax-type cooperative enzyme phosphoglycerate dehydrogenase. *Nature Struct. Biol.* **2**, 69–76.
36. Niles, E. G. & Westhead, E. W. (1973). The variable subunit structure of lysine-sensitive aspartylkinase from *Escherichia coli* TIR-8. *Biochemistry*, **12**, 1715–1722.
37. Dixon, M. & Webb, E. C. (1979). *Enzymes*, 3rd edit., Longman Group Ltd, London.
38. Jones, T. A., Zou, J., Cowan, S. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
39. Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Anal. Biochem.* **72**, 248–254.
40. Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680–685.
41. Kelley, L. A., MacCallum, R. M. & Sternberg, M. J. (2000). Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.* **299**, 499–520.
42. Alexandrov, N. N., Nussinov, R. & Zimmer, R. M. (1995). Fast protein fold recognition via sequence to structure alignment and contact capacity potentials. In *Pacific Symposium on Biocomputing '96* (Hunter, L. & Klein, T. E., eds), pp. 53–72, World Scientific Publishing Co., Singapore.
43. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
44. Merritt, E. A. & Murphy, M. E. P. (1994). Raster3D version 2.0. A program for photorealistic molecular graphics. *Acta Crystallog. sect. D*, **50**, 869–873.

Edited by R. Huber

(Received 29 May 2003; received in revised form 15 September 2003; accepted 19 September 2003)

Capítulo 2

First-time crystallization and preliminary X-ray crystallographic analysis of a bacterial-archaeal type UMP kinase, a key enzyme in microbial pyrimidine biosynthesis.

Trabajo publicado en

Biochim. Biophys. Acta. **(2005) 1747, pags. 271-275**

Available online at www.sciencedirect.com

SCIENCE DIRECT

Biochimica et Biophysica Acta 1747 (2005) 271–275

BIOCHIMICA ET BIOPHYSICA ACTA BBA

http://www.elsevier.com/locate/bba

Short crystallization paper

First-time crystallization and preliminary X-ray crystallographic analysis of a bacterial-archaeal type UMP kinase, a key enzyme in microbial pyrimidine biosynthesis

Clara Marco-Marín, Juan Manuel Escamilla-Honrubia, Vicente Rubio*

Instituto de Biomedicina de Valencia, Consejo Superior de Investigaciones Científicas (IBV-CSIC), C/Jaime Roig 11, 46010-Valencia, Spain

Received 1 October 2004; received in revised form 16 November 2004; accepted 17 November 2004
Available online 30 November 2004

Abstract

UMP phosphorylation, a key step for pyrimidine nucleotide biosynthesis, is catalyzed in bacteria by UMP kinase (UMPK), an enzyme specific for UMP that is dissimilar to the eukaryotic UMP/CMP kinase or to other nucleoside monophosphate kinases. UMPK is allosterically regulated and participates in pyrimidine-triggered gene repression. As first step towards determining UMPK structure, the putative UMPK-encoding gene of the hyperthermophilic archaeon *Pyrococcus furiosus* was cloned and overexpressed in *Escherichia coli*. The protein product was purified and confirmed to be a genuine UMPK. It was crystallized at 294 K in hanging drops by the vapor diffusion technique using 3.5–4 M Na formate. Cubic 0.2-mm crystals diffracted synchrotron X-rays to 2.4-Å resolution. Space group was *I*23 ($a=b=c=144.95$ Å), and the asymmetric unit contained two monomers, with 52% solvent content. The self-rotation function suggests that the enzyme is hexameric, which agrees with biochemical studies on bacterial UMPKs.

Keywords: Pyrimidine nucleotide biosynthesis; Amino acid kinase family; Uridylate kinase; UMP kinase

In the de novo biosynthesis of pyrimidine nucleotides, the phosphoryl transfer from ATP to UMP plays a key role, opening the way for synthesis of all other pyrimidine nucleotides [1]. The enzyme responsible for UMP phosphorylation in eukaryotes (including animals), UMP/CMP kinase, similarly to other nucleoside monophosphate kinases, resembles adenylate kinase in both amino acid sequence and 3-D structure [2,3]. In contrast, bacteria have a dedicated UMP kinase (UMPK) of unknown structure, which is highly specific for UMP and which, on the basis of sequence comparisons, does not resemble either UMP/CMP kinase, adenylate kinase, or any other nucleoside monophosphate kinase [4,5], but which resembles proteins of the amino acid kinase family [5]. This family(PF00696, http://www.sanger.ac.uk/Software/Pfam) includes the homodimeric enzymes carbamate kinase and acetylglutamate kinase, which have been found to share a characteristic fold [6,7].

Bacterial UMPK appears to be a key point of feed-back control of pyrimidine biosynthesis, since it has been found to be inhibited by UTP and activated by GTP [4,5]. Surprisingly, studies in *Escherichia coli* have shown [8] that UMPK plays a crucial role in the repression by pyrimidines of the expression of the *carAB* genes, the genes that encode the two subunits of the first enzyme of the pyrimidine biosynthesis pathway, carbamoyl phosphate synthetase. This important gene regulatory role involves also two DNA binding proteins, PEPA and IHF, but UMPK is the element that renders the complex sensitive to pyrimidines [8]. Thus, UMPK associates the triple quality of being a catalyst, an allosterically regulated device and a component of a gene expression regulatory complex. These characteristics, and the fact that UMPK plays a key metabolic function and has an exclusive microbial distribution, render highly desirable the elucidation

* Corresponding author. Tel.: +34 963 391 772; fax: +34 963 690 800.
E-mail address: rubio@ibv.csic.es (V. Rubio).

1570-9639/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbapap.2004.11.010

```
ECOL   ATNAKPVYKRILLKLSGEALQGTEGFGIDASILDRMAQEIKELVELGIQVGVVIGGGNLF 60
PFUR   --------MRIVFDIGGSVLVPENP---DIDFIKEIAYQLTKVSED-HEVAVVVGGGKLA 48
BSUB   --MEKPKYKRIVLKLSGEALAGEQGNGINPTVIQSIAKQVKEIAELEVEVAVVVGGGNYG 58

ECOL   RGAGLAKAGMNR--VVGDHMGMLATVMNGLAMRDALHRAYVNARLMSAIPLNGVCDSYSW 118
PFUR   RKYIEVAEKFNSSETFKDFIGIQITRANAMLLIAALREKAYPVVVED------------F 96
BSUB   AEKTGSDLGMDR--ATADYMGMLATVMNSLALQDSLETLGIQSRVQTSIEMRQVAEPYIR 116

ECOL   AEAISLLRNNRVVILSAGTGNPFFTTDSAACLRGIEIEADVVLKATKVDGVFTADPAKD 177
PFUR   WEAWKAVQLKKIPVMGG--THPGHTTDAVAALLAEFLKADLLVVITNVDGVYTADPKKD 153
BSUB   RKAIRHLEKKRVVIFAAGTGNPYFSTDTTAALRAAEIEADVILMAKNNVDGVYNADPRKD 176

ECOL   PTATMYEQLTYSEVLEKE---------LKVMDLAAFTLARDHKLPIRVFNMNKPGALRRV 228
PFUR   PTAKKIKKMKPEELLEIVGKGIEKAGSSSVIDPLAAKIIARSGIKTIVIGKEDAKDLFRV 213
BSUB   ESAVKYESLSYLDVLKDG---------LEVMDSTASSLCMDNDIPLIVFSIMEEGNIKRA 227

ECOL   VMGEKEGTLITE- 240
PFUR   IKGDHNGTTIEP- 225
BSUB   VIGESIGTIVRGK 240
```

Fig. 1. Amino acid sequence alignment of *E. coli* (ECOL), *P. furiosus* (PFUR) and *B. subtilis* (BSUB) UMP kinases. Residues that are identical to those in *P. furiosus* are in white letters over a dark grey background and those that are conservatively replaced are black letters over light grey shadowing. The alignment was constructed with ClustalW, version 1.8 [21].

of UMPK structure at atomic resolution. As initial steps towards such goal, we report here the cloning of the putative gene for UMPK (*pyrH*) of the hyperthermophilic archaeon *Pyrococcus furiosus*, its heterologous overexpression in *E. coli*, the purification of the protein product, confirming by enzyme assays that it really corresponds to a highly specific UMPK, and the crystallization and preliminary X-ray diffraction of the enzyme. In addition, this is the first report on an archaeal UMPK.

The putative *pyrH* gene of *P. furiosus* appeared a good candidate to encode a genuine UMPK. The 225-residue putative protein encoded by this gene resembles in length and in sequence (27–28% identity) the UMPKs of *E. coli* and *Bacillus subtilis* [4,5] (Fig. 1). In contrast, no significant sequence identity was found with UMP/CMP kinases. For example, sequence identity with the 204- and 194-residue yeast and *Dyctiostelium discoideum* UMP/CMP kinases was <11% (data not shown).

We PCR-amplified the putative *pyrH* gene (nucleotides 1321789 to 1322466 of the *P. furiosus* genome, corresponding to gene PF 1407; http://www.tigr.org) from genomic DNA of *P. furiosus*, using a high fidelity proofreading thermostable DNA polymerase (Deep Vent; from New England Biolabs) and the direct and reverse primers 5′-GGTAATTTGGTGAA*CAT*ATGAGGATAGTG-3′ and 5′-TCGTGACC*GG*ATCCATACATCAC-3′. These primers are designed to amplify the entire open reading frame and to introduce by mutation (bases shown in italic) a *Nde*I site (direct primer) and a *Bam*HI site (reverse primer) just upstream of the initiator ATG and downstream of the stop codon, respectively. The PCR product was digested with *Nde*I and *Bam*HI and was ligated into the corresponding sites of plasmid pET-15b (from Novagen), followed by electrotransformation of *E. coli* DH5α (from Clontech) and isolation of plasmid pUKPFU, which contains the properly inserted *pyrH* gene (corroborated by sequencing). The gene was overexpressed at 37 °C by induction with 1 mM isopropyl-β-D-thiogalactoside of 2 l, late-exponential (A_{600}=0.5) cultures (in LB broth containing 0.1 mg/ml ampicillin and 0.05 mg/ml chloramphenicol) of BL21(DE3) *E. coli* cells (Novagen) carrying plasmids pUKPFU and pSJS1240 (a pRI952 [9] derivative encoding tRNAs for the rare codons AGA, AGG and ATA). SDS-PAGE [10] of centrifuged sonicates of the collected cells (suspended in 2.5 volumes of HEPES 20 mM pH 8.0, 1 mM dithiothreitol) revealed the presence in the supernatants of a prominent protein band of 26.5 kDa when the cells carried pUKPFU, but not when they had been transformed with the parental pET15b plasmid (Fig. 2, lanes 1 and 2). The mass of the

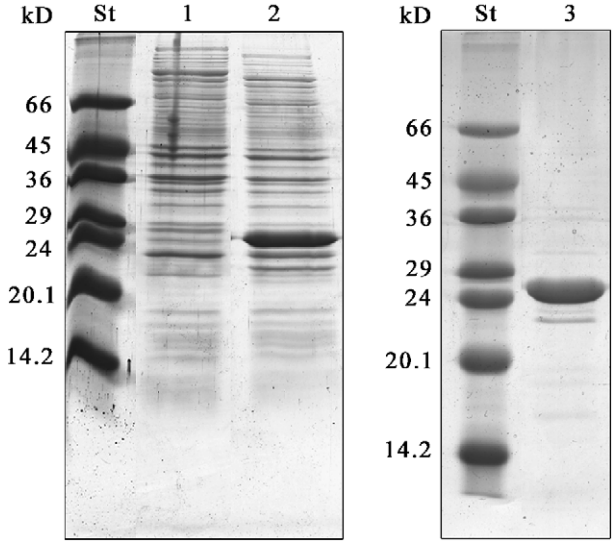


Fig. 2. Expression and purification of *P. furiosus* UMP kinase. SDS-PAGE (15% polyacrylamide gels) of cell extracts of *E. coli* BL21 (DE3) transformed with plasmid pSJS1240 and with either pET15b (lane 1) or pUKPFU (lane 2). Lane 3, purified enzyme. The lanes labeled St correspond to protein standards of the masses (in kDa) indicated on the side.

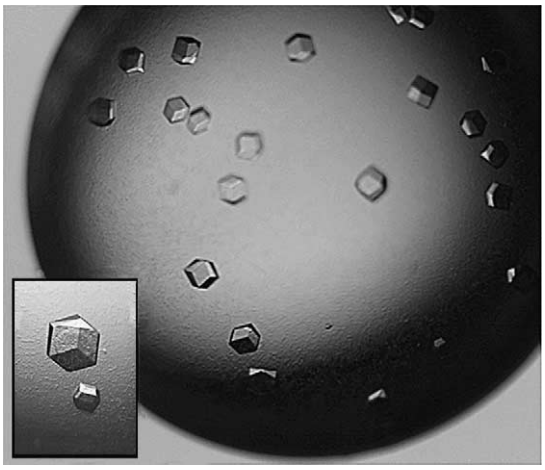


Fig. 3. Typical crystals of *P. furiosus* UMP kinase.

overexpressed protein agrees with the sequence-deduced expected mass of 26,653 Da, a mass that corresponds to the *pyrH*-encoded product supplemented with the 20-residue 6-His-containing N-terminal extension that is introduced by cloning at the *Nde*I site of pET15b. A single nucleotide change (thymine to adenine) at position 973 of the coding sequence was detected in the cloned gene, relative to the sequence in the database (http://www.tigr.org), causing the conservative amino acid change S195T. The expressed protein was purified easily without having to use the 6-His tag, since it was enough to heat for 5 min at 90 °C the supernatant of the centrifuged (30 min, 25,000×*g*) sonicate, followed by chilling to 4 °C and centrifugation, and by two passages (at 4 °C) in succession through a 1×18-cm Q-Sepharose Fast Flow (from Amersham Biosciences) column, using as protein solvent and as column and washing buffer the sonication buffer (first column passage) and Na Bicine 20 mM pH 8.5 containing 1 mM dithiothreitol (second column passage). The *pyrH* product was not retained in any of the two passages, resulting in a near-homogeneous preparation of the protein after the second passage (Fig. 2, lane 3).

The purified product was proven to catalyze the phosphorylation of UMP by using the standard NADH-coupled spectrophotometric assay for UMPK activity [11] at 37 °C: there was activity (monitored continuously by the fall in A_{280}) only when UMP (used at a concentration of 0.3 mM) was added, and the activity was nil when UMP was replaced by either CMP, TMP, AMP or GMP. This high specificity for UMP is characteristic for bacterial-type UMPKs [4,5]. The specific activity was 13 U/mg, a value that is not too far from the specific activity of the purified *B. subtilis* enzyme (26 U/mg) [5], particularly taking into consideration that, at mesophilic temperatures, the enzymes from hyperthermophiles frequently exhibit lower activities than the corresponding enzymes of mesophiles [12]. Thus, given the similarities in the sequence and the specificity, *P. furiosus* (and possibly other archaea) uses for phosphorylating UMP a UMPK of bacterial type (or, from our results, of bacterial-archaeal type) rather than a UMP/CMP kinase as is used by eukaryotes.

For crystallization experiments, pure UMPK, in Na Bicine 20 mM pH 8.5, 1 mM dithiothreitol, was concentrated to 11 mg/ml by centrifugal ultrafiltration (Microsep 10K device, from Pall Filtron). The initial crystallization screening at 293 K was carried out using a commercial sparse-matrix screening kit (Crystal screen, from Hampton Research, Aliso Viejo, CA, USA) [13] and the hanging-drop vapor diffusion method. Drops were prepared by mixing 1.5 μl of reservoir fluid and 1.5 μl of the UMPK solution. Crystals appeared within 2 days when the crystallization solution was 4 M Na formate. Improvement trials using graded concentrations of Na formate yielded in 3–4 days good-diffracting crystals of cubic appearance and 0.2-mm maximum dimension (Fig. 3) when using 3.5–4 M Na formate. One crystal, cryoprotected with 10% glycerol (supplemented to the reservoir fluid containing 3.5 M Na formate), was flash-frozen in liquid nitrogen and diffracted at a temperature of 100 K with a cryo-cooling system (Oxford Cryosystems, England) at the BM16 beamline (λ=0.9794 Å) of the ESRF synchrotron (Grenoble, France), using a Quantum ADSC Q4 CCD detector. The crystal to detector distance was set to 170 mm, with a rotation angle of 1.0° and an exposure time of 100 s. A complete data set was collected at 2.4-Å resolution and was indexed and processed with programs MOSFLM [14], SCALA [15] and TRUNCATE [16]. The space group of the UMPK crystals was determined to be *I*23, with cubic unit-cell dimensions $a=b=c$=144.94 Å. The data set is summarized in Table 1. The asymmetric unit probably contains two monomers, since this would give 52% solvent content and a Matthews coefficient [17] of 2.59 $Å^3$ Da^{-1}. The presence of two monomers in the asymmetric unit would fit an organization in which the enzyme basic molecular unit is a homodimer, as is the case with the homologous enzymes [5] carbamate kinase [6,18] and acetylglutamate kinase [7]. However the results of analytical ultracentrifugation experiments with *E. coli* and *B. subtilis* UMP kinases [4,5] indicate that these enzymes are hexameric. Calculation of the self-rotation function with MOLREP [19] yielded results that are

Table 1
X-ray data collection statistics

Beamline	ESRF BM16
Wavelength (Å)	0.979471
Space group	I23
Unit-cell parameters (Å)	$a=b=c$=144.95
Resolution range (Å)	45.64–2.40 (2.53–2.40)
Total number of reflections	148,826
Unique reflections	19,936
Data completeness (%)	100 (100)
R_{sym} (%)[a]	6.6 (37.2)
I/σ (I)	9.5 (2)

Values in parentheses are for the highest resolution shell.

[a] $R_{sym}=\Sigma|I-<I>|/\Sigma I$, where I is the observed intensity, and <I> is the average intensity of multiple observations of symmetry-related reflections.

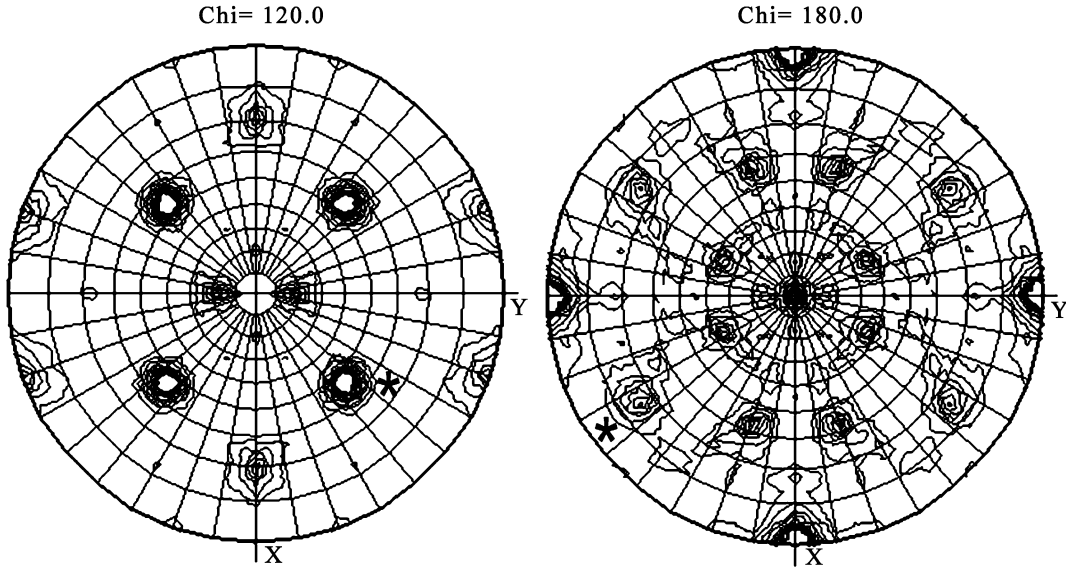


Fig. 4. Representation of the χ=180° and χ=120° sections of the self-rotation function for *P. furiosus* UMPK crystals. Start level is 1σ and the step size is 0.75σ. The ortogonalization code follows the PDB convention. The strong peaks in the χ=120° section correspond to the crystal threefold axes (the diagonals of the cubic cell). A peak for a non-crystallographic twofold axis, at θ=73.82° and ϕ=−56.85°, is labeled with an asterisk in the χ=180° section. This axis appears perpendicular to the crystal threefold axis reflected in the peak at θ=54.74° and ϕ=45.0° (also labeled with an asterisk) in the χ=120° section.

suggestive of an hexameric quaternary structure formed by the association of three dimers, since the sections χ=120° and χ=180° of this function (Fig. 4), in addition to showing strong peaks according to the crystallographic symmetry of cubic space group *I*23, contained in the section χ=180° a well-defined non-crystallographic peak perpendicular to a crystal threefold axes (see the legend to Fig. 4). This is compatible with the packing of molecular hexamers, presenting 32 symmetry, having the molecular threefold axis coincident with a crystal threefold axis and two subunits related by a non-crystallographic twofold axis in the crystal asymmetric unit (as it is inferred from density packing considerations, see above).

Attempts to find an initial solution by molecular replacement using programs MOLREP [19] and AMoRe [20] and a polyalanine model generated from *E. coli* acetylglutamate kinase [7] have given no conclusive solutions, and the structure determination using the multiple-wavelength anomalous dispersion (MAD) approach is under way.

Acknowledgements

The BM16 beamline of ESRF (Grenoble) was provided by the Spanish Ministry of Education and Science. This work was supported by grants from the Ministerio de Ciencia y Tecnología (BMC2001-2182) and the Instituto de Salud Carlos III of the Ministerio de Sanidad y Consumo (RCMN C03/08 and REDEMETH G03/54). C.M.-M. is a I3P fellow of CSIC-Bruker España S.A. We thank F.E. Jenney, Jr. (Dept of Biochemistry, University of Georgia, Athens) for providing the genomic DNA from *Pyrococcus furiosus* and S.J. Sandler (Dept. Microbiology, University of Massachussets) for providing plasmid pSJS1240. Staff members of ESRF beamline BM16, Grenoble (France) are thanked for their help.

References

[1] G. Michal, Biochemical Pathways, John Wiley and Sons, Inc., New York, 1999.
[2] H.J. Muller-Dieckmann, G.E. Schulz, The structure of uridylate kinase with its substrates, showing the transition state geometry, J. Mol. Biol. 236 (1994) 361–367.
[3] K. Scheffzek, W. Kliche, L. Wiesmuller, J. Reinstein, Crystal structure of the complex of UMP/CMP kinase from *Dictyostelium discoideum* and the bisubstrate inhibitor P1-(5′-adenosyl) P5-(5′-uridyl) pentaphosphate (UP5A) and Mg^{2+} at 2.2 Å: implications for water-mediated specificity, Biochemistry 35 (1996) 9716–9727.
[4] L. Serina, C. Blondin, E. Krin, O. Sismeiro, A. Danchin, H. Sakamoto, A.M. Gilles, O. Bârzu, *Escherichia coli* UMP-kinase, a member of the aspartokinase family, is a hexamer regulated by guanine nucleotides and UTP, Biochemistry 34 (1995) 5066–5074.
[5] C. Gagyi, N. Bucurenci, O. Sîrbu, G. Labessse, M. Ionescu, A. Ofiteru, A. Liliane, L. Stephanie, A. Danchin, O. Bârzu, A.M. Gilles, UMP kinase from the Gram-positive bacterium *Bacillus subtilis* is strongly dependent on GTP for optimal activity, Eur. J. Biochem. 270 (2003) 3196–3204.
[6] A. Marina, P.M. Alzari, J. Bravo, M. Uriarte, B. Barcelona, I. Fita, V. Rubio, Carbamate kinase: new structural machinery for making carbamoyl phosphate, the common precursor of pyrimidines and arginine, Protein Sci. 8 (1999) 934–940.
[7] S. Ramon-Maiques, A. Marina, F. Gil-Ortiz, I. Fita, V. Rubio, Structure of acetylglutamate kinase, a key enzyme for arginine biosynthesis and a prototype for the amino acid kinase enzyme family, during catalysis, Structure 10 (2002) 329–342.
[8] A. Kholti, D. Charlier, D. Gigot, N. Huysveld, M. Roovers, N. Glansdorff, *pyrH*-encoded UMP-kinase directly participates in pyrimidine-specific modulation of promoter activity in *Escherichia coli*, J. Mol. Biol. 280 (1998) 571–582.

[9] B.J. Del Tito Jr., J.M. Ward, J. Hodgson, C.J. Gershater, H. Edwards, L.A. Wysocki, F.A. Watson, G. Sathe, J.F. Kane, Effects of a minor isoleucyl tRNA on heterologous protein translation in *Escherichia coli*, J. Bacteriol. 177 (1995) 7086–7091.
[10] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature (Lond.) 227 (1976) 680–685.
[11] C. Blondin, L. Serina, L. Wiesmuller, A.M. Gilles, O. Bârzu, Improved spectrophotometric assay of nucleoside monophosphate kinase activity using the pyruvate kinase/lactate dehydrogenase coupling system, Anal. Biochem. 220 (1994) 219–221.
[12] S. Ramon-Maiques, H.G. Britton, V. Rubio, Molecular physiology of phosphoryl group transfer from carbamoyl phosphate by a hyperthermophilic enzyme at low temperature, Biochemistry 41 (2002) 3916–3924.
[13] J. Jancarik, S.H. Kim, Sparse matrix sampling: a screening method for crystallization of proteins, J. Appl. Crystallogr. 24 (1991) 409–411.
[14] A.G.W. Leslie, Joint CCP4-ESF-EAMCB Newsletter on Protein Crystallography, vol. 26, 1992.
[15] P.R. Evans, Joint CCP4-ESF-EAMCB Newsletter on Protein Crystallography, vol. 33, 1997, pp. 22–24.
[16] CCP4 (Collaborative Computational Project Number 4), The CCP4 suite: programs for protein crystallography, Acta Crystallogr., D Biol. Crystallogr. 50 (1994) 760–763.
[17] B.W. Matthews, Solvent content of protein crystals, J. Mol. Biol. 33 (1968) 491–497.
[18] S. Ramon-Maiques, A. Marina, M. Uriarte, I. Fita, V. Rubio, The 1.5 Å resolution crystal structure of the carbamate kinase-like carbamoylphosphate synthetase from the hyperthermophilic archaeon *Pyrococcus furiosus*, bound to ADP, confirms that this thermostable enzyme is a carbamate kinase, and provides insight into substrate binding and stability in carbamate kinases, J. Mol. Biol. 299 (2000) 463–476.
[19] A. Vagin, A. Teplyakov, MOLREP: an automated program for molecular replacement, J. Appl. Crystallogr. 30 (1997) 1022–1025.
[20] J. Navaza, AMoRe: an automated package for molecular replacement, Acta Crystallogr., A 50 (1994) 157–163.
[21] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.

Capítulo 3

The crystal structure of *Pyrococcus furiosus* UMP kinase provides insight into catalysis and regulation in microbial pyrimidine nucleotide biosynthesis.

Trabajo publicado en

Journal of Molecular Biology **(2005) 352, pags. 438-454**

doi:10.1016/j.jmb.2005.07.045 J. Mol. Biol. (2005) 352, 438–454

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

The Crystal Structure of *Pyrococcus furiosus* UMP Kinase Provides Insight into Catalysis and Regulation in Microbial Pyrimidine Nucleotide Biosynthesis

Clara Marco-Marín, Fernando Gil-Ortiz and Vicente Rubio*

Instituto de Biomedicina de Valencia (IBV-CSIC), Jaume Roig 11,Valencia-46010, Spain

UMP kinase (UMPK), the enzyme responsible for microbial UMP phosphorylation, plays a key role in pyrimidine nucleotide biosynthesis, regulating this process *via* feed-back control and *via* gene repression of carbamoyl phosphate synthetase (the first enzyme of the pyrimidine biosynthesis pathway). We present crystal structures of *Pyrococcus furiosus* UMPK, free or complexed with AMPPNP or AMPPNP and UMP, at 2.4 Å, 3 Å and 2.55 Å resolution, respectively, providing a true snapshot of the catalytically competent bisubstrate complex. The structure proves that UMPK does not resemble other nucleoside monophosphate kinases, including the UMP/CMP kinase found in animals, and thus UMPK may be a potential antimicrobial target. This enzyme has a homohexameric architecture centred around a hollow nucleus, and is organized as a trimer of dimers. The UMPK polypeptide exhibits the amino acid kinase family (AAKF) fold that has been reported in carbamate kinase and acetylglutamate kinase. Comparison with acetylglutamate kinase reveals that the substrates bind within each subunit at equivalent, adequately adapted sites. The UMPK structure contains two bound Mg ions, of which one helps stabilize the transition state, thus having the same catalytic role as one lysine residue found in acetylglutamate kinase, which is missing from *P. furiosus* UMPK. Relative to carbamate kinase and acetylglutamate kinase, UMPK presents a radically different dimer architecture, lacking the characteristic 16-stranded β-sheet backbone that was considered a signature of AAKF enzymes. Its hexameric architecture, also a novel trait, results from equatorial contacts between the A and B subunits of adjacent dimers combined with polar contacts between A or B subunits, and may be required for the UMPK regulatory functions, such as gene regulation, proposed here to be mediated by hexamer–hexamer interactions with the DNA-binding protein PepA.

Keywords: pyrimidine nucleotide biosynthesis; amino acid kinase family; uridylate kinase; UMP kinase; protein structure

*Corresponding author

Introduction

In the "*de novo*" biosynthesis of pyrimidine nucleotides, the phosphorylation of UMP by ATP, yielding UDP, opens the way for synthesis of all other pyrimidine nucleotides (Figure 1).[1,2] The enzyme responsible for UMP phosphorylation differs in eukaryotes and in prokaryotes. Whereas the eukaryotic enzyme has dual specificity for UMP and CMP (UMP/CMP kinase) and, as other nucleoside monophosphate kinases, resembles adenylate kinase,[3] prokaryotes have a dedicated and strictly UMP-specific UMP kinase[1,4] (UMPK), which appears to be essential for growth,[5–8] but which does not resemble UMP/CMP kinase, adenylate kinase, or any other nucleoside monophosphate kinase.[5] Thus, prokaryotic UMPK is a candidate target for developing antimicrobials. Sequence

Present address: F. Gil-Ortiz, Centro de Investigación Príncipe Felipe (FVIB-CSIC), Avda. Autopista del Saler 16, Valencia-46013, Spain.

Abbreviations used: UMPK, UMP kinase (also called uridylate kinase); NAGK, *N*-acetyl-L-glutamate kinase; CK, carbamate kinase; PepA, aminopeptidase A; IHF, integration host factor; AAKF, amino acid kinase family.

E-mail address of the corresponding author: rubio@ibv.csic.es

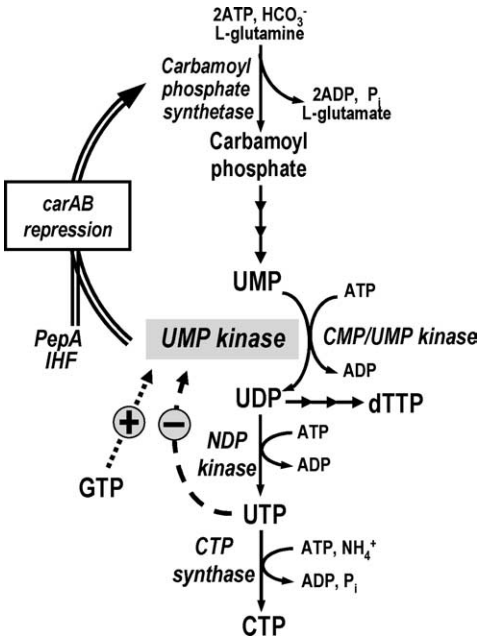


Figure 1. Pyrimidine nucleotide biosynthesis and the roles of UMPK demonstrated in *Escherichia coli*. Prokaryotic UMPK is highlighted within a grey box. The eukaryotic CMP/UMPK is included also. Multiple arrows denote multiple steps. The broken and dotted arrows denote feed-back inhibition and activation, respectively. The double-lined arrow indicates repression by UMPK of the *carAB* genes, which encode carbamoyl phosphate synthetase. The involvement of the DNA-binding proteins PepA and IHF in this process is indicated.

comparisons indicate that this enzyme belongs to the amino acid kinase protein family (AAKF)[9] (PF00696†), a family in which the two structurally characterized members, carbamate kinase (CK) and *N*-acetyl-L-glutamate kinase (NAGK),[9–11] have a characteristic fold and a homodimeric architecture, and catalyse acyl group phosphorylation. However, UMPK phosphorylates a phosphate group, it is hexameric[4,8,12] and is the subject of feed-back regulation,[4,8,12,13] mediated by UTP (inhibitor) and GTP (activator). These characteristics render prokaryotic UMPK particularly interesting for structural study, specially since other enzymes of the AAKF, including aspartokinase, glutamate-5-kinase,[14,15] *N*-acetyl-L-glutamate synthase,[16] and even the NAGK of most organisms (but not that of *Escherichia coli*, the NAGK for which the structure was reported[17]) are the targets of feed-back regulation, and since some of these enzymes are hexameric.[17,18] Thus, the UMPK structure might provide a broader view of substrate binding and catalysis, and might help understand control and hexameric architecture among the enzymes of the AAKF.

† http://www.sanger.ac.uk/Software/Pfam

An additional interesting characteristic of UMPK is its key role, demonstrated in *E. coli*,[19] in the repression by pyrimidines of the expression of the *carAB* genes, the genes that encode the two subunits of the first enzyme of the pyrimidine biosynthesis pathway, carbamoyl phosphate synthetase. This important gene regulatory role involves also two DNA binding proteins of known structure,[20,21] aminopeptidase A (PepA) and integration host factor (IHF), but UMPK is needed to render the complex sensitive to pyrimidines.[19] In addition to UMPK, another two enzymes of the AAKF, NAGK and glutamate 5-kinase, were reported to be involved in gene regulatory processes,[22,23] but, as with UMPK, the mechanism of the gene controlling functions of these enzymes remains to be elucidated. Thus, information gathered with UMPK might bear on gene regulation by UMPK itself, and by these two other enzymes.

Overall, neither catalysis, feed-back control or gene expression control effected by UMPK are presently understood in structural terms. As a first step towards this understanding, we determine here the crystal structure of *Pyrococcus furiosus* UMPK, ligand-free and complexed with the ATP analogue AMPPNP or with both AMPPNP and UMP, at resolutions of 2.4 Å, 3 Å and 2.55 Å, respectively. Our results with this enzyme, which has a close similarity to bacterial UMPKs in amino acid sequence and in the strict specificity for UMP,[24] clarify substrate binding and catalysis, with involvement of two Mg ions, reveal dramatic differences in homodimer architecture with respect to the other AAKF enzymes of known structure, and demonstrate a hexameric architecture in which three dimers are organized as a trimer of dimers having a hollow nucleus. The structure shows that small changes in intersubunit interactions might alter the active centres of the subunits and influence enzyme activity. This might be the basis for regulation by allosteric effectors. Further, the shape and size of the hexamer suggests ways of interaction with aminopeptidase A (PepA) in the gene regulatory complex involving UMPK.

Results

The polypeptide fold

Three crystal structures, for ligand-free UMPK, for the binary complex of the enzyme with AMPPNP, and for the ternary complex with UMP and AMPPNP, have been determined at resolutions of 2.4 Å, 3 Å and 2.55 Å, respectively (Table 1). The three structures have two monomers in the asymmetric unit, A and B, related by a rotation of 180°, forming a dimer. Not visible in any structure are, the 20 residue N-terminal His_6 tag, and, in the substrate-free enzyme, residues 174–181 (both subunits) and residues 147–155 (subunit B only). Residues 174–181 and 147–155 belong to two substrate-interacting loops, and are clearly visible

Table 1. Data collection, phasing and refinement statistics

		SeMet-substituted UMPK				
		Uncomplexed form used for MAD[a]			UMPK complexed with	
	Wild-type UMPK	Inflection	Peak	Remote	AMPPNP	AMPPNP and UMP
A. *Data collection and phasing statistics*						
Beamline	BM16 (ESRF)		BL1 (BESSY)		BM16 (ESRF)	ID23 (ESRF)
Space group	I23		I23		I23	I23
Unit cell ($a=b=c$) (Å)	144.95		144.71		146.68	144.56
Resolution range (Å)	45.6–2.40 (2.53–2.40)	30.0–2.60 (2.64–2.60)	30.0–2.60 (2.64–2.60)	30.0–2.65 (2.70–2.65)	46.6–3.00 (3.16–3.00)	45.7–2.55 (2.69–2.55)
Wavelength (Å)	0.9795	0.9800	0.9799	0.9649	0.9795	0.9686
Reflections total/unique	148,826/19,936	87,842/15,490	175,408/15,558	792,52/14,569	74,237/10,688	92,2381/16,529
Completeness (%)	100 (100)	98.9 (100)	99.4 (100)	98.4 (100)	100 (100)	100 (100)
I/σ	9.5 (2.0)	33.6 (3.7)	75.3 (9.0)	26.2 (3.1)	7.1 (1.8)	6.8 (1.9)
R_{sym}[b] (%)	6.6 (37.2)	4.6 (35.2)	8.9 (29.7)	5.9 (39.1)	9.4 (41.6)	8.9 (39.3)
FOM[c]			0.531 (0.34)			
B. *Refinement statistics*						
Resolution range (Å)	46–2.4		20–2.6		50–3.0	50–2.55
R-factor (%)[d]	20.59		20.97		21.85	19.32
R_{free}[d] (%)	24.02		26.40		27.91	24.35
No. protein atoms	3286		3268		3452	3446
No. water molecules	94		107		14	54
No. other molecules					2 AMPPNP Two Mg ions	2 AMPPNP 4 Mg ions 2 UMP
rmsd from ideal						
Bond lengths (Å)	0.006		0.006		0.006	0.012
Bond angles (deg.)	1.24		1.26		1.23	1.27
Average *B*-factor (Å^2)						
Protein	40.83		51.90		48.47	38.99
Water	41.27		48.46		33.03	37.43
AMPPNP					62.37	62.71
Magnesium					64.98	56.75
UMP						27.05
Ramachandran plot[e]						
Most favored (%)	93.8		89.8		82.0	90.1
Additionally allowed (%)	5.4		8.9		16.8	9.4
Generously allowed (%)	0.5		0.8		1.3	0.5
Disallowed regions (%)	0.3		0.5		0	0

Values in parentheses are data for the highest-resolution shell.
[a] MAD, phasing method based on the multiple anomalous dispersion of Se.
[b] $R_{sym}=\Sigma|I-\langle I\rangle|/\Sigma I$, where I is the observed intensity and $\langle I\rangle$ is the average intensity of multiple observations of symmetry-related reflections.
[c] FOM, figure of merit calculated with the program SOLVE.[40]
[d] $\sum_h ||F_{obs}|-|F_{calc}||/\sum_h |F_{obs}|$, where $|F_{obs}|$ and $|F_{calc}|$ are observed and calculated structure factors amplitudes for all reflections (*R*-factor) and the reflections applied in the test are R_{free} set (reflections not used in the structure refinement), respectively.
[e] Calculated using PROCHECK[44].

in both complexes having AMPPNP, but one of these loops (loop αF-αG, residues 171–184) adopts different conformations, depending on whether UMP is present or absent. In the two complexes that contain AMPPNP, the definition of the entire polypeptide (residues 1–225) is good (Figure 2(a)) although a few side-chains are not visible (Lys173, Ile175 and Glu176 in subunit B of the AMPPNP–UMP complex; Lys173 in subunit A and Ile175 in subunit B of the complex with AMPPNP). In each structure, both subunits are very similar (rmsd value for superposition of all C^{α} atoms of subunit A with those in subunit B, 0.39–0.73 Å; mean value 0.52 Å), although ∼10% of the residues exhibit different side-chain conformations in the two subunits. Superposition of the individual subunits from the different structures yield rmsd values for all C^{α} atoms ranging, for subunit A, between 0.46 Å and 0.79 Å (mean value 0.65 Å), and, for subunit B, between 0.50 Å and 0.97 Å (mean 0.75 Å), without significant changes in the position of loops or secondary structure elements, except in the case of the αF-αG loop (see below). Unless indicated, the complex with AMPPNP and UMP will be described, referring to the other structures when pertinent.

Each subunit resembles a nest (dimensions, 32 Å × 35 Å × 45 Å) holding the substrates (Figure 2(b)), in which the bottom is formed (Figure 2 (c) and (d)) by an open, extended, mainly parallel eight-stranded β sheet (strand order, β3, β4, β2, β1, β5, β7, β8 and β6; where β8 and β6 are antiparallel to the other strands) sandwiched between one layer of three α helices (helices C, A and H) and another layer of four helices (helices D, E, G and F). The rim of the opening around the substrates is formed by elements protruding from the β sheet C-edge: the β5′-β6′ hairpin loop (residues 141–159), the αB-αC helix hairpin (residues 45–73), and the loops β4-αE (residues 112–118), αF-αG (residues 171–184), β1-αA (residues 9–17) and β7-αH (residues 203–208). The β1-αA and

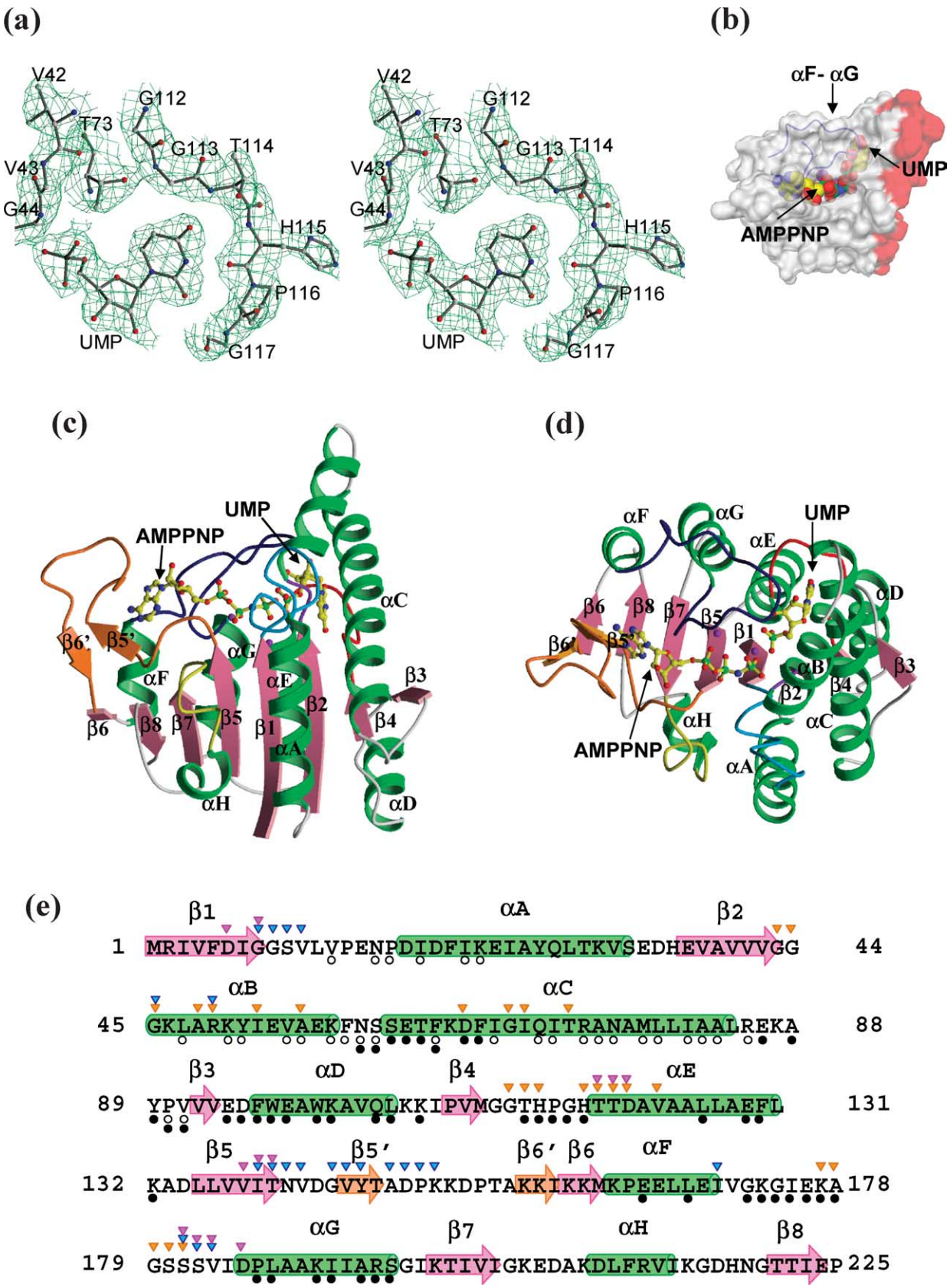


Figure 2. The *Pyrococcus furiosus* UMPK subunit structure. (a) Final ($2F_{obs} - F_{calc}$) electron density map, contoured at the 1.0 σ level, around the bound UMP molecule. The corresponding molecular model is superimposed. Protein residues are indicated in the single-letter code. (b) Semitransparent surface representation of the subunit containing Mg_2AMPPNP and UMP. The surface involved in dimer formation is coloured reddish and the backbone trace of the mobile αF-αG loop is shown in blue. (c) and (d) Ribbon representation of two views of the UMPK subunit with bound

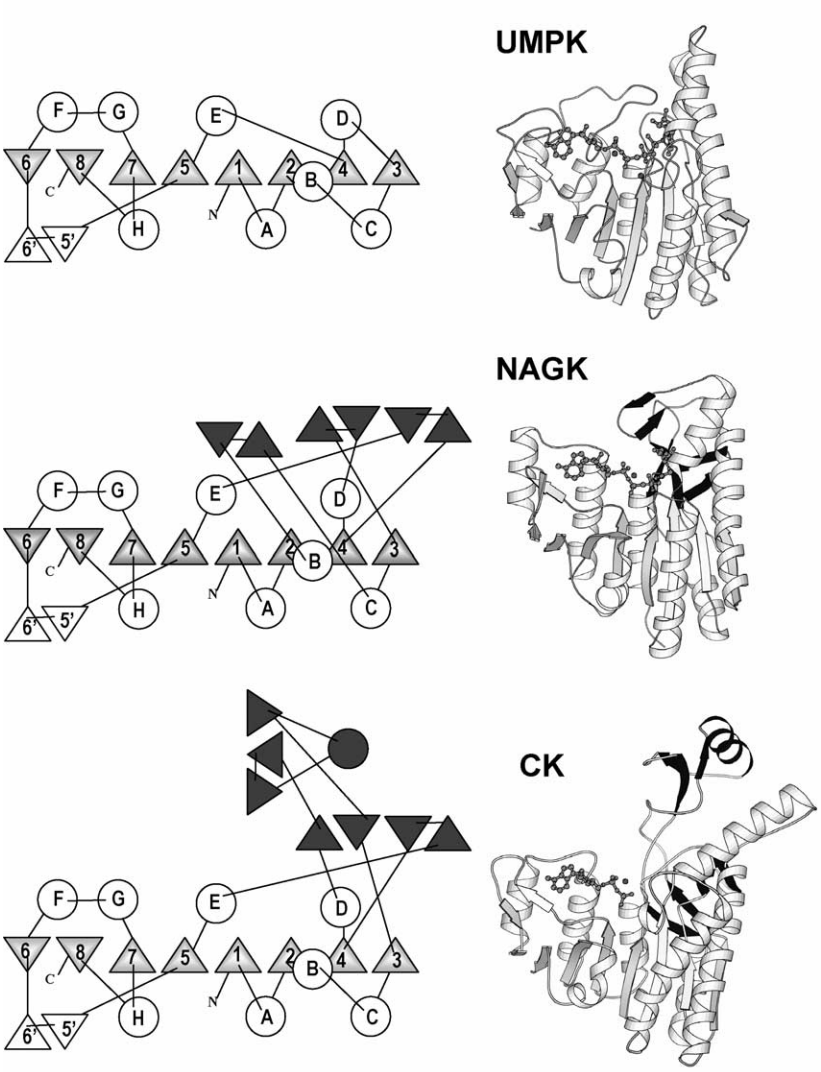


Figure 3. Scheme of the topology of secondary structure elements (left) and ribbon representation (right) of the subunits of *P. furiosus* UMPK with bound Mg_2AMPPNP and UMP, of *E. coli* NAGK with MgAMPPNP and acetylglutamate,[9] and of *P. furiosus* CK with MgADP.[11] In the scheme of the topology, β strands and α helices are represented as triangles and circles, respectively, the strands of the main β sheet are shadowed, elements of NAGK or of CK that are missing from UMPK are filled in black, and the designation of the other secondary structure elements is as in UMPK. In the ribbon representation of the subunits, elements missing from UMPK are also coloured black.

β7-αH loops include one α helical turn at the N terminus.

The subunit fold and the topology conform to those of the other two enzymes of the AAKF for which the structure was determined,[9–11] acetylglutamate kinase and carbamate kinase (Figure 3, and see Discussion). By analogy with these, the polypeptide can be divided into an N-terminal lobe (residues 1–132) formed by the four N-terminal β strands (β1-β4) and associated α helices (αA-αE), which binds the UMP and forms all the intersubunit interactions in the homodimer; and a C-terminal lobe (residues 133–225) formed by the remaining elements, which binds the ADP moiety of the ATP molecule and provides the mobile loop that covers the active centre in the ternary complex.

UMP binding

UMP binds in the N-lobe of both subunits (Figures 2(a)–(d) and 4(a)), in a bent conformation, with the uracil in *anti* orientation and with the phosphate group emerging on the same side of the ribose plane as the uracil, sitting over the C-edge of the main β sheet in a pocket formed between the αB-αC hairpin and the β4-αE and αF-αG loops. The phosphate group is encircled by (a) the β2-αB connection, with which it is hydrogen bonded *via*

Mg_2AMPPNP and UMP. The central β-sheet is coloured magenta, α-helices are green and loops are grey, except the loops β1-αA, β4-αE, β5-β6 (including the β5′-β6′ hairpin), αF-αG and β7-αH, which are coloured light blue, red, orange, dark blue and yellow, respectively. (e) Correspondence between the amino acid sequence (in the single-letter code) and the secondary structure. Superimposed green cylinders denote α helices, and arrows denote β strands, coloured pink for the main sheet and orange for the β5′-β6′ hairpin. Residues having in the ternary complex decreased accessibility upon the binding of UMP, AMPPNP or Mg are indicated with orange, blue or violet triangles, respectively. White and black circles indicate residues that exhibit decreased accessibility upon homodimer and hexamer formation, respectively.

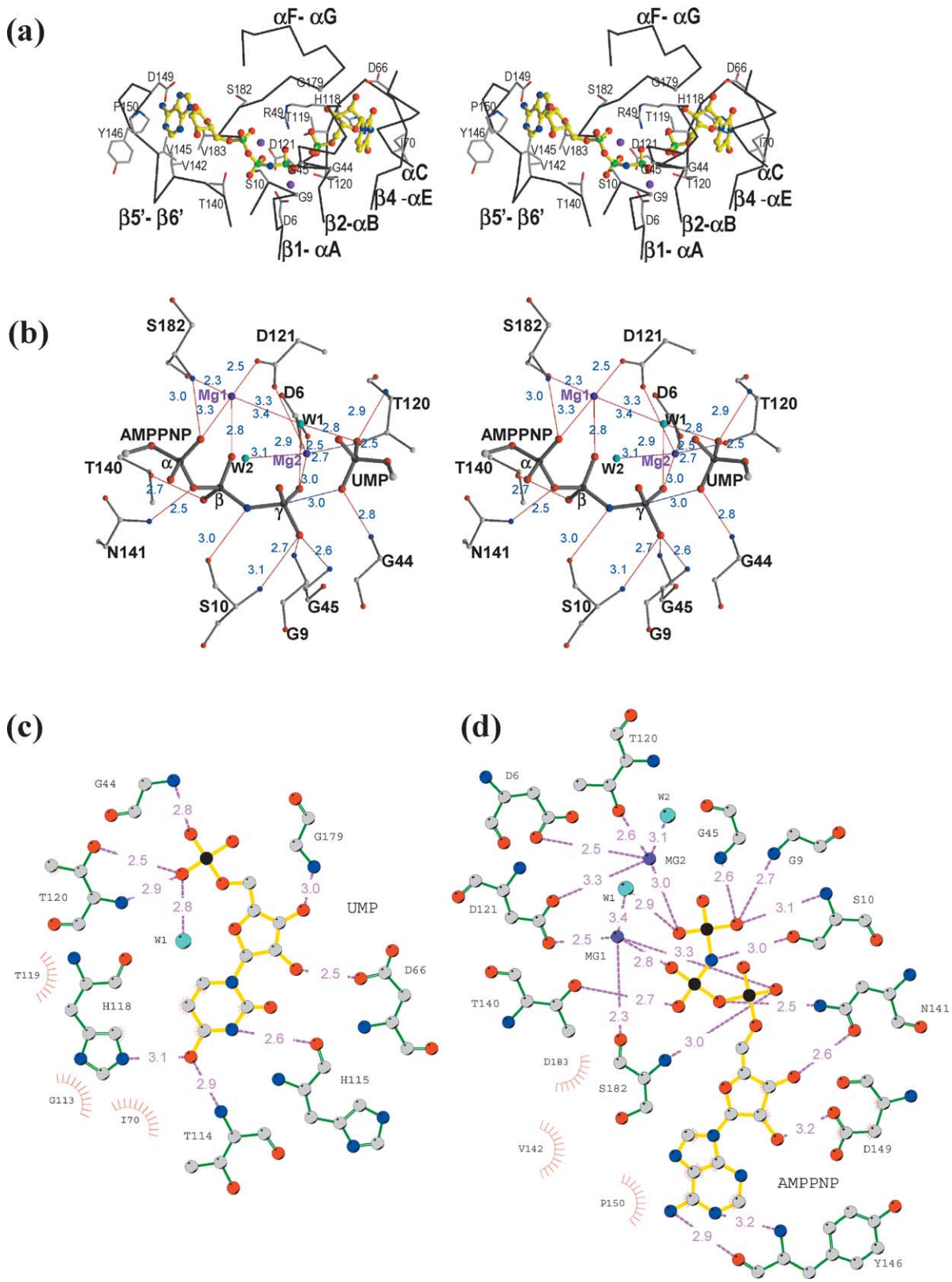


Figure 4. Substrate binding in UMPK. (a) Stereo view of the C^{α} trace of the substrate-binding sites, with bound Mg_2AMPPNP and UMP coloured. Amino acid side-chains are also in colour, in thinner trace. (b) Stereoscopic detailed representation of the phosphoryl group transfer site in the complex with Mg_2AMPPNP and UMP. Mg ions and water molecules are drawn as purple and cyan spheres, respectively. Nearby protein residues are shown in thinner trace. Hydrogen bonds and coordination bonds with Mg are shown as red lines, indicating the interatomic distances (in Å). The interatomic distance between the attacking O atom of UMP and the γ-P atom is represented with a blue line. (c) and (d) Plots of the interactions between the protein and (c) UMP or (d) Mg_2AMPPNP in the ternary complex. The letter W denotes a water molecule. Distances are in Å. Parts (c) and (d) were drawn with LIGPLOT.[50]

the N atom of Gly44; (b) Thr119 and Thr120 of the β4-αE loop, making two hydrogen bonds with the N and O main-chain atoms of Thr120; (c) Gly179 and Ser180 of the αF-αG mobile loop; and (d) the side-chain of Arg49. In this way, the UMP phosphate group is fixed precisely in position for attack on the nearby ATP γ-phosphate group, with the attacking O atom at only 3 Å from the attacked γP atom of the purine nucleotide.

The exquisite specificity for UMP of *P. furiosus* UMPK[24] that is a characteristic trait also of bacterial UMPKs,[1,4] but not of the eukaryotic CMP/UMP kinases,[3] is accounted for by the binding of the uracil base, which sits over the side-chain of Ile70 and is enveloped by the β4-αE loop (residues 112–118, Figures 2(a) and 4(a)), making three hydrogen bonds that cannot be formed with the usual amino form of cytosine: between the O atom at C4 of uracil and the ε2 N and αN atoms of His118 and Thr114, respectively; and between the N3 atom of uracil and the αO atom of His115 (Figure 4(c)).

In both subunits of the ternary complex, the sites for the AMPPNP γ-phosphate group and for the entire UMP molecule are isolated from the outside by a lid formed by the mobile αF-αG loop, which sits extended over the bound substrates (Figure 2(b)). In contrast, in the binary complex having AMPPNP, the empty UMP site is exposed and the αF-αG loop residues 172–181 are displaced outwards form the active centre (Figure 5). Since the αF-αG loop is disordered and could not be traced unless AMPPNP was bound, the purine nucleotide fixes this loop in a retracted conformation, and the additional presence of UMP favours a second, extended conformation covering the phosphoryl transfer site. The ribose of UMP may play a key role in triggering this extended loop conformation, since it makes a hydrogen bond between its 3′ hydroxyl group and Gly179 (Figure 4(c)), anchoring the loop on the third turn of αC, on which the ribose lies.

Binding of ATP and of two Mg ions

In the binary complex with AMPPNP and in the ternary complex with AMPPNP and UMP, AMPPNP binds extended in each of the two subunits (Figures 2 (b)–(e) and 4 (a), (b) and (d)), in a groove that follows the direction of the central β sheet over the sheet C-edge. One side of the groove is formed by the β5′-β6′ hairpin loop and the β1-αA loop, and the other side is formed by the αF-αG and β4-αE loops. The adenine moiety, in the *syn* configuration, is sandwiched between the αF-αG loop and the base-contouring β5′-β6′ hairpin loop (Figure 4(a)), and points with the free edge of the hexameric ring towards the groove bottom, forming hydrogen bonds between its N1 and 6-amino groups and the main-chain N and O atoms of Tyr146, respectively (Figure 4(d)). These interactions, and the lack of extra space to accommodate the C2 amino group of guanine, account for the high specificity of UMPK for ATP reflected in our failure to observe enzyme activity when the 1 mM ATP used in the standard assay[24] was replaced by either 1 mM GTP or 1 mM UTP (detection limit, 2% of the activity with ATP, data not illustrated).

The ribose sits over the side-chain of Val142, (Figure 4(a)), and the extended polyphosphate chain crosses transversally over the C-ends of β5 and β1 in the direction of the UMP phosphate group (Figures 2(c) and (d) and 4(a) and (b)). A rich set of interactions is established (Figure 4(b) and (d)) between the γ-phosphate group and the protein, particularly hydrogen bonds with small residues, mainly from the β1-αA turn, fixing this phosphate group properly for attack, near the phosphate group of UMP. One Mg ion (Mg2) is observed only in the ternary complex with AMPPNP and UMP, and is coordinated (Figure 4(b) and (d)) with one non-bridging O atom of the γ-phosphate group, neutralizing the negative charge on this phosphate group and bridging this phosphate group to the protein, since the γ-carboxylate groups of Asp6 and Asp121, and the OH group of Thr120 also belong to the Mg coordination sphere. Another Mg ion (Mg1) is found in the AMPPNP complexes, irrespective of the presence or absence of UMP, and is coordinated with a non-bridging O atom of the AMPPNP β-phosphate group and with the main-chain O atom of Ser182 and the γ-COO^- of Asp121 (Figure 4(b) and (d)). Thus, the γ-carboxylate group of Asp121 links both Mg ions and may have an important organizing role in the active centre.

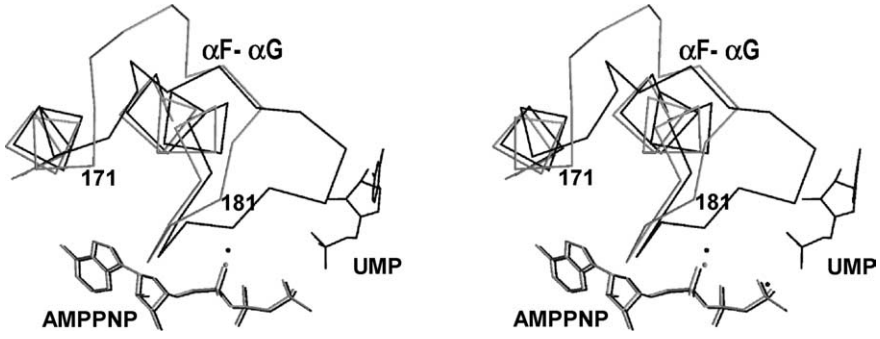


Figure 5. Movement of the αF-αG loop. Stereo view of the superposition of the C^{α} trace of the UMPK complexes with MgAMPPNP (lighter) or with both Mg_2AMPPNP and UMP (darker). The Mg ions are shown as dots. Only the substrates and helices F, G and the αF-αG loop are represented. The numerals denote residue numbers.

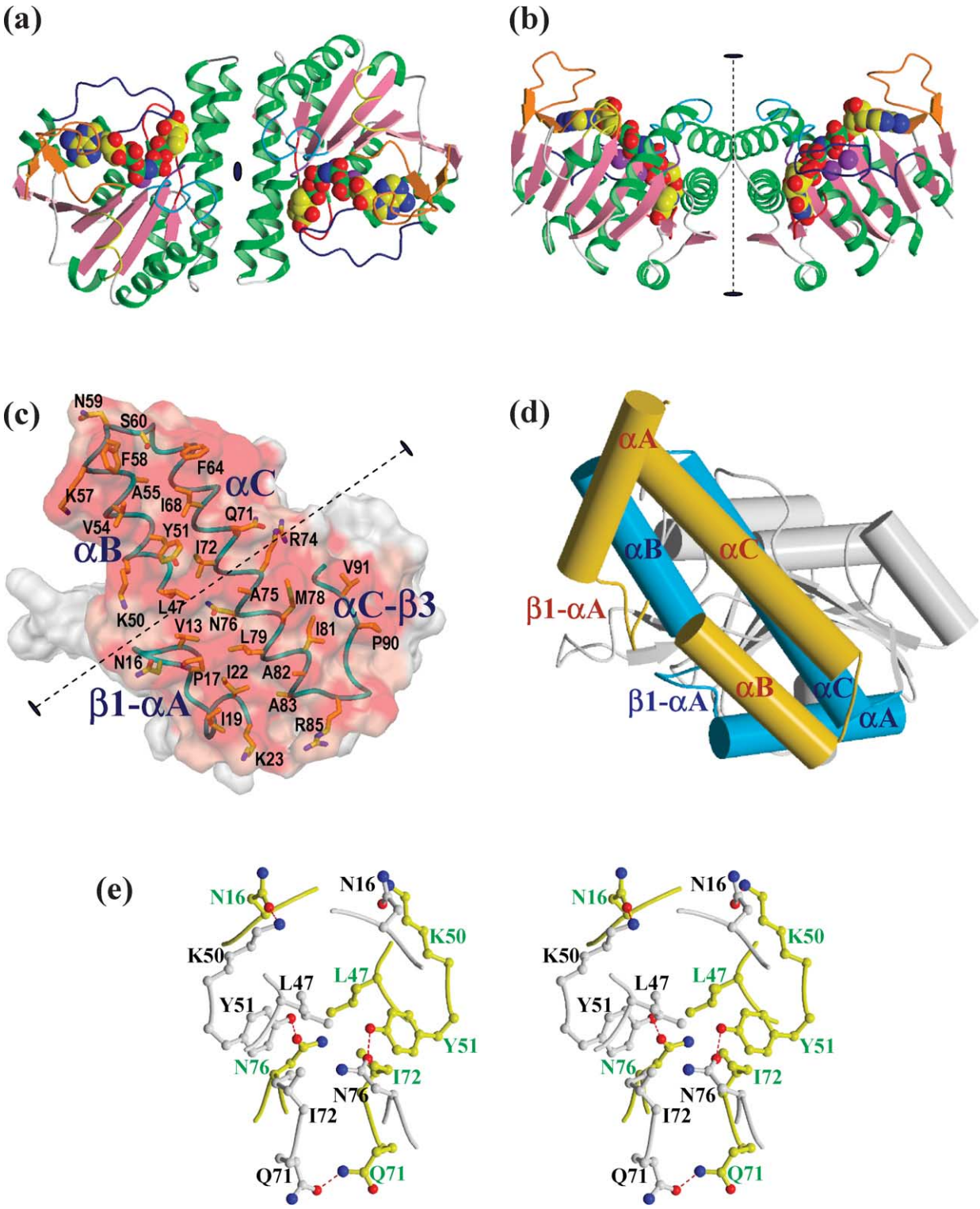


Figure 6. The UMPK dimer. (a) and (b) Ribbon diagrams of two perpendicular views of the UMPK dimer drawn as in Figure 2(c) and (d), with the substrates in space-filling representation. The 2-fold molecular symmetry axis of the dimer is perpendicular to the page (black ellipse in (a)) or parallel with the page(broken line ended in two black ellipses, in (b)). (c) Semitransparent representation of the dimerization surface of one subunit, viewed from the other subunit, with its non-polar or polar atoms coloured reddish or pink, respectively. Secondary structure elements involved in intersubunit interactions and important interfacial residues are represented and identified. The 2-fold molecular axis of the dimer is represented as a broken line. (d) Cartoon representation of the relative disposition of the main secondary structure elements of the two subunits that define the intersubunit surface, in the same orientation as in (c). Helices are represented as cylinders. The subunit that is farther away from the viewer is shown grey, except the elements involved in the intersubunit interactions, which are blue. Of the subunit that is closer to the viewer, only the elements that interact across the dimeric junction are shown, coloured yellow. (e) Stereoscopic detailed view of intersubunit contacts close to the 2-fold axis (which runs centrally and vertically). Each subunit is in one colour.

The homodimer

The two subunits found in the asymmetric unit are related by a 180° rotation and form a dimer of ~72 Å × 42 Å × 45 Å (Figure 6(a) and (b)). In projection, the dimer has the appearance of the letter W, with one valley between the two subunits on one side, whereas the opposite side, where the substrates are bound, is continuous, slightly convex and approximately rectangular, with a protuberance emerging on opposite ends of this surface and corresponding to the β5′-β6′ hairpin loop. Extensive, mainly hydrophobic contacts between the two subunits are mediated by a flat and continuous intersubunit surface of 1324 Å^2 (determined using a probe radius of 1.4 Å), representing 12% of the subunit surface (Figure 6(c)). The interaction surface is grossly rectangular, and the 2-fold molecular symmetry axis of the dimer crosses perpendicularly the long sides of the rectangle at the middle of these sides. In the dimer, one half of the intersubunit junction (Figure 6(d)) is formed by the mutual contacts of the C helices from both subunits, which run antiparallel to each other. The other half of the junction consists of helix B from subunit A, the β1-αA junction and the beginning of αA from subunit B, and of the same elements of the converse subunits. In this way, the intersubunit junction is composed of two symmetric halves, where each half consists of a four-helix bundle: helix B and the N-terminal half of helix C, both from one subunit, and the C-terminal half of helix C and helix A from the other subunit, although helix A is quite displaced (Figure 6(d)). The two helix bundles are connected across the symmetry axis by (Figure 6(e)), the mutual interactions of the side-chains of Gln71 from both C-helices, and by the interactions between the side-chains of Leu47 from the beginning of both B-helices. The only polar bond within the secluded surface of interaction is that between the phenolic OH group of Tyr51 and the amidic nitrogen atom of Asn76 (Figure 6(e)).

The hexamer

Application of the 3-fold crystallographic symmetry reveals that the enzyme is organized as a hexamer, having the architecture of a trimer of dimers (Figure 7(a) and (b)). The three dimers, with their long axes tilted approximately 45° relative to the 3-fold axis (Figure 7(a)), have the substrate sites looking outwards, and thus the A subunit of one dimer lies on the B subunit of the adjacent dimer and the valleys of the three dimers merge into a central cavity of ~30 Å diameter (Figure 7(c)). This hollow nucleus of the hexamer is limited laterally by the interactions between the A and B subunits of the adjacent dimers, and is limited on the poles around the 3-fold axis by interactions between the three A subunits or between the three B subunits (Figure 7(d)). These interactions at the poles are mediated by helix D, and by the preceding β3-αD and αC-β3 connections (Figure 7(b) and (d)) and include a sealing, circularly closed, ion pair network around the 3-fold axis, formed by the side-chains of Lys101 from the three subunits, which point towards the 3-fold axis and are flanked by the side-chains of Glu98 and Tyr89, so that the positive charge on the lysine residues is neutralized by the negative charge on the nearby glutamate residues.

The lateral interdimeric contacts between the A subunit of one dimer and the B subunit of the adjacent dimer are extensive, representing 1100 Å^2, or 10% of the surface of each subunit. They involve (Figure 2(e)) helices C, E, G and F, and the associated loops β4-αE and αF-αG of each subunit (Figure 7(e), coloured elements). These elements of the two subunits that form an interdimeric junction are related by a 2-fold symmetry axis. This axis is the prolongation of the molecular symmetry axis of the dimer that is not involved in this particular interdimeric junction (Figure 7(b)). The axis (Figure 7(f)) runs in a direction approximately perpendicular to that of the axes of helices F, G and E, passing between the two walls formed by these four-helix layers of both subunits and, more specifically, between the C termini of helices F and between the N termini of helices G and E. The interdimeric interactions are largely hydrophobic (Figure 7(f)), although they include one ion pair between Asp66 and Arg194, and hydrogen bonds between the ξ N of Lys190 and the O atoms of Gly172 and Gly174, and between one η N of Arg194 and the main-chain O atom of Glu176. The imidazole rings of the residues of His115 and the side-chains of Leu187 of both subunits surround the 2-fold axis and interact mutually. The extensive interdimeric interactions provide a stable architecture for the hexamer, and may stabilize the conformation of the loops involved in the interactions between dimers, particularly those of the β4-αE loop.

Discussion

The present structures confirm that UMPK does not resemble adenylate kinase or the UMP/CMP kinases of eukaryotes, and show that its polypeptide fold conforms to the amino acid kinase fold.[9] Thus, rmsd values of 2.1 Å are obtained for the superposition of 151 and 154 C$^\alpha$ atoms of UMPK with C$^\alpha$ atoms of *P. furiosus* CK[11] and *E. coli* NAGK,[9] two enzymes of the AAKF for which the structures are known.[9–11] All the α helices and the strands of the central β sheet found in CK and NAGK are preserved in UMPK (Figure 3), although strand β3 (secondary structure elements of CK and NAGK are designated here as in UMPK, according to Figure 3) has only two residues and plays no interfacial role in UMPK, whereas in NAGK and CK it has six residues and plays a central role in the interactions for making the dimer. Nevertheless, the fold of the N-lobe is somewhat simpler in UMPK than in NAGK and CK (Figure 3), since the four-stranded secondary β sheet found in these other enzymes is

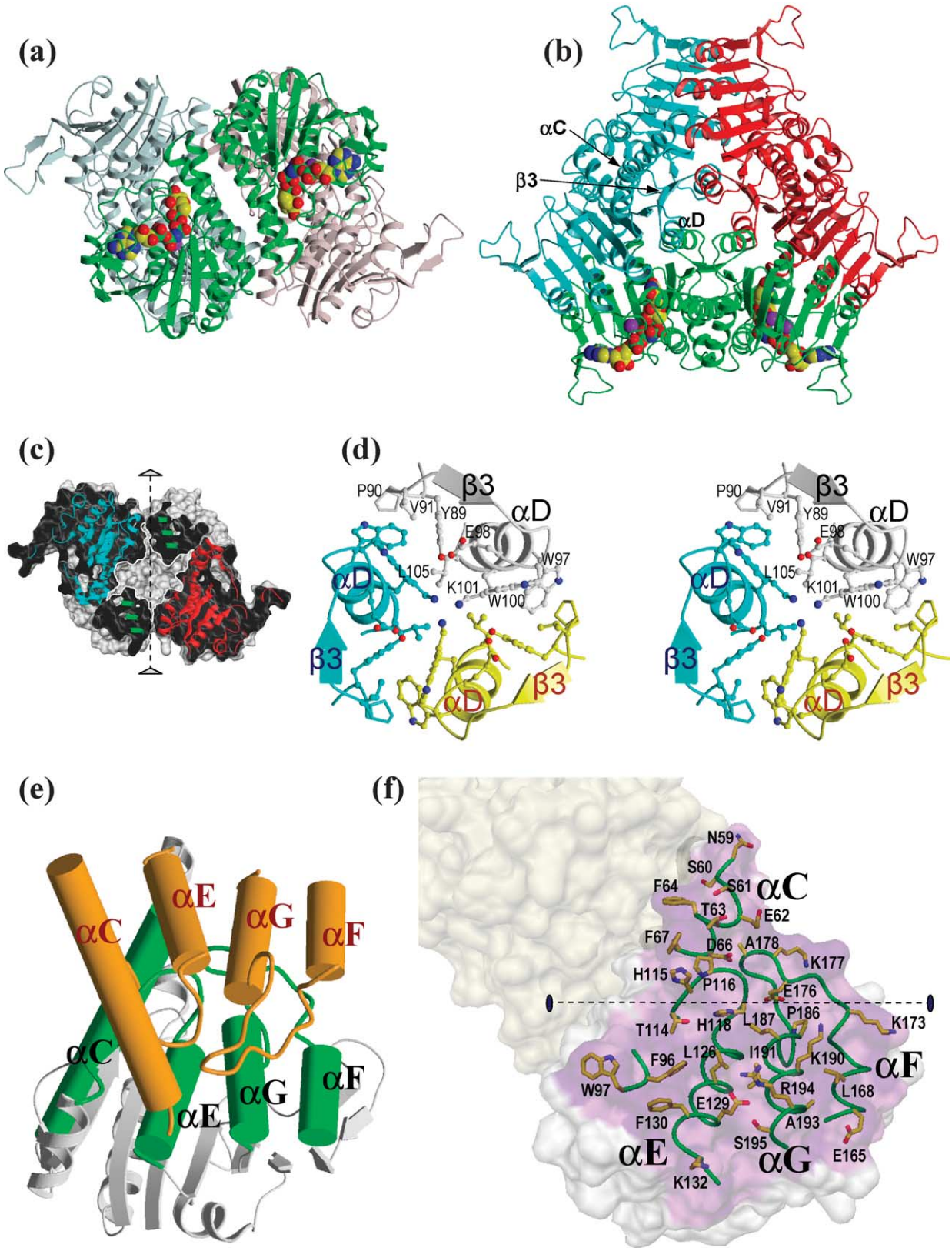


Figure 7. The UMPK hexamer. (a) and (b) Ribbon diagrams of two perpendicular views of the UMPK hexamer, with (a) the 3-fold axis vertical or (b) perpendicular to the page. For clarity, each homodimer is coloured differently, and substrates have been represented (in space-filling view) in only one dimer. (c) Cross-section through the surface of UMPK, oriented as in (a), showing the central cavity. The 3-fold axis is shown. (d) Stereo view showing the interactions between the A subunits around the 3-fold axis. Each subunit is shown in one colour. (e) Cartoon representation of the relative disposition of the elements that define the surface between the A and B subunits of adjacent dimers. The subunit that is farther away from the viewer is shown grey, except the elements involved in the intersubunit interactions, which

missing from UMPK, as well as the distant subdomain (three β strands and one α helix) found in CK and the β hairpin between αB and αC found in NAGK. These missing elements are in CK and NAGK at or near the site for the substrate that is phosphorylated (Figure 3) but, since their lack in UMPK does not hamper either the specificity[24] or the affinity of this enzyme for the substrate that is phosphorylated ($K_m^{UMP} = 50$ μM , data not shown; compare for example with $K_m^{NAG} =$ 0.2 mM for *E. coli* NAGK[14]), these elements appear not to be essential for substrate binding and may play other roles in CK and NAGK.

The folding of the C-lobe of UMPK (Figure 3) conforms closely to the fold observed in NAGK and CK, but the αF-αG loop found in this lobe is larger in UMPK and is mobile, covering the active site when both substrates are bound. The much shorter αF-αG loops of NAGK or CK do not shield the active site.[9–11,25] In fact, amino acid sequence alignments[24] among UMPKs (further alignments not shown) indicate that the long αF-αG loop (typically 14 residues long) is found only in the UMPK of the archaea, whereas in bacteria, including the hyperthermophilic bacteria *Thermotoga maritima*, this loop only has four residues. Thus, active centre shielding by this loop appears not to be an essential requirement of substrate binding, catalysis or transition-state protection among UMPKs, and therefore the specific role of this loop in archaea requires clarification.

Substrate binding also appears quite similar in CK, NAGK[9,11] and UMPK. The purine nucleotide binds in the three cases (Figure 3) in the C-lobe, in an extended conformation, with one Mg ion sitting between the α and β phosphate groups, with the polyphosphate group pointing towards the molecular 2-fold axis of the dimer, and with the adenine base looking in the opposite direction and being sandwiched between the β5′-β6′ hairpin and αF-αG loop. Similarly, UMP binding in UMPK and acetylglutamate binding in NAGK[9,25] take place at equivalent sites in the N-lobe (Figure 8(a)), with the group that is to be phosphorylated occupying essentially identical positions in both enzymes, between the C terminus of β2 and the N terminus of αE. The ribose moiety of UMP and the C^{α} and associated substituents of acetylglutamate bind near the β4-αE loop, in an equivalent region that is limited towards the 2-fold axis by helix C. The uracil base is the only component of UMP that sits in a place without correspondence in the acetylglutamate site of NAGK. In this last enzyme, this part of the site is obliterated by the β4-αE loop, which in NAGK[9] is much more extended and has an orientation different from that in UMPK (Figure 8(a)). These observations, taken together with the structural data on CK[11] and with the conclusions of structure-directed or modelling-directed mutagenesis studies with NAGK and aspartokinase,[14] support the view that the substrate that is phosphorylated sits in the N-lobe of all the enzymes of the AAKF. Therefore, both substrates may be bound in the same way by the different enzymes of this family.

In the structure of the complex of UMPK with AMPPNP and UMP the attacking O atom of UMP (O_{UMP}) is clearly identified (Figure 4(b)). The angle formed between this atom, the Pγ of AMPPNP and the bridging N atom of AMPPNP ($N_{\beta\gamma}$) is 162–167° (depending on the subunit), close to the 180° expected for in-line attack on phosphorus with formation of a trigonal bipyramidal transition state.[26,27] The distance O_{UMP}–Pγ is 3 Å, 1.9 Å too short for a dissociative mechanism, corresponding to fractional bond formation between Pγ and O_{UMP} of 0.08.[27] Assuming that the P atom in the bipyramidal transition-state is localized at the midpoint between the positions observed in the structure of the entering and leaving atoms, the bond number for the transition state would be 0.10, and thus the mechanism would be, in the terminology of Mildvan,[27] 10% associative.

An interesting observation made in the present studies is that in the ternary complex with AMPPNP and UMP there are two Mg ions associated with the polyphosphate chain of the triphosphate, one complexed to the β phosphate group and the other to the γ-phosphate group (Figure 4(b)). In the corresponding ternary complex of NAGK,[9] only one Mg ion is observed, at essentially the same position as the Mg ion that sits in UMPK next to the β phosphate group. The extra Mg ion found in UMPK near the γ-phosphate group occupies a position corresponding to that of the positively charged amino group of Lys8 in NAGK. This lysine residue is found in NAGK, aspartokinase, glutamate 5-kinase and the bacterial UMPKs,[9] and was demonstrated in NAGK and aspartokinase to be involved in the catalysis of phosphoryl group transfer.[14,25] In CK, a lysine residue emerging from a different site has its ξ-amino group at this position and possibly plays the same role in catalysis.[9] Thus, since *P. furiosus* UMPK has no equivalent lysine residue, but it has a Mg ion at this site, this Mg ion must play the catalytic role of Lys8 of NAGK, helping stabilize the equatorial negative charge in the pentavalent

are green. Of the subunit that is closer to the viewer, only the elements that interact across the junction are shown, coloured orange. (f) Surface representation of a dimer, in the same orientation as in (e), with the interdimeric interaction surface between A and B subunits in the hexamer made semitransparent in one subunit. The other subunit of the dimer is coloured yellowish. Polar and non-polar atoms on the represented interaction surface are coloured in light or dark violet, respectively. Secondary structure elements involved in intersubunit interactions and important residues of the interface are represented and identified. The 2-fold axis at the interdimeric junction is represented with a broken line.

phosphorus transition state. Curiously, in contrast to the UMPK from *P. furiosus* and from most archaea, bacterial UMPKs do have[24] the lysine residue (further alignments not illustrated) and thus they may use this amino acid residue, and not a Mg ion, for catalysis of phosphoryl group transfer. Anyway, the role of the γ-phosphate group-complexed Mg in the mechanism of archaeal UMPK would differ importantly from the role of the second Mg ion found in other phosphoryl transferring enzymes that use two divalent metal ions in the reaction,[28,29] since in these enzymes one

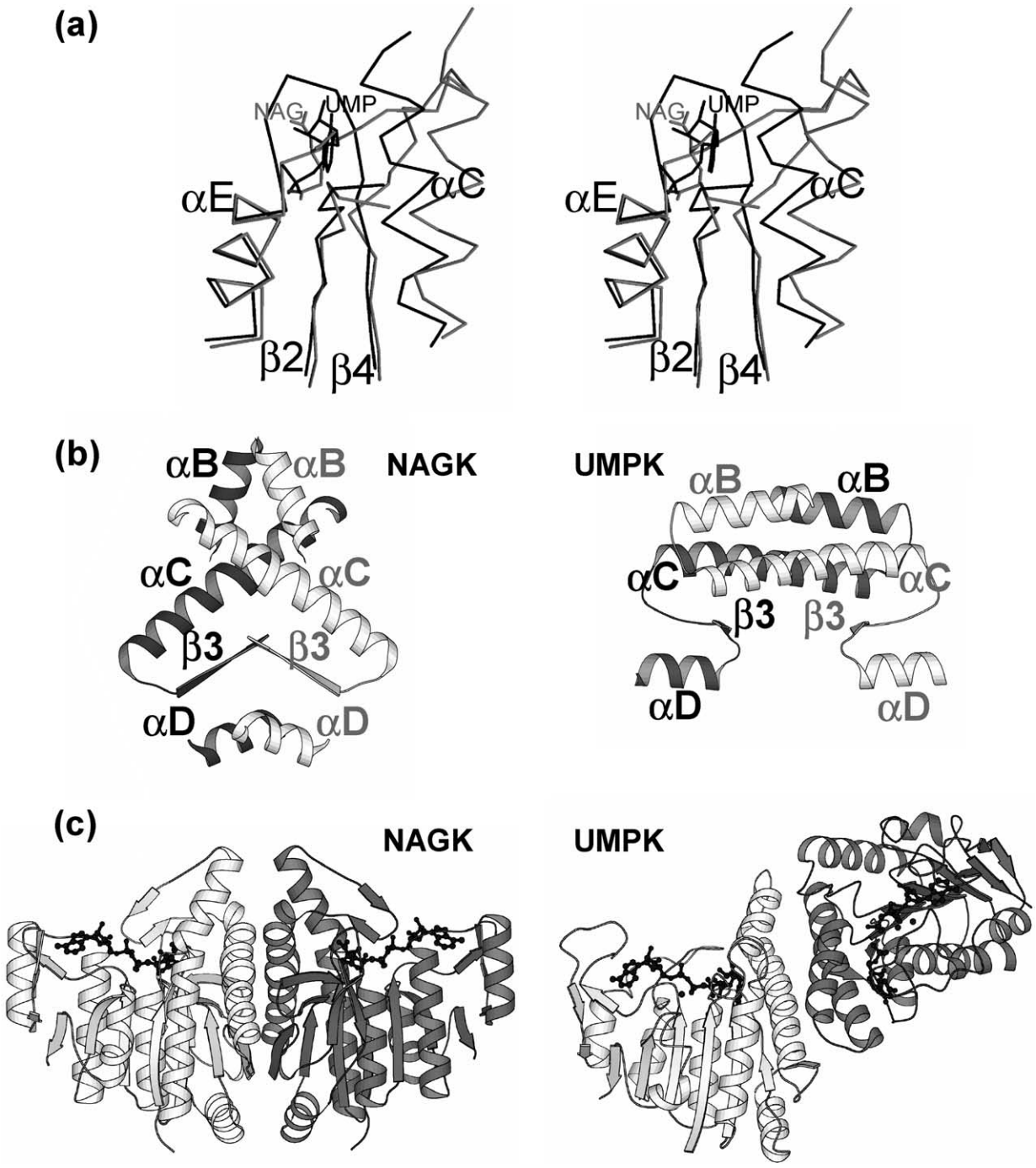


Figure 8. Comparison between *E. coli* NAGK and *P. furiosus* UMPK. (a) Superposition in stereo view of NAG or UMP and surrounding protein elements of NAGK (lighter) and UMPK (darker) to show the similar location in both enzymes of the site for the substrate that is phosphorylated. Substrates and the C^{α} trace of the surrounding protein chains are shown in bond representation. (b) Elements of the two subunits involved in dimer formation, to show their different relative orientations in the two enzymes. (c) Ribbon representation of the homodimers of NAGK and UMPK, to highlight the drastically different dimer architectures. The subunit on the left (shown lighter) has been fixed in the same orientation, in both enzymes.

Mg ion appears to stabilize negative charge development in the leaving group (a role that would be played also by the Mg ion that is complexed to the β-phosphate group of ATP in UMPKs, Figure 4(b)), but the second Mg ion activates a hydroxyl group to render it more acidic and thus a better nucleophile. This cannot be the function of the second Mg ion in UMPK, since in this enzyme the attacking group (the phosphate group of UMP) is already acidic and would not be expected to require activation and, in addition, it does not interact with the Mg ion (Figure 4(b)).

Other catalytic elements identified in UMPK that withdraw negative charge from the phosphoryl group that is transferred are (Figure 4(b)): (1) the donor hydrogen bonds between Gly9, Ser10 and Gly45, and one non-bridging O atom of the γ-phosphate group; (2) the positive charge on the N-end of the short α-helical segment of the β1-αA turn, since this turn is close to the γ-phosphate group (Figure 2(c)); and (3) the nearby (3.6 Å) positive charge on the side-chain of the strictly conserved Arg49 (Figure 4(a)). The catalytic role of Arg49 is supported by the observation of a two orders of magnitude decrease in the V_{max} of *E. coli* UMPK when the corresponding Arg62 was mutated to histidine.[30] The network of donor hydrogen bonds with the γ-phosphate group is observed also in NAGK.[25] In this enzyme, the N-ends of helices B and E provide additional positive charge around the phosphate group,[9] but the somewhat different orientations of these two helices in UMPK precludes their participation in abstracting negative charge from the transition state during phosphoryl group transfer.

A radical difference between the structure of UMPK and those of NAGK and CK is the way in which the subunits associate in the homodimer. In CK and NAGK the central β sheet continues across the dimerization surface through the mutual contacts between strand β3 of the two subunits,[9–11] leading to the formation of a continuous 16-stranded molecular β sheet that spans the dimer from end to end as its backbone. Helices αC and αD, which run parallel with β3 and flank it on both sides, also participate in the interactions. Thus, these three secondary structure elements form with the same elements of the other subunit a cross-grid with a crossover angle of approximately 110° (Figure 8(b), left). In contrast, in UMPK there is no continuity between the central β sheets of the two subunits, because in this enzyme the C helices across the homodimer junction interact mutually along their entire length, running antiparallel to each other, and forming an angle between their axes of approximately 190° (Figure 8(b), right). The differences in dimer architecture can be highlighted by placing one subunit of UMPK in the same orientation as in NAGK (Figure 8(c)): the other UMPK subunit is oriented very differently from that in NAGK, because of rotation of this subunit around an axis crossing perpendicularly the inter-subunit interface at the crossover point between the C helices, until both C helices across the junction are nearly totally antiparallel (Figure 8(b), right). Nevertheless, the same surface of the N-lobe is used for dimer formation in UMPK and in NAGK and CK. There is no obvious functional reason related to substrate binding or to catalysis for this novel form of organization of the UMPK homodimer, since, as with the CK or NAGK homodimers, there is no contact between the substrate sites in the two subunits of the homodimer and, in fact, the substrate kinetics is hyperbolic for these three enzymes (data not shown for *P. furiosus* UMPK).[4,10–11,14] As postulated previously for NAGK and CK, dimer formation may simply be related to the need to provide structural stabilization of the active centres by the extensive interactions across the intersubunit junction, particularly since the UMP site is anchored on this junction.

Another novel and highly spectacular trait of the UMPK structure is its trimer of dimers hexameric architecture. A high degree of oligomerization appears to increase the thermal stability of enzymes,[31] but formation of hexamers is not a peculiarity of the hyperthermophilic *P. furiosus* UMPK, since mesophilic UMPKs are also hexameric.[4,8,12] Hexamer formation appears to be essential for activity, since gel-filtration experiments with *E. coli* UMPK revealed a substantial low molecular mass fraction, consistent with monomers or dimers, which exhibited no activity.[8,32] The presence of low molecular mass forms suggests that the stability of the hexamer is limited, as might be predicted from its hollow nature. It is reasonable to assume that hexamer formation is needed for activity because of the central participation of the β4-αE loop both in the interdimeric junctions (Figure 7(e) and (f)) within the hexamer, and in the active centre of the enzyme (Figure 4(a)). In this way, the mutual interactions between adjacent dimers in the hexamer should help position this loop correctly for activity. This may provide an alternative to the lack of other stabilizing elements that are present in NAGK and CK but not in UMPK, since in the other two enzymes the β4-αE (UMPK denomination) loop belongs to a four-element secondary β sheet (Figure 3(b) and (c)) that may provide it with structural stability.[9–11]

The hexameric structure may be essential for the *carAB* gene regulatory activity reported for *E. coli* UMPK.[19] This activity requires PepA, another hexameric protein organized as a dimer of trimers,[20] which binds DNA.[33] A possible interaction between PepA and UMPK within the *carAB* regulatory complex was suggested on the basis of the observations with PepA mutants.[33] The results of yeast two-hybrid analyses support this proposal.[34] Figure 9 compares the surfaces and their charge potentials for *E. coli* PepA and for the UMPK from the same organism, modeled[35] on the basis of the structure of *P. furiosus* UMPK and of the reported[24] sequence alignment between the *P. furiosus* and *E. coli* enzymes. The view along the 3-fold axis of UMPK and PepA shows that the dimension of the

two hexamers would be appropriate for their interaction with their 3-fold axes aligned. Surface potentials appear nicely complementary. Three symmetric, approximately flat platforms represented by the C-domains of the three subunits around the 3-fold axis of PepA are suggestive of being appropriate for interaction with UMPK. The *E. coli* UMPK mutation Ala94Glu (indicated with circles in Figure 9), which abolishes UMPK control of *carAB* expression without impairing UMPK activity,[19] affects a residue at the end of helix C, a partially superficial residue at the proposed surface of interaction with PepA according to the present model. This coarse model for the UMPK–PepA complex may be used to guide site-directed mutagenesis work, to probe these potential interactions. Nevertheless, the determination of the structure of the regulatory complex including UMPK would appear essential for clarifying how *carAB* transcription is repressed by pyrimidines.

While this manuscript was being submitted, the structure of the D159N mutant of *E. coli* UMPK, bound to either UMP, UDP or UTP, was reported.[36] The bacterial and archaeal enzymes are highly similar hexamers, having UMP bound in the same way and at equivalent sites, supporting the possibility to generalize the conclusions of the present study for both archaea and bacteria. Nevertheless, there are some differences between the two enzymes. Thus, the *E. coli* enzyme does not have the mobile αF-αG loop, as predicted above from sequence comparisons, confirming our view that this loop is not essential. Also as proposed above, the ξ-amino group of Lys15 occupies in *E. coli* NAGK an equivalent position to that of the Mg2 ion of *P. furiosus* NAGK, supporting the suggestion that Lys15 and the Mg2 ion play the same role in the catalysis of phosphoryl group transfer in the two enzymes. Interestingly, in the *E. coli* enzyme, the UMP moiety of the negative effector UTP sits in the same place as UMP, and the UTP β and γ phosphate groups are approximately in the sites occupied by the γ and β phosphate groups of AMPPNP in *P. furiosus* NAGK. Thus, UTP may be a competitive inhibitor, rather than an allosteric inhibitor, *versus* both UMP and ATP, as already suggested for UMP by the results of kinetic studies with *E. coli* UMPK,[32] and for ATP in studies with *Streptococcus pneumoniae* UMPK.[8] The binding of any other nucleotide triphosphate on the same site as UTP should be inhibitory, and thus, the suggestion made on the bases of kinetic and fluorescence results that both nucleoside triphosphate effectors bind to the same allosteric site cannot be correct.[32]

Materials and Methods

Enzyme expression, SeMet substitution, purification, crystallization and crystal soaking

Overexpression and purification of plasmid-encoded *P. furiosus* UMPK having a 20 residue His_6 N-terminal extension was carried out as described,[24] except for the use, when indicated, of SeMet-substituted enzyme.[37] The recombinant, His-tagged enzyme was previously shown to be active and highly specific for UMP.[24] Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (using the Voyager DE-Pro instrument, Applied Biosystems) demonstrated a mass excess consistent with the replacement by SeMet of all the methionine residues of the UMPK monomer. Cubic crystals of approximately 0.2 mm in the largest dimension were grown as reported,[24] at 294 K, using the vapour-diffusion approach, in hanging drops prepared by mixing 1.5 μl of a solution of 11 mg/ml of UMPK in20 mM NaBicine (pH 8.5), containing 1 mM dithiothreitol, and 1.5 μl of 3.5 M sodium formate. To prepare enzyme–ligand complexes, crystals were soaked for four hours at 294 K in 3.5 M sodium formate containing 1.5 mM AMPPNP and 4 mM $MgCl_2$, or a mixture of 0.5 mM UMP, 1.5 mM AMPPNP and 4 mM $MgCl_2$.

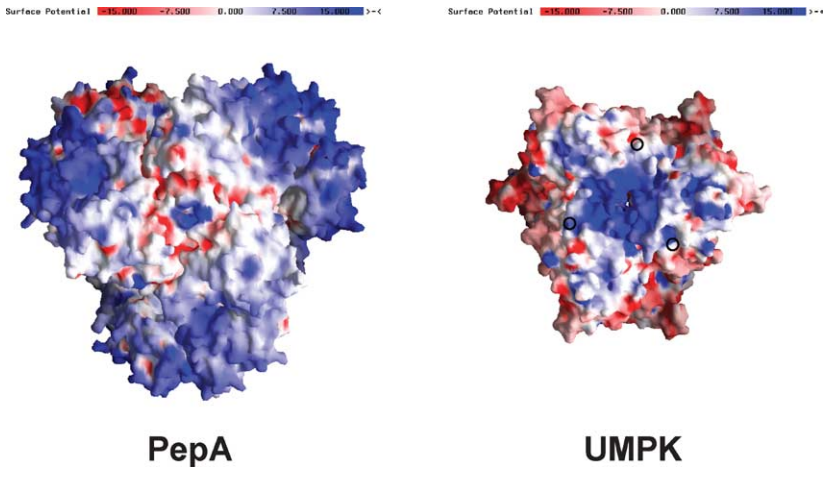


Figure 9. Molecular surface of PepA (PDB entry 1gyt) and UMPK from *E. coli*. The molecules are oriented with their 3-fold axes pointing towards the viewer. The UMPK *E. coli* structure was modelled from that of the *P. furiosus* enzyme (see the text). Electrostatic potentials were calculated with the program GRASP.[51] The polar surfaces are coloured blue (positively charged) and red (negatively charged). The circles on UMPK indicate the position of Ala94. The lack of perfect 3-fold symmetry in the surface potential of PepA results from small differences between the different subunits of the hexamer, given the non-crystallographic symmetry, with two hexamers in the asymmetric subunit (resolution, 2.5 Å).

Data collection and processing

Except the diffraction data collected at three wavelengths for multiple anomalous dispersion (MAD) phasing using SeMet-substituted enzyme, which were obtained in the Protein Structure Factory at the BESSY synchrotron at Berlin Adlershof, all diffraction data were collected at the European Synchrotron Radiation Facility, Grenoble, France, using the beamlines indicated in Table 1. In all cases, crystals were cooled to 100 K during diffraction. Cryoprotection was accomplished by brief immersion of the crystals in the 10% (v/v) glycerol-supplemented crystallization or soaking solution. Data were processed either using DENZO and SCALEPACK,[38] or MOSFLM and SCALA,[39] and the structure factors were obtained with TRUNCATE.[39]

All crystals belong to the cubic space group $I23$, where $a=b=c=144.6$–145.0 Å, although in the crystal of the complex with AMPPNP the edges were slightly larger ($a=146.7$ Å). Further details are given in Table 1.

Phasing, model building and refinement

To determine the structure of the unligated enzyme and of two UMPK–ligand complexes (Table 1) initial phases were obtained from a SeMet-substituted enzyme crystal containing no ligands by using the MAD approach. The crystal had two protein chains in the asymmetric unit, and we determined the positions of eight Se atoms with program SOLVE.[40] Initial MAD phases were improved by density modification carried out with RESOLVE,[41] obtaining a model with 281 residues (out of 490 residues in the asymmetric unit) with unambiguous identification of 183 amino acid side-chains. The model was checked using program O.[42] Then rigid body refinement and simulated annealing with program CNS[43] were applied, followed by several rounds of model building with O, alternating with positional and individual *B*-factor refinement with CNS. *B*-factor restraint was released gradually as the refinement progressed. All the diffraction data were used throughout the refinement process, except the 5% randomly selected data for calculating R_{free}. Refinement converged to a final R value of 21.0% ($R_{free}=26.4\%$) for a model of 427 residues at 2.6 Å resolution having excellent stereochemistry (tested with PROCHECK;[44] Table 1). This model was utilized in searching the phases by molecular replacement with the program MOLREP[45] for the higher resolution (2.4 Å) dataset obtained with a ligand-free crystal of native UMPK. The process of model building with O and refinement with CNS was as above, with automatic assignment of solvent positions with program WATERPICK of the CNS suite, followed by visual confirmation of the correctness of the assignations. Again, structure analysis with PROCHECK yielded highly satisfactory results (Table 1) for the final model having 427 residues (subunit A: His(–1) and residues 1–173 and 182–225; subunit B His(–1) and residues 1–146, 156–173 and 182–225). This last model, devoid of solvent molecules, was that used for molecular replacement with the two enzyme–ligand complexes analyzed here. Model building and refinement were as described above, yielding models again having excellent stereochemistry and encompassing the entire polypeptide chain (residues 1–225) for the complexes having AMPPNP and UMP (final resolution, 2.55 Å) or AMPPNP only (3 Å resolution).

Calculations and illustrations

Superposition of structures was carried out with program O (lsq option) using default parameters. Buried surface areas were calculated using NACCESS†. Figures were generated using MOLSCRIPT,[46] BOBSCRIPT,[47] RASTER3D,[48] MSMS[49] and DINO (Phillippsen‡).

Protein Data Bank accession codes

Coordinates and structure factors of ligand-free, AMPPNP-complexed and AMPPNP and UMP-complexed UMPK are deposited in the Protein Data Bank (PDB) with accession codes 2brx, 2bri and 2bmu, respectively.

Acknowledgements

This work was supported by grants from the Spanish Ministerio de Ciencia y Tecnología (BMC2001-2182), the Generalitat Valenciana (GV04B587) and the Instituto de Salud Carlos III of the Ministerio de Sanidad y Consumo (RCMN C03/08 and REDEMETH G03/54). Data collection at beamlines ID23 and BM16 of ESRF was under the auspices of the EU and the Spanish Ministry of Science and Technology, respectively. We thank BESSY for diffraction, and the personnel of the ESRF and BESSY for help. C. M.-M. and F.G.-O. had fellowships of I3P CSIC-Bruker España S. A. and Fundación Ferrer, respectively. We thank J. J. Calvete for mass spectrometry, A. Marina for helping with phasing and D. Charlier for pointing out the two-hybrid data on PepA and UMPK.

References

1. Neuhard, J. & Kelln, R. A. (1996). Biosynthesis and conversions of pyrimidines. In *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology (Neidhardt, F. C., ed.), vol. 1, pp. 580–599, ASM Press, Washington, DC.
2. Michal, G. (1999). *Biochemical Pathways: An Atlas on Biochemistry and Molecular Biology*, Wiley, New York.
3. Yan, H. G. & Tsai, M. D. (1999). Nucleoside monophosphate kinases: structure, mechanism, and substrate specificity. *Advan. Enzymol. Relat. Areas Mol. Biol.* **73**, 103–134.
4. Serina, L., Blondin, C., Krin, E., Sismeiro, O., Danchin, A., Sakamoto, H. *et al*. (1995). *Escherichia coli* UMP-kinase, a member of the aspartokinase family, is a hexamer regulated by guanine nucleotides and UTP. *Biochemistry*, **34**, 5066–5074.
5. Smallshaw, J. D. & Kelln, R. A. (1992). Cloning, nucleotide sequence and expression of the *Escherichia coli* K-12 pyrH gene encoding UMP kinase. *Life Sci. Advan.* **11**, 59–65.
6. Yamanaka, K., Ogura, T., Niki, H. & Hiraga, S. (1992).

† http://wolf.bms.umist.ac.uk/naccess
‡ http://cobra.mih.unibas.ch/dino

Identification and characterization of the *smbA* gene, a suppressor of the *mukB* null mutant of *Escherichia coli*. *J. Bacteriol.* **174**, 7517–7526.
7. Akerley, B. J., Rubin, E. J., Novick, V. L., Amaya, K., Judson, N. & Mekalanos, J. J. (2002). A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*. *Proc. Natl Acad. Sci. USA*, **99**, 966–971.
8. Fassy, F., Krebs, O., Lowinski, M., Ferrari, P., Winter, J., Collard-Dutilleul, V. & Salahbey Hocini, K. (2004). UMP kinase from *Streptococcus pneumoniae*: evidence for co-operative ATP binding and allosteric regulation. *Biochem. J.* **384**, 619–927.
9. Ramón-Maiques, S., Marina, A., Gil-Ortiz, F., Fita, I. & Rubio, V. (2002). Structure of acetylglutamate kinase, a key enzyme for arginine biosynthesis and a prototype for the amino acid kinase enzyme family, during catalysis. *Structure*, **10**, 329–342.
10. Marina, A., Alzari, P. M., Bravo, J., Uriarte, M., Barcelona, B., Fita, I. & Rubio, V. (1999). Carbamate kinase: new structural machinery for making carbamoyl phosphate, the common precursor of pyrimidines and arginine. *Protein Sci.* **8**, 934–940.
11. Ramón-Maiques, S., Marina, A., Uriarte, M., Fita, I. & Rubio, V. (2000). The 1.5 Å resolution crystal structure of the carbamate kinase-like carbamoyl phosphate synthetase from the hyperthermophilic Archaeon *Pyrococcus furiosus*, bound to ADP, confirms that this thermostable enzyme is a carbamate kinase, and provides insight into substrate binding and stability in carbamate kinases. *J. Mol. Biol.* **299**, 463–476.
12. Gagyi, C., Bucurenci, N., Sîrbu, O., Labessse, G., Ionescu, M., Ofiteru, A. *et al*. (2003). UMP kinase from the Gram-positive bacterium *Bacillus subtilis* is strongly dependent on GTP for optimal activity. *Eur. J. Biochem.* **270**, 3196–3204.
13. Sakamoto, H., Landais, S., Evrin, C., Laurent-Winter, C., Barzu, O. & Kelln, R. A. (2004). Structure-function relationships of UMP kinases from pyrH mutants of Gram-negative bacteria. *Microbiology*, **150**, 2153–2159.
14. Marco-Marín, C., Ramón-Maiques, S., Tavarez, S. & Rubio, V. (2003). Site-directed mutagenesis of *Escherichia coli* acetylglutamate kinase and aspartokinase III probes the catalytic and substrate-binding mechanisms of these amino acid kinase family enzymes and allows three-dimensional modelling of aspartokinase. *J. Mol. Biol.* **334**, 459–476.
15. Pérez-Arellano, I., Gil-Ortiz, F., Cervera, J. & Rubio, V. (2004). Glutamate-5-kinase from *Escherichia coli*: gene cloning, overexpression, purification and crystallization of the recombinant enzyme and preliminary X-ray studies. *Acta Crystallog. sect. D*, **60**, 2091–2094.
16. Leisinger, T. & Haas, D. (1975). N-Acetylglutamate synthase of *Escherichia coli* regulation of synthesis and activity by arginine. *J. Biol. Chem.* **250**, 1690–1693.
17. Fernández-Murga, M. L., Gil-Ortiz, F., Llácer, J. L. & Rubio, V. (2004). Arginine biosynthesis in *Thermotoga maritima*: characterization of the arginine-sensitive N-acetyl-L-glutamate kinase. *J. Bacteriol.* **186**, 6142–6149.
18. Marvil, D. K. & Leisinger, T. (1977). N-acetylglutamate synthase of *Escherichia coli*: purification, characterization, and molecular properties. *J. Biol. Chem.* **252**, 3295–3303.
19. Kholti, A., Charlier, D., Gigot, D., Huysveld, N., Roovers, M. & Glansdorff, N. (1998). pyrH-encoded UMP-kinase directly participates in pyrimidine-specific modulation of promoter activity in *Escherichia coli*. *J. Mol. Biol.* **280**, 571–582.
20. Strater, N., Sherratt, D. J. & Colloms, S. D. (1999). X-ray structure of aminopeptidase A from *Escherichia coli* and a model for the nucleoprotein complex in Xer site-specific recombination. *EMBO J.* **18**, 4513–4522.
21. Rice, P. A., Yang, S., Mizuuchi, K. & Nash, H. A. (1996). Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell*, **87**, 1295–1306.
22. Hall, D. A., Zhu, H., Zhu, X., Royce, T., Gerstein, M. & Snyder, M. (2004). Regulation of gene expression by a metabolic enzyme. *Science*, **306**, 482–484.
23. Ogura, M. & Tanaka, T. (1996). Transcription of *Bacillus subtilis degR* is sigma D dependent and suppressed by multicopy *proB* through sigma D. *J. Bacteriol.* **178**, 216–222.
24. Marco-Marín, C., Escamilla-Honrubia, J. M. & Rubio, V. (2005). First-time crystallization and preliminary X-ray crystallographic analysis of a bacterial-archaeal type UMP kinase, a key enzyme in microbial pyrimidine biosynthesis. *Biochim. Biophys. Acta*, **1747**, 271–275.
25. Gil-Ortiz, F., Ramón-Maiques, S., Fita, I. & Rubio, V. (2003). The course of phosphorus in the reaction of N-acetyl-L-glutamate kinase, determined from the structures of crystalline complexes, including a complex with an AF_4-transition state mimic. *J. Mol. Biol.* **331**, 231–244.
26. Knowles, J. R. (1980). Enzyme-catalyzed phosphoryl transfer reactions. *Annu. Rev. Biochem.* **49**, 877–919.
27. Mildvan, A. S. (1997). Mechanisms of signaling and related enzymes. *Proteins: Struct. Funct. Genet.* **29**, 401–416.
28. Patel, S., Martinez-Ripoll, M., Blundell, T. L. & Albert, A. (2002). Structural enzymology of Li^+-sensitive/Mg^{2+}-dependent phosphatases. *J. Mol. Biol.* **320**, 1087–1094.
29. Steitz, T. A. & Steitz, J. A. (2002). A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl Acad. Sci. USA*, **90**, 6498–6502.
30. Bucurenci, N., Serina, L., Zaharia, C., Landais, S., Danchin, A. & Barzu, O. (1998). Mutational analysis of UMP kinase from *Escherichia coli*. *J. Bacteriol.* **180**, 473–477.
31. Villeret, V., Clantin, B., Tricot, C., Legrain, C., Roovers, M., Stalon, V. *et al*. (1998). The crystal structure of *Pyrococcus furiosus* ornithine carbamoyltransferase reveals a key role for oligomerization in enzyme stability at extremely high temperatures. *Proc. Natl Acad. Sci. USA*, **95**, 2801–2806.
32. Serina, L., Bucurenci, N., Gilles, A. M., Surewicz, W. K., Fabian, H., Mantsch, H. H. *et al*. (1996). Structural properties of UMP-kinase from *Escherichia coli*: modulation of protein solubility by pH and UTP. *Biochemistry*, **35**, 7003–7011.
33. Charlier, D., Kholti, A., Huysveld, N., Gigot, D., Maes, D., Thia-Toong, T. L. & Glansdorff, N. (2000). Mutational analysis of *Escherichia coli* PepA, a multifunctional DNA-binding aminopeptidase. *J. Mol. Biol.* **302**, 411–426.
34. Nguyen, M. (1998). On the role of PyrH, PepA and ArgR in the pyrimidine-specific regulation of the *Escherichia coli* carAB operon encoding carbamoyl phosphate synthetase. Master thesis, Molecular Biology GGS IPMB. Vrije Universiteit Brussel, Brussels.
35. Sali, A. & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815.
36. Briozzo, P., Evrin, C., Meyer, P., Assairi, L., Joly, N., Bârzu, O. & Gilles, A. M. (2005). Structure of

Escherichia coli UMP kinase differs from that of other NMP kinases and sheds new light on enzyme regulation. *J. Biol. Chem* **280**, 25533–25540.
37. Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J. & Huber, R. (1995). High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in *Escherichia coli*. *Eur. J. Biochem.* **230**, 788–796.
38. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscilation mode. *Methods Enzymol.* **276**, 307–326.
39. Collaborative Computational Project Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
40. Terwillinger, T. C. & Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallog. sect. D*, **55**, 849–861.
41. Terwilliger, T. C. (2000). Maximum-likelihood density modification. *Acta Crystallog. sect. D*, **56**, 965–972.
42. Jones, T. A., Zou, J.-Y., Cowan, S. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
43. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al*. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
44. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
45. Vagin, A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Crystallog.* **30**, 1022–1025.
46. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
47. Esnouf, R. M. (1999). Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallog. sect. D*, **55**, 938–940.
48. Merritt, E. A. & Murphy, M. E. P. (1994). Raster3D Version 2.0. A program for photorealistic molecular graphics. *Acta Crystallog. sect. D*, **50**, 869–873.
49. Sanner, M. F., Olson, A. J. & Spehner, J. C. (1996). Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers*, **38**, 305–320.
50. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. *Protein Eng.* **8**, 127–134.
51. Honig, B. & Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science*, **268**, 1144–1149.

Edited by R. Huber

(Received 28 May 2005; received in revised form 12 July 2005; accepted 14 July 2005)
Available online 10 August 2005

Capítulo 4

A novel two-domain architecture within the amino acid kinase enzyme family revealed by the crystal structure of *Escherichia coli* glutamate 5-kinase.

Trabajo enviado a

Journal of Molecular Biology

A NOVEL TWO-DOMAIN ARCHITECTURE WITHIN THE AMINO ACID KINASE ENZYME FAMILY REVEALED BY THE CRYSTAL STRUCTURE OF *Escherichia coli* GLUTAMATE 5-KINASE

Clara Marco-Marín[1], Fernando Gil-Ortiz[1], Isabel Pérez-Arellano[2], Javier Cervera[2], Ignacio Fita[3] and Vicente Rubio[1,*]

[1]*Instituto de Biomedicina de Valencia (IBV-CSIC) and CIBERER (Instituto de Salud Carlos III), Jaume Roig 11, Valencia-46010, Spain*

[2]*Centro de Investigación Príncipe Felipe (FVIB-CSIC), Avda. Autopista del Saler 16, Valencia-46013, Spain*

[3]*Instituto de Biología Molecular de Barcelona(IBMB-CSIC). Institut for Research in Biomedicine. Parc Científic, Josep Samitier 1-5, 08028-Barcelona, Spain.*

Present address: F. Gil-Ortiz, Centro de Investigación Príncipe Felipe (FVIB-CSIC), Avda. Autopista del Saler 16, Valencia-46013, Spain

* *Corresponding author*:

Vicente Rubio

Instituto de Biomedicina de Valencia

Jaume Roig 11, Valencia-46010, Spain

E-mail: rubio@ibv.csic.es Tel. +34 963 391 772 Fax. +34 963 690 800

Short title: Structure of γ-glutamyl kinase of *Escherichia coli*

Summary. Glutamate 5-kinase (G5K) makes the highly instable product glutamyl-5-phosphate (G5P) in the initial, controlling step of proline/ornithine synthesis, being feed-back inhibited by proline or ornithine, and causing, when defective, clinical hyperammonaemia. In higher eukaryotes G5K belongs to a bifunctional polypeptide that includes the next enzyme of the route, glutamyl-5-phosphate reductase (G5PR), effecting internal G5P channelling, but bacterial G5Ks are monofunctional and must associate with G5PR. The bacterial enzymes generally contain, judged from the sequence, a constant amino acid kinase (AAK) domain typical of acylphosphate-forming enzymes, and a C-terminal PUA domain of obscure function that is typical of RNA-modifying enzymes and that might be implicated in gene regulatory functions demonstrated for G5K in *Bacillus subtilis*. We determine the crystal structure of a PUA domain-containing G5K, from *Escherichia coli*, at 2.9- and 2.5-Å-resolution, having bound at the active centres either glutamate and sulphate, or G5P, sulphate and the proline analog 5-oxoproline. *E.coli* G5K presents a novel tetrameric (dimer of dimers) architecture that is well suited for G5P channelling. The canonic AAK and PUA domains of different subunits associate non-canonically with the same domains of the other subunit to build the dimer. Two dimers interact exclusively through the AAK domains. The tetramer is flat and elongated and has in each face, pericentrically, two exposed active centres in accessible craters in the AAK domains of alternate subunits, permitting the close apposition of two active centres of one interacting bacterial G5PR tetramer. The structures clarify substrate binding and catalysis, justify the high glutamate specificity and K_m value, agree with the requirement for free Mg for activity, explain the effects on the activity of known point mutations, rationalize some of the effects of PUA domain-deletion, and support that proline triggers inhibition by binding

near glutamate. RNA is modelled binding to G5K as in other PUA domain-containing enzymes.

Keywords: Δ^1-Pyrroline 5-carboxylate synthetase; proline synthesis; ornithine synthesis; amino acid kinase family; PUA domain.

Abbreviations used: AAK, amino acid kinase; CK, carbamate kinase; G5K, L-glutamate 5-kinase; G5P, glutamate 5-phosphate; G5PR, glutamyl-5-phosphate reductase; MAD, multiple anomalous dispersion; MIR, multiple isomorphous replacement; NAGK, N-acetyl-L-glutamate kinase; PUA, a domain named after pseudouridine synthases and archaeosine-specific transglycosylases; P5CS, Δ^1-pyrroline 5-carboxylate synthetase; rmsd, root mean square deviation; SAD, single anomalous dispersion; SIR, single isomorphous replacement.

INTRODUCTION

Proline is not only a protein building block, but it also fulfills other important functions, protecting cells from osmotic stress, being a scavenger of free radicals, a transient form of nitrogen storage, a source of reducing power, and being involved in pH regulation.[1-5] Microorganisms and plants make proline from glutamate in three steps catalyzed by glutamate 5-kinase (G5K), glutamyl 5-phosphate reductase (G5PR) and pyrroline 5-carboxylate reductase (Figure 1).[6-8] Mammals also use G5K and G5PR to synthetize ornithine (Figure 1),[9-10] what is crucial for proper ammonia detoxification, as highlighted in Δ^1-pyrroline 5-carboxylate synthetase deficiency (OMIM 138250; http://www.ncbi.nlm.nih.gov), an inborn error in which a G5K-inactivating mutation causes clinical hyperammonaemia, with hypoornithininemia and hypoprolinemia.[11] G5K plays a key role in proline or ornithine biosynthesis not only because it catalyzes the first step of these biosynthetic routes, but also because it is the subject of feed-back inhibition by the final products, proline in microorganisms and plants,[6,7] and ornithine in animals.[10]

In animals and plants G5K is the N-terminal moiety of a bifunctional polypeptide called pyrroline 5-carboxylate synthetase which also contains, as its C-terminal moiety, the enzyme G5PR.[7,10] The bifunctionality apparently serves the purpose of channelling between G5K and G5PR of the G5K product glutamyl 5-phosphate (G5P), a product that is highly instable[12] because of the rapid intramolecular reaction between its amino and γ-acylphosphate groups to yield 5-oxoproline (also called pyroglutamate) (Figure 1).[13] Channelling of G5P between G5K and G5PR also appears necessary in prokaryotes,[13] but, nevertheless, bacteria have monofunctional G5Ks[14-15] that may be expected to associate in as yet undetermined ways with the G5PR present in these organisms, which is also a monofunctional enzyme.[14,16] Possibly

reflecting this association, in the activity assay of *Escherichia coli* G5K using hydroxylamine it is essential to add G5PR (in absence of the G5PR substrate NADPH) for efficient convertion of the G5P produced to the corresponding hydroxamate.[13]

In any case, in both prokaryotes and eukaryotes G5K contains a domain of ~260-residues that exhibits substantial sequence identity over its entire length with carbamate kinase (CK) and N-acetyl-L-glutamate kinase (NAGK).[17] The latter two enzymes catalyze the same type of reaction as G5K, consisting in the synthesis of a carboxylic-phosphoric anhydride using ATP as the phosphoryl group donor, and they belong to the same structural family, called the amino acid kinase (AAK) family (Pfam: PF00696; http://www.sanger.ac.uk/Software/Pfam), a family that is characterized by a structure conforming to an $\alpha_3\beta_8\alpha_4$ sandwich fold having a typical constant topology.[17] Most bacterial G5Ks such as the archetypical *E. coli* enzyme[18] not only have an AAK domain, but, similarly to plant and animal G5Ks, they are linked C-terminally to another domain which in the case of the bacterial enzymes has only ~110 residues and appears to be on the basis of its sequence a PUA domain.[19,20] PUA domains, named after pseudouridine synthases and archaeosine-specific transglycosylases (Pfam: PF01472),[19] are ~80-residue domains that exhibit a characteristic β-sandwich fold and one or two α helices.[21-22] They are found in RNA-modifying enzymes, and, therefore, their identification in bacterial G5Ks was surprising, although it was speculated[19] that they might be involved in the gene expression regulation of obscure mechanism that was reported for G5K in *Bacillus subtilis*.[23-24] Although the actual function of the PUA domain of bacterial G5Ks is not known, this domain is not needed for substrate binding, catalysis or regulation, since either the natural or artificial lack of this domain, respectively, in *Streptococcus thermophilus*[15] G5K or in the PUA domain-deleted *E. coli* G5K,[20] does not prevent catalysis or proline inhibition of these G5Ks. Further, the

PUA domain-deleted *E. coli* enzyme appears on gel filtration to be tetrameric, similarly to the wild-type enzyme,[20] and, therefore, the PUA domain is not essential for dimer formation. Nevertheless, the deletion of this domain increased and decreased, respectively, the proline and Mg requirements for inhibition and for activity of *E. coli* G5K, and it also abolished proline-triggered aggregation of the enzyme tetramers into higher oligomers,[20] indicating that the PUA domain might have some role in modulating the enzymatic properties of bacterial G5Ks.

It appears important to determine the structure of a bacterial G5K that has a PUA domain, to confirm that this domain as well as the AAK domain fit the canonic criteria for such types of domains, and also to try to understand the roles of both domains, their respective interactions in the building of the apparently tetrameric architecture having no precedent within the AAK family, the potential of both domains for mediating interactions with G5PR and possibly for mediating gene regulation, and, of course, to understand substrate binding, catalysis and regulation in G5Ks. We answer here some of these questions, by determining two crystal structures for *E. coli* G5K, at 2.9 and 2.5 Å resolution, having in their active centres either bound glutamate and sulphate or G5P, sulphate and the proline analog 5-oxoproline. Our structures reveal a novel architecture within the AAK and PUA domain families which is well adapted for G5P channelling and they largely clarify substrate binding and catalysis, accounting for the reported effects of some mutations.[25] Further, the structures provide hints on the mechanism of allosteric regulation by proline, pointing to the binding of the effector near the substrate glutamate. The structure allows plausible models to be proposed for macromolecular association with bacterial G5PR to allow product channelling, and to model interaction with an RNA on the basis of the interaction of another PUA domain-containing enzyme.[26]

RESULTS AND DISCUSSION

Structure determination

E. coli G5K posed big challenges to crystallization and phasing. Tetragonal and monoclinic crystals (Table 1) diffracting at 2.5 Å or 2.9 Å resolution and belonging to the space groups $P4_12_12$ or $P2_1$, respectively, either grew after several months and could not be reproduced (tetragonal crystals) or grew slowly and their reproducibility was poor (monoclinic crystals). Phasing attempts failed using MAD, SAD, MIR or SIR or selenomethionine or heavy atom derivatives including Br or I derivatives, or when using molecular replacement with models of *E. coli* NAGK[17] or *Pyrococcus furiosus* CK[27] (23 and 19 % sequence identity, respectively, with residues 1-257 of *E. coli* G5K). Phases were finally obtained for the tetragonal crystal data by molecular replacement using as a model a putative G5K from *Campylobacter jejuni* having no PUA domain and exhibiting ~33% sequence identity with residues 1-257 of *E. coli* G5K, deposited in the PDB (file 2AKO) by a structural genomics consortium. After introduction of a model of the archetypical PUA domain of the archaeosine tRNA-guanine transglycosylase from *Pyrococcus horikoshii*,[21] a dimer was obtained in the asymmetric unit, with each subunit being composed of one AAK and one PUA domain that interact with the same domains of the other subunit. Application of the $P4_12_12$ crystallographic symmetry produced a tetramer in agreement with our previous conclusion based on gel filtration data that *E. coli* G5K is tetrameric.[20] In this tetramer two dimers interact by less extensive interfaces mediated exclusively by the AAK domains. A large mass of non-protein electron density was found at the active centre of each subunit that was interpreted as one molecule of each, L-glutamyl-5-phosphate, 5-oxoproline and sulphate. The model for both subunits lacked the 41 or 43 residues beginning in residue 172, because of the absence of defined electron density for these residues. However, the

model for the complete polypeptide (except residues 1 and 2; residues in the model, 3-367) could be obtained at 2.9 Å in two subunits (D and E) of the monoclinic crystal. The solution for this monoclinic crystal, obtained by molecular replacement using as a model the dimer determined with the tetragonal crystal, consisted of two tetrameric molecules (each one exhibiting non-crystallographic point group *222* symmetry) formed by subunits A-D and E-H, respectively (Figure 2 (a) to (c)). Non-protein densities in this monoclinic crystal corresponded to glutamate and one sulphate in the active centre, two Mg ions in a negatively charged hole found between domains at the dimer interface (Figure 2(d)), and one bound glutamate in the PUA domain (Figure 2 (e)). The two tetramers in the monoclinic crystal and the tetramer generated in the tetragonal crystal, as well as all enzyme subunits, are essentially identical. The root mean square deviation (rmsd) values for the superimpositions of C^{α} atoms are 0.64-1.74 Å (mean 1.19 Å), when comparing complete tetramers, and 0.14-0.57 Å (mean 0.47 Å), when comparing individual subunits (residues 172-213, missing in the subunits from the tetragonal crystal, were excluded in all the superimpositions).

A novel subunit architecture made of canonical AAK and PUA domains.

The *E. coli* G5K subunit is composed of an N-terminal catalytic domain (with the modelled polypeptide chain spanning from Asp3 to Ala257) and a C-terminal domain (residues from Ala275 to Arg367) connected by a hinge that includes the short and irregular αI helix. The catalytic domain, with a rmsd of 1.5 Å for 216 equivalent residues after the superimposition of Ca atoms with *C. jejuni* G5K, exhibits a typical AAK fold (Figures 3 (a), (b) and (c)).[17] In turn, the C-terminal domain exhibits, as expected, the topology of the archetypical PUA domain from *P. horikoshii* archaeosine tRNA-guanine transglycosylase[21] with a rmsd value of 1.4 Å for the superimposition of

the C^{α} atoms from 71 equivalent residues. The *E. coli* G5K AAK domain contains a mainly parallel eight-stranded β-sheet sandwiched by two layers of three and four α-helices which look, respectively, towards the other dimer and towards the C-terminal domain of the same subunit. The AAK domain can be further divided into the N-terminal lobe and the C-terminal lobe (starting with Lys164), and exhibits two prominent bulges which are found also in all other AAK family enzymes and which emerge from the two lobes, on opposite ends of the β sheet C-edge. In contrast with most other enzymes of the AAK family,[17,27,28] G5K does not have any lid covering the active centre. The loops that connect helices B and C and strand β3 and helix D are not long and do not cover the site for the phosphorylatable substrate as in NAGK,[17] being oriented away from this site. Thus, the active centre is very wide and open and is surrounded by the two bulges, having the aspect of a large crater (Figure 2 (b)).

The *E. coli* G5K PUA domain is nucleated by a characteristic β sandwich that is its hallmark,[21,22] and which is composed of mainly antiparallel five-stranded and four-stranded sheets running at right angles. This domain has also two helices (J and K) found in other PUA domains,[21,29] and one extra helix (L) at the end of helix K which belongs to a short insertion spanning from residue His342 to Tyr354. At the free edge of the β sandwich a crevice is formed between helix J, strand β10 and the turn leading to β11, where one glutamate molecule was unexpectedly found in subunits A, D, E, F and G of the monoclinic crystal, but not in the tetragonal crystal (Figure 2 (e)). This glutamate molecule appears to interact almost exclusively through hydrophobic interactions, whereas polar atoms remain mostly exposed towards the solvent. On the opposite side of the β sandwich edge, helices K and L participate in the interactions with the PUA domain of a neighbour subunit. Therefore, both faces of the β sandwich

are exposed, and the PUA domain exhibits its five-stranded sheet on the same side of the subunit as the active centre of the AAK domain.

Finally, helix I, in the linker between the AAK and PUA domains, is embedded in a four-helix bundle with helices G, E and a helical turn preceding helix D. The face of helix I that looks towards the PUA domain concentrates most of the contacts with the latter domain, in particular with the β16 end from the five-stranded sheet (Figure 3 (b)).

The first tetramer within the AAK family.

The G5K structures determined in the two crystals correspond, as indicated before, to a molecular tetramer with accurate 222 symmetry and a global shape that can be approached by a flattened ellipsoid with radii of 34, 50 and 76 Å, respectively (Figure 2 (b) and (c)). Intersubunit interactions across the three molecular symmetry axes (named by convention and historical reasons P, Q and R)[30] vary widely, being absent across the P axis. Interactions, in turn, are very extensive across the twofold axis corresponding to the longest radius (the R axis) with 2150 $Å^2$ of buried surface per subunit (Figure 4(a)), while presenting a buried surface of only 430 $Å^2$ per subunit across the second axis (the Q axis) (Figure 4(b)). With intermolecular interactions across only two of the molecular symmetry axes it is consistent to define the G5K tetramer as the dimer across the P or Q axes of the dimer formed across the R axis (Figure 2 (b) and (c)). The tetramer leaves a central hole of ~5×8 Å, but appears to be highly stable, not dissociating even in dilute solution, as shown by gel filtration.[20] Subunits in the R dimer are situated in such a way that the AAK and the PUA domains interact only with the AAK and the PUA domain of the neighbour subunit, respectively. Since there are no intersubunit contacts between different types of domain, the strictures corresponding in each subunit to the interdomain boundary, merge in the dimer into a

central, negatively charged hole of ~8 Å diameter (Figure 2(d)). Within this negatively charged hole, two masses of density found in the monoclinic crystal have been interpreted as Mg ions that possibly play a structural role, since *E. coli* G5K needs free Mg for activity.[20] In agreement with the involvement of the PUA domain in forming this hole, PUA domain deletion altered the Mg requirement.[20] The centres of mass of the four domains of each dimer are approximately coplanar, but, given the larger size of the AAK domain, the dimer end corresponding to the AAK domains is thicker and wider (~100×50×40 $Å^3$) than the other end (~60×25×35 $Å^3$), which corresponds to the interacting PUA domains. The stricture between domains splits the flat and large surface of contact with the other subunit into a large portion corresponding to the AAK domain (1,663 $Å^2$, or ~80% of the total surface) and a smaller component (468 $Å^2$) corresponding to the PUA domain (Figure 4(a)). The surface mediating the contacts between both AAK domains is equivalent to the surface of dimerization of the other AAK family enzymes CK, NAGK and UMP kinase,[17,27,28] corresponding to the β3 end of the αβα sandwich of the AAK domain. The elements of this surface (β3, helices C and D and their interconnecting loops; the loop following β4; the C-end of helix B and residues of the helix B-helix C connection) form a cross-grid with the same elements of the other subunit (Figure 5), with intercrossing in all these enzymes of the long C helices of both subunits. However, the interacting surfaces represent in the various enzymes different degrees of rotation around an axis that penetrates perpendicularly the intersubunit interface at the crossover point of the C helices of both subunits, with rotations of ~110º for NAGK and CK,[17,27] ~190º for UMP kinase[28] and ~260º for *E. coli* G5K. The crossover point is also shifted from the third turn of helix C in NAGK and CK to the fifth turn in G5K. These changes in the mutual orientations of the interacting surfaces largely determine the shape of the dimer, and, in the case of G5K, results in the

exposition of the active centres of alternate subunits on opposite faces of the dimer and of the tetramer (Figure 2(b)), greatly facilitating channelling (see below).

Although G5K is the first known structure in which two PUA domains interact, the surface used for this interaction (the edge of the β sandwich corresponding to helix K and the β11-β12 connection, Figure 4(a)) is essentially the same surface that is utilized in archaeosine tRNA-guanine transglycosylase for interaction between the PUA domain and the catalytic domain.[21] In both enzymes the interactions leave exposed the 4-stranded sheet of the β sandwich, which is the face used for interaction with RNA in the PUA domains in which the structure of the PUA-RNA complex has been determined.[22,26,29] In G5K this face is flatter and smoother than the opposite side of the β sandwich and it has a more positively charged surface (Figure 6), which includes the positively charged exposed residues His360 and Arg361, fully conserved in the PUA domains of microbial G5Ks. In RNA-modifying enzymes that have a PUA domain, RNA interacts with the PUA domain and with the adjacent catalytic domain.[22,26,29] It is interesting that on the face of the G5K subunit exposing the 4-stranded sheet of the PUA domain, the AAK domain of G5K is also relatively smooth and presents a nearby patch of exposed positively charged residues that are invariant or highly conserved in PUA domain-containing G5Ks but not in PUA domain-lacking G5Ks: Arg111, Arg118 and Arg265 (Arg265 belongs to the interdomain linker and, as the PUA domain residues, is not present in the short G5Ks). Thus, it might be conceivable that these positively charged patches were used in potential RNA-mediated functions of microbial G5Ks having a PUA domain.

Since in the PUA domain-containing enzymes that have been crystallized in the presence of RNA, the surface of the domain that participates in the interaction is essentially the same,[22,26,29] the possible interaction between G5K and RNA was

modelled by a rigid body docking of the complex of the PUA domain of archaeosine tRNA-guanine transglycosylase from *P. horikoshii* with tRNA (PDB 1J2B)[26] (Figure 6). In this docking the tRNA sits on the surface of G5K, interacting with at least two clusters of positively charged residues and without presenting important steric clashes. Intriguingly, the 3´ end of the tRNA is located where the glutamate molecule bound to the PUA domain was found, suggesting possible regulatory mechanisms mediated by aminoacyl tRNAs that deserve further analysis, particularly given the involvement of G5K in gene regulation in *B. subtilis*.[23,24]

The interface between R dimers in the tetramer is provided exclusively by the 3-helix layer of the AAK domains, forming each two such domains across one interdimeric twofold axis a bundle of four antiparallel helices (helices A and C of each subunit) that is glued by hydrophobic contacts (involving Arg25, Ala26, Val 29, Ile 93 and Tyr94), four H bonds (between the Nη and Nα of Arg25 and Ala26 and the Oγ and Oα of Ser92 and Ile93, respectively), and two ion pairs (Glu30:Arg33) (Figure 4(b)). In the tetramer, the planes of both dimers, when viewed along the R axis, form an angle of ~20º, and thus the tetramer, as already indicated, is approximately planar and elongated, having a maximum dimension of ~140 Å, and being thicker and wider towards its centre, where the four AAK domains are found, than at both ends, corresponding to the two pairs of PUA domains. Two active centres are fully exposed in the central part of each face of the tetramer, on alternate subunits (Figure 2 (b) and (c)), arranged optimally for interaction with bacterial G5PR (see below). The protagonism of the AAK domain in the intersubunit contacts accounts for the persistence of the tetramer upon PUA domain deletion.[20] Further, the crystal contacts in the monoclinic crystal possibly justify the abolition of proline-triggered G5K aggregation into higher oligomers by the deletion of the PUA domain,[20] since, in the $P2_1$ crystal, the tetramers, running

approximately parallel and being shifted along their longer axis by a full subunit (Figure 2 (a)), interact by hybrid contacts between the AAK and PUA domains that would be eliminated by PUA domain deletion.

The active centre

Although both the monoclinic and the tetragonal G5K crystals were grown in the presence of glutamate and ADP, no nucleotide was found bound to any of the subunits. In the monoclinic crystal one sulphate and one glutamate were located in the large crater of the AAK domain in all the subunits (except in subunit H where the density for the glutamate was not well defined) (Figure 7(a)). In the tetragonal crystals the sulphate was also present, but the electron density corresponding to the glutamate was larger, extending both towards the sulphate and also, with discontinuity, in a region close to the Ca of the glutamate (Figure 7 (b) and (c)). This large density is well accounted for by a glutamyl 5-phosphate (G5P) and by a 5-oxoproline molecule. Since the crystals were grown in the presence of glutamate and ADP, the G5P may have been produced by the enzyme using contaminating ATP (ADP generally contains some ATP, and, in addition, it dismutates with time to ATP and AMP). The 5-oxoproline can then be spontaneously generated by intramolecular cyclisation of the G5P. Within the active site the binding pocket of glutamate and G5P is an open pit formed between β4, the β2-αB junction and the two ends of the large β4-αE loop all belonging to the N-lobe of the AAK domain.

The sulphate is located in the site that would be expected, from the structures of the homologous enzymes G5K from *C. jejuni* in complex with ADP (PDB code 2AKO) and of NAGK,[17] to accommodate the β-phosphate of ATP (Figure 8(a)). Indeed, the residues that in *E. coli* G5K interact with the sulphate, Ser14, Thr169 and Lys217, have equivalent residues in His11, Ser162 and Lys210 of G5K from *C. jejuni* and in Gly11,

Ser180 and Lys217 of *E. coli* NAGK.[17] The paramount role for catalysis of the positive charge in Lys217, previously proposed in NAGK to help stabilize the negative charge developing in the β-phosphate of the nucleotide as the γ-phosphoryl group is being transferred,[17,31] is strongly supported by the effect of the mutations Lys217Ala and Lys217Arg, both of which decrease >100-fold enzyme activity.[25] The hydrogen bond provided by Thr169 is also highly relevant for ATP binding and for catalysis, since the Thr169Ala mutation increases the K_m^{ATP} ~10-fold and decreases the apparent V_{max} for ATP to ~5 % of normal, without affecting much the $S_{0.5}^{Glu}$. In contrast, the mutation Thr169Ser, which preserves the side-chain hydroxyl, produces very modest changes in the enzyme kinetic parameters.[25]

The glutamate found in the monoclinic crystals presents an extended conformation with one of its molecular faces totally exposed (Figure 8(a)). The correspondence of this conformation with that of the diequatorial form of L-*cis*-cycloglutamate, and the lack of steric constrains to accommodate the three extra carbons of a cycloglutamate, seem to justify the >1-order of magnitude better K_m of G5K for the conformationally more rigid cycloglutamate.[13] The extended conformation of the glutamate, and the interactions mediated by the α and γ carboxylates and by the α-amino group of this amino acid justify the high specificity of G5K for L-glutamate, and in particular, its discrimination versus D-glutamate, L-aspartate, L-glutamine, glutarate, and γ-aminobutyrate.[13,20] The α and γ-carboxylates of the glutamate (Figures 7(a) and 8(a)) sit near the N-termini of helices B and E, respectively, making hydrogen bonds with main chain N-atoms of the first turn of these helices (either Gly51, Ala52 or Ile53, depending on the subunit, or, in helix E, Asn149). In addition, the γ-carboxylate makes a hydrogen bond with the side-chain of Ser50, from the β2-αB loop. The α-amino group anchors glutamate to the floor of the pit by forming one salt bridge with the side

chain of Asp137, from the initial part of the β4-αE loop. As expected, the loss of this bond, in the Asp147Gly mutant of tomato G5K (Asp147 corresponds in the tomato G5K sequence to Asp137 of *E. coli* G5K), decreases activity fivefold in the assay at 75 mM glutamate,[32] a result that could be accounted by a drastic decrease in the affinity for glutamate. The α-amino and α-carboxylate of glutamate also make hydrogen bonds (in some subunits) with the δ-O and δ-N groups from Asn134, respectively. Asn134 is part of the hydrogen bond chain linking the two catalytic centres in the R dimer and involving also the conserved residues Asn80, Gln100 and Asn149 (Figure 8(b)). Preliminary mutational studies (Carmona and Cervera, in preparation) support the involvement of this hydrogen bond chain in modulating cooperativity between subunits. Asn134 and Asn149 conform most of the floor of the glutamate binding pocket with the Asn149 side chain running antiparallel to the glutamate chain. The open character of the glutamate binding site might contribute to the very high $S_{0.5}^{Glu}$ of the *E. coli* enzyme (estimated in our hands as ~90 mM).[20,25]

There are little changes in the protein residues of the active centre of the tetragonal crystal, relative to the monoclinic crystal. The binding of the sulphate and of the α-amino and of the γ-carboxyl groups of the G5P is essentially the same as in the monoclinic crystal, with residues surrounding the bound ligands also identical in both crystals (Figure 7 (c)). However, in the tetragonal structure the a-carboxylate of G5P interacts with the ξ-N of Lys145, moving away from the carboxylate of the 5-oxoproline. Therefore, the glutamate and the glutamate moiety of G5P appear to be bound in different ways, possibly depending on the presence or absence of 5-oxoproline (or proline). The phosphate of G5P interacts with the ξ-N of the invariant Lys10, similarly to what happens with the γ-phosphate of ATP and Lys8 in *E. coli* NAGK.[17,31] The importance in catalysis of this lysine has been shown by site-directed mutagenesis

in both G5K and NAGK.[25,33] The 5-oxoproline molecule sits flat on the side chain of Ile53 with its amidic O group hydrogen-bonded to the hydroxyl group of Thr13, and with its carboxylate bound to the amino group and to the side-chain of Asn134 (Figures 7(b) and (c)). This carboxylate occupies approximately the same site used in the monoclinic crystal by the α-carboxylate of the glutamate.

The structure of G5K highlights a key role for Asp150 in the organization of the active centre of the enzyme. Although this residue does not interact with any of the two bound sulphates or with glutamate, it makes bonds with the ξN atoms of both Lys10 and Lys217, and, in this way, it appears to orient these two substrate-binding and catalytic lysine residues. In agreement with this important proposed role, mutation of this residue to either alanine or asparagine inactivated the enzyme.[25] The mutation in NAGK of the corresponding residue, Asp162 (a residue that also interacts with both catalytic lysines), to glutamate, also essentially inactivated the enzyme.[33]

Taking together our data on the complexes containing sulphate and glutamate or G5P, and the deposited structure of *C. jejuni* G5K containing bound ADP, it is clear that the G5K active centre (Figure 8(a)) replicates the basic traits of the active centre of NAGK, the best studied enzyme of the AAK family, which also exhibits a high structural similarity with G5K.[17,31] Thus, the nucleotide and the phosphoryl acceptor are bound, respectively, on the C- and N-lobes of these enzymes, over the main sheet C-edge, with the nucleotide oriented with its polyphosphate chain looking towards the N-domain and with the purine looking on the opposite direction. The phosphoryl group that is transferred, represented in the present structures by the phosphoryl group of G5P, sits at the junction between both lobes. Catalysis of the transfer involves two invariant lysine residues (Lys10/Lys8 and Lys217/Lys217 in G5K/NAGK) that are connected and mutually oriented towards the substrates by an aspartate (Asp150/Asp162 for

G5K/NAGK) that plays a key role in organizing the active centre. The β1-αA and β2-αB junctions include in both enzymes small residues that provide interactions with the ATP β phosphate (represented here by the sulphate), the transferred phosphoryl and the acceptor carboxylate. The latter carboxylate sits virtually identically in the two complexes studied here and in NAGK. Instead, the other end of the glutamate molecule, in the two conformations observed here, differs importantly from the way of binding to NAGK.[17,31] Therefore, while the basic catalytic machinery is fixed, the details of the binding of the specific substrate that is to be phosphorylated vary, as previously observed also with UMP and aspartate among other enzymes of the AAK family.[28,34]

G5K regulation and potential bases for channelling

We could not obtain G5K crystals containing proline, but the binding in the active centre of the tetragonal crystal, next to the glutamate site, of the proline analog 5-oxoproline fits the recent conclusion of kinetic and mutagenesis studies that proline and glutamate bind at overlapping sites.[25] Since oxoproline is a poor inhibitor of bacterial G5K,[33,35] the binding of proline must trigger effects that are not triggered by oxoproline. The positive charge on the proline imino group that is not present in oxoproline may trigger conformational changes mediated by Asp148, a residue that when mutated to asparagine dramatically increased the $I_{0.5}^{Pro}$,[25] and which belongs to the mobile β4-αE loop. The existence of very important conformational changes in this loop including Asp148 is supported by the comparison of the *E. coli* and *C. jejuni* structures (Figure 9). Many mutations found to hamper proline inhibition concentrate in the β4-αE loop (or in the equivalent region in other G5Ks),[25,32] which includes two residues involved in the hydrogen bond network that interlinks the glutamate sites of two subunits and which may be involved in the cooperativity for glutamate (Figure 8(b)). In this respect, proline

importantly increases the cooperativity for glutamate.[25] The tendency of the enzyme to aggregate in the presence of proline[20] may be the result of a change in the exposition of this external loop triggered by proline binding. Interestingly, mammalian G5K exists as two alternatively spliced forms, one having and the other not having two extra residues in the region corresponding to the β4-αE loop, and resulting in the shorter form being and the longer form not being feed-back inhibited by ornithine.[10] Thus, ornithine and proline may inhibit, respectively, the mammalian and the bacterial/plant enzymes in the same way involving changes in the β4-αE loop. In any case, the vast majority of the reported mutations that abolish proline inhibition map in the N-lobe of the AAK domain,[25,32] around the site for glutamate, again supporting the nearness of the glutamate and inhibitor sites. There is already another example among the enzymes of the AAK family in which the feed-back inhibitor binds overlapping with the substrate that is phosphorylated, although in the case of that enzyme, UMP kinase, the overlap is complete.[36] G5K and UMP kinase would represent one way of inhibition diametrally different from feed-back inhibition of NAGK by arginine, since in the latter enzyme arginine binds in a remote site relative to the active centre.[37] We find in G5K none of the traits of the arginine site of NAGK (not illustrated) excluding the possibility that G5K would resemble in this respect NAGK.

A novel feature of G5K is its tetrameric organization distant from the dimeric or hexameric architectures described for other AAK enzymes.[17,27,28,34,37] We interpret this organization, which appears highly stable despite the modest surface of interaction between the two dimers forming the tetramer, as being an adaptation to the needs of product channelling. The structure of the monoclinic G5K crystal containing bound G5P reveals that this product can reamin intact in the active centre of G5K. It should be possible to transfer G5P to G5PR without cyclization, if the active centres of G5K and

G5PR are apposed intimately and the thiol group of the active centre cysteine of G5PR reacts with G5K-bound G5P. This appears possible in the bienzymatic complex modelled in Figure 10. In this complex between both bacterial enzymes a mechanism of half of the sites reactivity would appear possible in which each one of the enzymes would interact with the other using, alternatively, one of its molecular faces. This proposed model agrees with the fact that, in higher eukaryotes, G5K and G5PR belong to a single polypeptide, and thus a model of the bienzymatic P5CS can be generated from the bacterial complex proposed above. This model assumes a tetrameric organization of mammalian P5CS with the four AAK domains organized as reported for *E. coli* G5K, and with a pair of G5PR domains oriented to each of the sides of the G5K tetramer, making interactions within G5PR that are equivalent to those observed in the recently determined structure of the G5PR domain of human P5CS (PDB file 2H5G). In the human enzyme, an Arg84Gln mutation was reported to cause pyrroline-5-carboxylate synthetase deficiency.[11] Although this residue is not constant in all G5Ks, being Gly18 the correspondin residue in the *E. coli* enzyme, the residue is at the β1-αA junction which is centrally located at the site of phosphoryl group transfer, and thus it is conceivable that structural changes triggered by mutations at this position may inactivate the enzyme. In any case it would be important to determine the crystal structure of human P5CS to document experimentally the deleterious effect of this mutation, and also to determine experimentally the architecture of the G5K-G5PR complex. Efforts in this direction are presently under way in our laboratory.

MATERIALS AND METHODS

Enzyme crystallization and data collection

E. coli (DH5α strain, from Clontech) *proB* cloning into pET-22b to yield pGKE, and overexpression and purification of G5K have been reported.[38] The sequence of the enzyme presents a single amino acid change of conservative nature (I129V), relative to the deposited sequence for *E. coli* G5K (file P0A7B5, Swissprot Database; http://www.expasy.org/uniprot). Bipyramidal-shaped crystals of approximately 0.3 mm maximal dimension were grown as reported,[38] in about four to five months, at 294 K, using the vapour-diffusion approach, in hanging drops prepared by mixing 1.5 μl of reservoir solution (1.6 M $MgSO_4$, 0.1 M KCl, 0.1 M MES pH 6.5) and 1.5 μl of enzyme solution (10 mg ml^{-1} G5K in 50 mM Tris-HCl pH 7.2, 1 mM dithiothreitol, 160 mM Na L-glutamate, 30 mM $MgCl_2$, 20 mM KCl and 6 mM ADP). In addition to these crystals, small plates of 0.1 mm maximal dimension were prepared in the same way, using a reservoir solution containing 1.45 M $MgSO_4$, 20 mM $CaCl_2$, 0.1 M MES pH 6.5. Cryoprotection of the crystals was accomplished by immersion in 10 % (v/v) glycerol-supplemented crystallization solution. Complete X-ray diffraction datasets were collected at 100 K from single crystals at the European Synchrotron Radiation Facility (ESRF), Grenoble, France (Table 1), attaining 2.5 Å and 2.9 Å resolution with the bipyramidal crystals and the plates, respectively. Data were processed with MOSFLM, SCALA and TRUNCATE.[39] Space group for the bipyramidal crystals was $P4_12_12$ with unit cell parameters a= b= 101.1 Å, c= 178.6 Å. The space group for the plates was $P2_1$ with unit cell parameters a= 96.3 Å, b= 124.1 Å, c= 144.9 Å, α= γ= 90º, β= 94º. Packing density calculations for the observed cell sizes and for a mass of 38.9 kDa for the G5K polypeptide, agree with the presence of two and eight molecules of the enzyme

polypeptide in the asymmetric unit of the $P4_12_12$ and $P2_1$ crystals, respectively, for corresponding solvent contents of 58 and 55 %.

Structure determination and refinement - Statistics for data collection and refinement are summarized in Table 1. Phasing was successful by molecular replacement with MOLREP[40] with diffraction data between 50 and 4 Å resolution for the $P4_12_12$ crystal, using a polyalanine search model derived from the structure of the *C. jejuni* putative G5K dimer (PDB file 2AKO), a protein that exhibits 33 % sequence identity with the N-terminal 257 amino acids of the 367-residue polypeptide chain of *E. coli* G5K. The solution corresponded to two subunits in the asymmetric unit forming one molecular dimer. Rigid body refinement and simulated annealing with program CNS[41] were applied to the initial model, followed by several rounds of model building with COOT[42] and O,[43] alternating with positional refinement with REFMAC5.[44] All the diffraction data (50-2.5 Å) were used throughout the refinement process, except the 5 % randomly selected data for calculating R_{free}. After several rounds of model building and refinement some of the elements of the PUA domain were incorporated into the model and these guided the fitting of the entire PUA domain, taken from the structure of archaeosine tRNA transglycosylase (PDB entry 1IQ8)[21], using MOLREP.[40] TLS refinement was performed assuming one rigid body per subunit. Water molecules were assigned with REFMAC5,[44] and were visually confirmed. One molecule of glutamyl phosphate and one of 5-oxoproline were included in each one of the subunits, and were refined using libraries generated with REFMAC5. After several rounds of refinement with REFMAC5 alternating with model building with O[43] R_{factor}/R_{free} converged to final values of 19.3/24.3. The final model for the two subunits, obtained at a resolution of 2.5 Å, encompassed residues 3 to 367, although residues 172-213 and 172-211 in subunits

A and B, respectively, could not be traced because of lack of electron density The model presented satisfactory stereochemistry (checked with Procheck).[45]

To determine the phases for the $P2_1$ crystal, a partially refined model of the complete G5K dimer found in the $P4_12_12$ crystal was used for molecular replacement. The correct solution yielded four dimers in the asymmetric unit, organized as two tetramers, one of them constituted by subunits A to D and the other by subunits E to H. Model building and refinement was performed essentially as described for the $P4_12_12$ crystal, applying non-crystallographic symmetry restraints that were progressively relaxed, mainly for amino acid side-chains, in the final rounds of refinement. R_{factor}/R_{free} converged to final values of 19.7/24.7. The final model, at 2.9 Å resolution, for the eight subunits in the asymmetric unit, encompassed the entire polypeptide chain from residue 3 in subunits E and D, but lacked the following residues in the other subunits: 202-211 (subunit A); 202-213 and 367 (B); 203-211 (G and C); 176-185 and 196-213 (H); 203-213 (F). One glutamate and one sulphate were found in each active centre, except in subunit H, where the glutamate was missing. In five subunits, a density corresponding to a glutamate molecule was found associated to the PUA domain.

Calculations and illustrations

Buried surface areas were calculated with NACCESS (http://wolf.bms.umist.ac.uk/naccess) using a probe of radius 1.4 Å. Superposition of structures was carried out with program LSQKAB.[46] Figures were generated using MOLSCRIPT,[47] BOBSCRIPT,[48] RASTER3D,[49] PYMOL (DeLano, http://www.pymol.org), MSMS[50] and DINO (Phillippsen, http://cobra.mih.unibas.ch/dino).

Protein Data Bank accession codes

Coordinates and structure factors are deposited in the Protein Data Bank (PDB) with accession codes 2J5T and 2J5V.

Acknowledgements. This work was supported by grants BFU2004-05159, and PI052838 from the Spanish Ministries of Education and Science and of Health, respectively. Data collection at beamline ID14.4 and BM16 of ESRF were under the auspices of the EU and the Spanish Ministry of Science and Technology, respectively. We thank the personnel of the ESRF for help.

REFERENCES

1. Csonka, L. N. & Hanson, A. D. (1991). Prokaryotic osmoregulation: genetics and physiology. *Annu. Rev. Microbiol.* **45**, 569-606.

2. Delauney, A. J. & Verma, D. P. S. (1990). Proline biosynthesis and osmoregulation in plants. *Plant J.* **4**, 215-223.

3. Smirnoff, N. & Cumbes, Q. J. (1989). Hydroxyl radical scavenging activity of compatible solutes. *Phytochemistry*, **28**, 1057-1060.

4. Kohl, D. H., Schubert, K. R., Carter, M. B., Hagedorn, C. H. & Shearer, G. (1988). Proline metabolism in N_2-fixing root nodules: energy transfer and regulation of purine synthesis. *Proc. Natl. Acad. Sci. U S A,* **85**, 2036-2040.

5. Venekamp, J. H. , Lampe, J. E. M. & Koot, J. T. M. (1989). Organic acids as sources of drought-induced proline synthesis in field bean plants, *Vicia faba* L. *J. Plant Physiol.* **133**, 654-659.

6. Leisinger T. (1996). Biosynthesis of proline in: *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology (Neidhardt, F.C., Ed.), pp. 434-441, ASM Press, Washington, DC.

7. Hu, C. A., Delauney, A.J. & Verma D.P. (1992). A bifunctional enzyme (Δ^1-pyrroline-5-carboxylate synthetase) catalyzes the first two steps in proline biosynthesis in plants. *Proc. Natl. Acad. Sci. U S A*., **89**, 9354-9358.

8. Aral, B. & Kamoun, P. (1997). The proline biosynthesis in living organisms. *Amino Acids,* **13**, 189-217.

9. Alonso, E. & Rubio, V. (1989). Participation of ornithine aminotransferase in the synthesis and catabolism of ornithine in mice. Studies using gabaculine and arginine deprivation. *Biochem. J.* **259**, 131-138.

10. Hu, C. A. Lin, W. W., Obie, C. & Valle, D. (1999). Molecular enzymology of mammalian ? [1]-pyrroline-5-carboxylate synthase. Alternative splice donor utilization generates isoforms with different sensitivity to ornithine inhibition. *J. Biol. Chem.* **274,** 6754–6762.

11. Baumgartner, M. R., Hu, C. A., Almashanu, S., Steel, G., Obie, C., Aral, B., *et al.* (2000). Hyperammonemia with reduced ornithine, citrulline, arginine and proline: a new inborn error caused by a mutation in the gene encoding ? [1]-pyrroline-5-carboxylate synthase. *Hum. Mol. Genet.* **9,** 2853-2858.

12. Katchalsky, A. & Paecht, M. (1954). Phosphate anhydrides of amino acids. *J. Am. Chem. Soc.* **76**, 6042-6044.

13. Seddon, A. P., Zhao, K. Y. & Meister, A. (1989). Activation of glutamate by ?-glutamate kinase: formation of ?-cis-cycloglutamyl phosphate, an analog of ?-glutamyl phosphate. *J. Biol. Chem.* **264**, 11326-11335.

14. Deutch, A. H., Rushlow, K. E. & Smith, C. J. (1984). Analysis of the *Escherichia coli proBA* locus by DNA and protein sequencing. *Nucleic Acids Res.* **12,** 6337-6355.

15. Massarelli, I., Forlani, G., Ricca, E. & De Felice, M. (2000). Enhanced and feedback-resistant ?-glutamyl kinase activity of an *Escherichia coli* transformant carrying a mutated *proB* gene of *Streptococcus thermophilus*. *FEMS Microbiol. Lett*. **182,** 143-147.

16. Page, R., Nelson, M. S., von Delft, F., Elsliger, M. A., Canaves, J. M., Brinen, L. S., *et al.* (2004). Crystal structure of γ-glutamyl phosphate reductase (TM0293) from *Thermotoga maritima* at 2.0 Å resolution. *Proteins,* **54**, 157-161.

17. Ramón-Maiques, S., Marina, A., Gil-Ortiz, F., Fita, I. & Rubio, V. (2002). Structure of acetylglutamate kinase, a key enzyme for arginine biosynthesis and a prototype for the amino acid kinase enzyme family, during catalysis. *Structure,* **10,** 329-342.

18. Smith, C. J., Deutch, A. H. & Rushlow, K. E. (1984) Purification and characteristics of a γ-glutamyl kinase involved in *Escherichia coli* proline biosynthesis. *J. Bacteriol*. **157**, 545-551.

19. Aravind, L. & Koonin E. V. (1999). Novel predicted RNA-binding domains associated with the translation machinery. *J. Mol. Evol.* **48**, 291-302.

20. Perez-Arellano, I., Rubio, V. & Cervera, J. (2005). Dissection of *Escherichia coli* glutamate 5-kinase: functional impact of the deletion of the PUA domain. *FEBS Lett*. **579,** 6903-6908.

21. Ishitani, R., Nureki, O., Fukai, S., Kijimoto, T., Nameki, N., Watanabe M., *et al.* (2002). Crystal structure of archaeosine tRNA-guanine transglycosylase. *J. Mol. Biol.* **318**, 665-677.

22. Pan, H., Agarwalla, S., Moustakas, D. T, Finer-Moore J., & Stroud R. M. (2003). Structure of tRNA pseudouridine synthase TruB and its RNA complex: RNA recognition through a combination of rigid docking and induced fit, *Proc. Natl. Acad. Sci. USA*, **100**, 12648-12653.

23. Ogura, M., Kawata-Mukai, M., Itaya, M., Takio, K. & Tanaka, T. (1994). Multiple copies of the *proB* gene enhance degS-dependent extracellular protease production in *Bacillus subtilis*. *J. Bacteriol*. **176**, 5673-5680.

24. Ogura, M. & Tanaka, T. (1996). Transcription of *Bacillus subtilis degR* is sigma D dependent and suppressed by multicopy *proB* through sigma D. *J. Bacteriol.* **178**, 216-222.

25. Perez-Arellano, I., Rubio, V. & Cervera, J. (2006). Mapping active site residues in glutamate-5-kinase. The substrate glutamate and the feed-back inhibitor proline bind at overlapping sites. *FEBS Lett*. **580**, 6247-6253

26. Ishitani, R., Nureki, O., Nameki, N., Okada, N., Nishimura, S. & Yokoyama, S. (2003). Alternative tertiary structure of tRNA for recognition by a posttranscriptional modification enzyme. *Cell,* **113**, 383-394.

27. Ramón-Maiques, S., Marina, A., Uriarte, M., Fita, I. & Rubio, V. (2000). The 1.5 Å resolution crystal structure of the carbamate kinase-like carbamoyl phosphate synthetase from the hyperthermophilic archaeon *Pyrococcus furiosus*, bound to ADP, confirms that this thermostable enzyme is a carbamate kinase, and provides insight into substrate binding and stability in carbamate kinases. *J. Mol. Biol.* **299,** 463-476.

28. Marco-Marin, C., Gil-Ortiz, F. & Rubio V. (2005). The crystal structure of *Pyrococcus furiosus* UMP kinase provides insight into catalysis and regulation in microbial pyrimidine nucleotide biosynthesis. *J. Mol. Biol.* **352**, 438-54.

29. Li, L. & Ye, K. (2006). Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature*, **443**, 302-307.

30. Rossmann, M. G., Adams, M. J., Buehner, M., Ford, G. C., Hackert, M. L., Liljas, A. *et al.* (1973). Letter: Molecular symmetry axes and subunit interfaces in certain dehydrogenases. *J. Mol. Biol.* **76**, 533-537.

31. Gil-Ortiz, F., Ramon-Maiques, S., Fita, I. & Rubio, V. (2003). The course of phosphorus in the reaction of N-acetyl-L-glutamate kinase, determined from the structures of crystalline complexes, including a complex with an AlF_4^- transition state mimic. *J. Mol. Biol.* **331**, 231-244.

32. Fujita, T., Maggio, A., Garcia-Rios, M., Stauffacher, C., Bressan R. A. & Csonka L. N. (2003). Identification of regions of the tomato ?-glutamyl kinase that are involved in allosteric regulation by proline. *J. Biol. Chem.* **278**, 14203-14210.

33. Marco-Marín, C., Ramón-Maiques, S., Tavárez, S. & Rubio, V. (2003) Site-directed mutagenesis of *Escherichia coli* acetylglutamate kinase and aspartokinase III probes the

catalytic and substrate-binding mechanisms of these amino acid kinase family enzymes and allows three-dimensional modelling of aspartokinase. *J. Mol. Biol*. **334**, 459-476.

34. Kotaka, M., Ren, J., Lockyer, M., Hawkins, A. R. & Stammers, D.K. (2006). Structures of R- and T-state *Escherichia coli* aspartokinase III: Mechanisms of the allosteric transition and inhibition by lysine. *J. Biol. Chem*. **281**, 31544-31552.

35. Krishna, R.V. & Leisinger, T. (1979). Biosynthesis of proline in *Pseudomonas aeruginosa*. Partial purification and characterization of ?-glutamyl kinase. *Biochem. J.* **181**, 215-222.

36. Briozzo, P., Evrin, C., Meyer, P., Assairi, L., Joly, N., Barzu, O. & Gilles, A. M. (2005). Structure of *Escherichia coli* UMP kinase differs from that of other nucleoside monophosphate kinases and sheds new light on enzyme regulation. *J. Biol. Chem*. **280**, 25533-25540.

37. Ramón-Maiques, S., Fernandez-Murga, M. L., Gil-Ortiz, F., Vagin, A., Fita, I. & Rubio, V. (2006). Structural bases of feed-back control of arginine biosynthesis, revealed by the structures of two hexameric N-acetylglutamate kinases, from *Thermotoga maritima* and *Pseudomonas aeruginosa*. *J. Mol. Biol*. **356**, 695-713.

38. Perez-Arellano , I., Gil-Ortiz, F., Cervera, J. & Rubio, V. (2004) Glutamate-5-kinase from *Escherichia coli*: gene cloning, overexpression, purification and crystallization of the recombinant enzyme and preliminary X-ray studies. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2091-2094.

39. Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50**, 760-763.

40. Vagin, A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30,** 1022-1025.

41. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54**, 905-921.

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

43. Jones, T. A., Zou, J.-Y., Cowan S., & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. Sect. A,* **47**, 110-119.

44. Murshudov, G. N., Vagin, A.A.& Dodson, E. J. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Cryst. D Biol. Crystallogr.* ***53***, 240-255.

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

46. Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. *Acta Crystallog. Sect. A,* **32**, 922-923.

47. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946-950.

48. Esnouf, R. M. (1999). Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55**, 938-940.

49. Merritt, E. A., & Murphy, M. E. P. (1994). Raster3D Version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50**, 869-873.

50. Sanner, M. F., Olson, A. J., & Spehner, J. C. (1996). Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers*, **38**, 305-320.

51. Honig, B. & Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science*, **268**, 1144-1149.

FIGURE LEGENDS

Figure 1. Route of synthesis of proline in microorganisms and plants and of ornithine in mammals. Enzymes are enclosed in grey boxes. Feed-back inhibition of microbial and plant G5Ks by proline and of animal G5Ks by ornithine is indicated with broken arrows. The dotted arrow indicates the spontaneous cyclization of G5P to 5-oxoproline that is an abortive side-reaction.

Figure 2. The structure of *E. coli* glutamate 5-kinase. (a) Backbone representation of two views of the asymmetric unit of the $P2_1$ crystal, formed by two tetramers colored blue and red. The unit cell is shown as a black lined box. (b) and (c) Surface representation of two perpendicular views of the glutamate 5-kinase tetramer, with the substrates in space-filling representation. Different colours are used for each subunit with different intensities for the amino acid kinase (AAK, darker) and PUA (lighter) domains. The molecular P, Q and R axes are identified and represented with broken lines. (d) Magnesium binding site represented with the electrostatic potential of the solvent accessible surface. Positively charged regions are colored blue and negatively charged regions are colored red, with the intensity of the color being proportional to the local potential. The $2F_{obs}$-F_{calc} omit electron density map contoured at the 1 σ level, is also shown (in green) with the bound magnesium ions as the purple spheres. (e) Semitransparent surface representation of the PUA domain showing the bound glutamate (ball and stick representation) and its corresponding $2F_{obs}$-F_{cal} omit electron density map contoured at the 1 σ level.

Figure 3. Structure of the *E. coli* G5K subunit (a) Stereo view of the superposition of the C^{α} traces of G5K from *E. coli* (green) and *C. jejuni* (pink). The absence of the PUA

domain is evident in the *C. jejuni* enzyme. (b) The *E. coli* G5K subunit structure with the catalytic AAK (green) and the PUA (orange) domains. Secondary elements and the link between domains (purple) are also indicated. Two molecules of glutamate, a sulphate ion and a magnesium ion bound to the G5K subunit are represented with balls and sticks. (c) Correspondence between the amino acid sequence (single-letter code) and the secondary structure. Superposed cylinders denote a-helices, and arrows denote ß-strands. Color code is as in (b). Residues having decreased accessibility upon the binding of glutamyl 5-phosphate, sulphate, 5-oxoproline or glutamate in the amino acid kinase domain, are indicated with cyan, violet, brown or blue triangles respectively. Residues having decreased accessibility upon the binding of glutamate in the PUA domain are indicated with yellow triangles and residues that form the negatively charged hole where magnesium binds, with red triangles. Open and closed circles indicate residues that exhibit decreased accessibility upon homodimer and tetramer formation, respectively.

Figure 4. Oligomerization surfaces of one subunit across the (a) R and (b) Q dimers (see in the text). Contacting non-polar (reddish) and polar (pink) atoms are indicated together with the corresponding structural elements.

Figure 5. Comparison between the dimers of *E. coli* NAGK, *P. furiosus* UMP kinase (UMPK) and *E. coli* G5K. Left panel, ribbon representation of the homodimers of NAGK, UMPK and G5K, to highlight the drastically different dimer architectures. Subunits on the left (lighter) have the same orientation, in the three enzymes. Right panel, elements of the two subunits involved in dimer formation, to show their different relative orientations in the three enzymes.

Figure 6. Possible binding of RNA to G5K. Surface representation of two perpendicular views of the G5K tetramer. Electrostatic potentials were calculated with the program GRASP[51] and represented with Pymol (DeLano, http://www.pymol.org). The polar surfaces are colored blue (positively charged) and red (negatively charged). Cartoon of a tRNA molecule modelled on the G5K surface according to the complex of the PUA domain of the archaeosine tRNA-guanine transglycosylase from *P. horikoshii* and a tRNA (PDB 1J2B). It is remarkable the proximity of the glutamate found bound to the PUA domain of G5K (space filling representation) and the 3' terminus of the tRNA molecule. The positively charged cluster 1 on the AAK domain includes residues *Arg111, Arg118*, Arg122, *Arg265*, Arg267, while cluster 2 on the PUA domain includes Arg309, Arg329, *His360* and *Arg361* (in italics conserved residues among G5Ks).

Figure 7. Substrate binding. (a) Stereoview of the glutamate 5-kinase (G5K) active centre, in the monoclinic crystal. The bound glutamate and sulphate are shown in ball and stick representation. Amino acid side chains are shown in thinner trace. The ($2F_{obs}$-F_{cal}) omit electron density map, contoured at the 1 σ around the bound substrates is also shown (green). Hydrogen bonds and ion pairs are shown as broken black lines. (b) Similarly for the tetragonal crystal, with the corresponding bound molecules of glutamyl 5-phosphate, 5-oxoproline and sulphate (c) Stereo view of the superimposition of the substrate binding region and of the bound ligands in the monoclinic (colored) and tetragonal crystals of E. coli G5K. The only significant differences are between the glutamate and glutamylphosphate and the presence of oxoproline only in the tetragonal crystal.

Figure 8. The active centre. (a) Semitransparent surface representation of a detailed view of the G5K binding site, showing bound glutamate and sulphate in stick representation as well as an ADP molecule modelled from the structure of *C. jejuni* G5K in the presence of this nucleotide. Important side chains are represented in bonds and identified. (b) Stereo view showing the hydrogen bond network between the glutamate binding sites of the two subunits of the R dimer. Each subunit is shown in a different colour.

Figure 9. Movement of the β4-αE loop. Stereo view of the superposition of the *E. coli* (green) and *C. jejuni* (violet) loops and neighbouring elements including (in stick representation) bound 5-oxoproline. The side chain of the important residue for proline binding, Asp148 (Asp141 of *C. jejuni* G5K), is represented also to illustrate its radical change of orinetation from looking outwards to looking into the site.

Figure 10. Possible interaction between the glutamate 5-kinase (G5K) and the glutamyl 5-phoshate reductase (G5PR)**.** Docking was performed superimposing the symmetry axes, approaching the active centres and minimizing steric clashes between the two molecules. Semitransparent surface representation of two perpendicular views of the G5K tetramer, with the substrates in space-filling representation. A dimer of the G5PR from *T. maritima* is shown (ribbon representation) with one subunit (colored orange) corresponding to the open conformation, as observed in the crystal structure of this enzyme in the absence of substrates (PDB 1O20), and the second subunit (colored yellow) corresponding to a closed conformation modelled according to the structure of the homologous class 3 aldehyde dehydrogenase in the presence of NAD (PDB 1AD3). The catalytic and NADPH binding domains of G5PR are identified.

Table 1. Data collection and refinement statistics

	Tetragonal crystals	Monoclinic crystals
A. Data collection		
ESRF Beamline/Wavelenght (Å)	BM16/0.979	ID14.4/0.979
Space group	$P4_12_12$	$P2_1$
Unit cell (Å or º)	a = b = 101.1, c = 178.6	a = 96.30, b = 124.11, c = 144.93; ? = 93.96
Resolution range (Å)	48.8-2.5 (2.64-2.50)	50-2.9 (3.06-2.90)
Reflections, total/unique	367,219/ 61,566 (29,004/ 8,941)	339,415/ 74,908 (49,569/ 10,882)
Completeness (%)	99.5 (99.5)	99.3 (99.6)
I/σ	6.8 (1.7)	9.7 (2.9)
R_{sym} (%)[a]	6.8 (41.6)	16.5 (49.4)
B. *Refinement statistics*		
Resolution range (Å)	49 - 2.5	50 - 2.9
R-factor/ R_{free} (%)[b]	19.3/24.3	19.7/24.7
Molecules and atoms refined	Polypeptide chains: 2	Polypeptide chains : 8
	Protein atoms: 4,872	Protein atoms: 21,344
	Glutamyl-5-phosphate molecules: 2	Glutamate molecules : 12
	5-Oxoproline molecules:2	Sulphate ions: 8
	Sulphate ions: 2	Magnesium ions: 8
	Magnesium ions: 1	Chloride ions: 20
	Water molecules: 166	Water molecules: 103

Rmsd from ideal		
Bond lengths (Å)l	0.011	0.009
Bond angles (º)	1.35	1.16
Average B-factor ($Å^2$)		
Protein	36.94	19.04
Water	35.86	5.67
Glutamate		28.37
Glutamyl 5-phosphate	41.99	
5-Oxoproline	38.57	
Sulphate	55.75	54.45
Magnesium	19.25	39.90
Chloride		29.20
Ramachandran plot (%)[c]		
Most favoured	94.1	90.3
Additionally allowed	5.9	9.4
Generously allowed	0	0.3
Disallowed	0	0

Values in parentheses are data for the highest resolution shell

[a]$R_{sym} = \Sigma | I - <I> | / \Sigma I$, where I is the observed intensity and <I> is the average intensity of multiple observations of symmetry-related reflections.

[b]$\Sigma_h | | F_{obs} | - | F_{calc} | | / \Sigma_h | F_{obs} |$, where $| F_{obs} |$ and $| F_{calc} |$ are observed and calculated structure factors amplitudes for all reflections (R-factor) and the reflections applied in the test are R_{free} set (reflections not used in the structure refinement), respectively.

[c]Calculated using PROCHECK

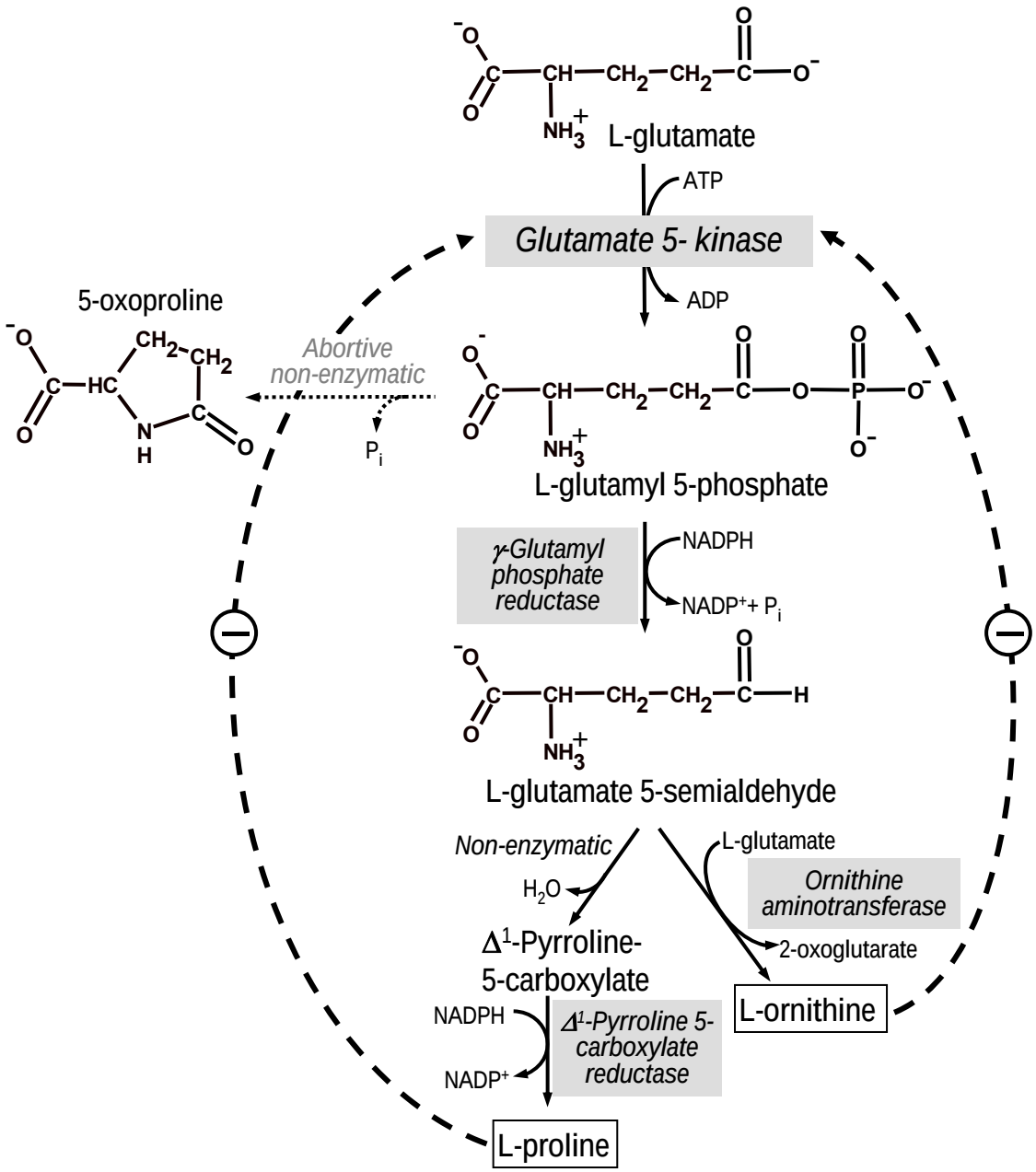

L-glutamate
ATP
Glutamate 5- kinase
ADP
5-oxoproline
Abortive non-enzymatic
P_i
L-glutamyl 5-phosphate
γ-Glutamyl phosphate reductase
NADPH
$NADP^+ + P_i$
L-glutamate 5-semialdehyde
Non-enzymatic
H_2O
Δ^1-Pyrroline-5-carboxylate
L-glutamate
Ornithine aminotransferase
2-oxoglutarate
L-ornithine
NADPH
$NADP^+$
Δ^1-Pyrroline 5-carboxylate reductase
L-proline

Figure 1

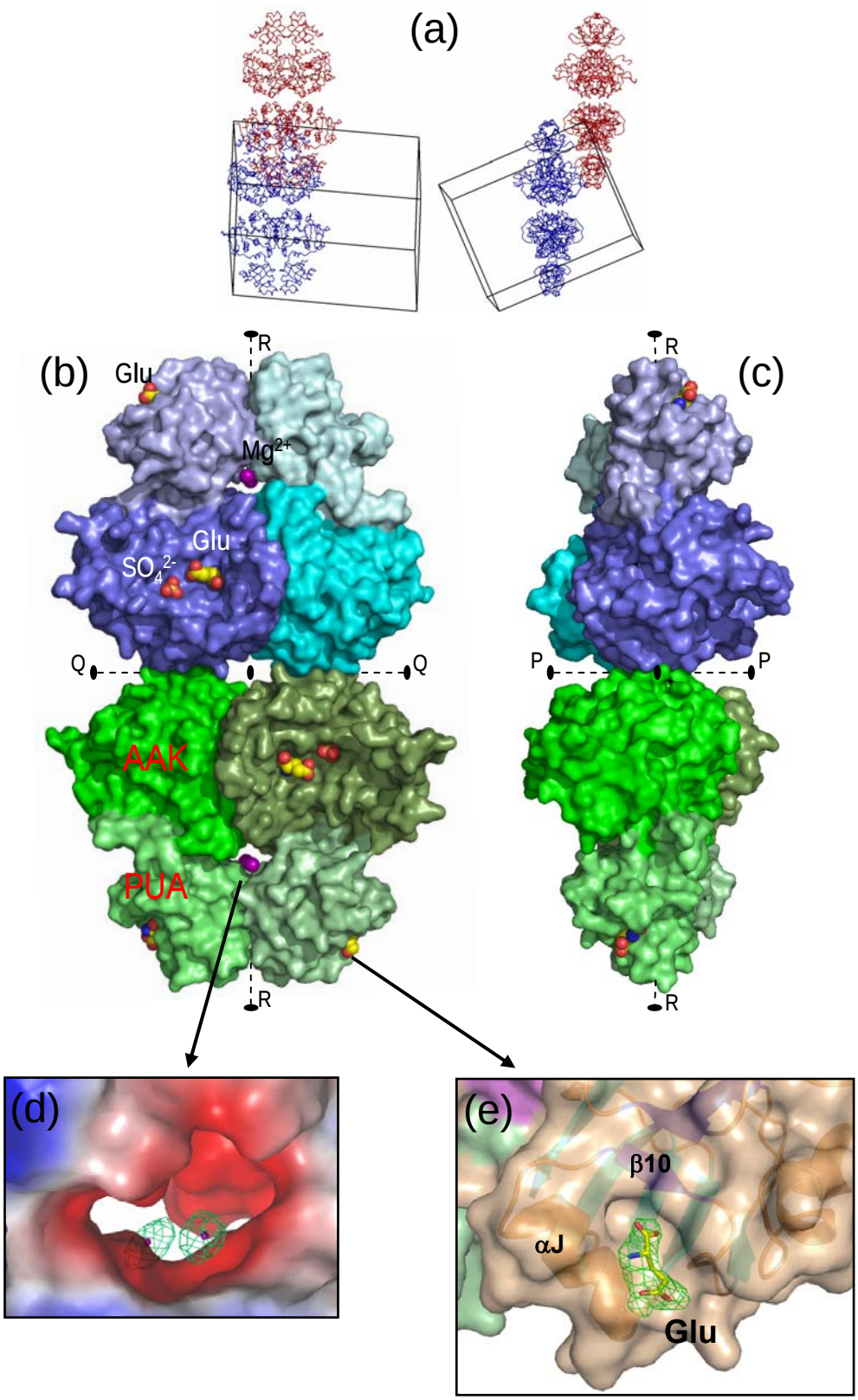


Figure 2

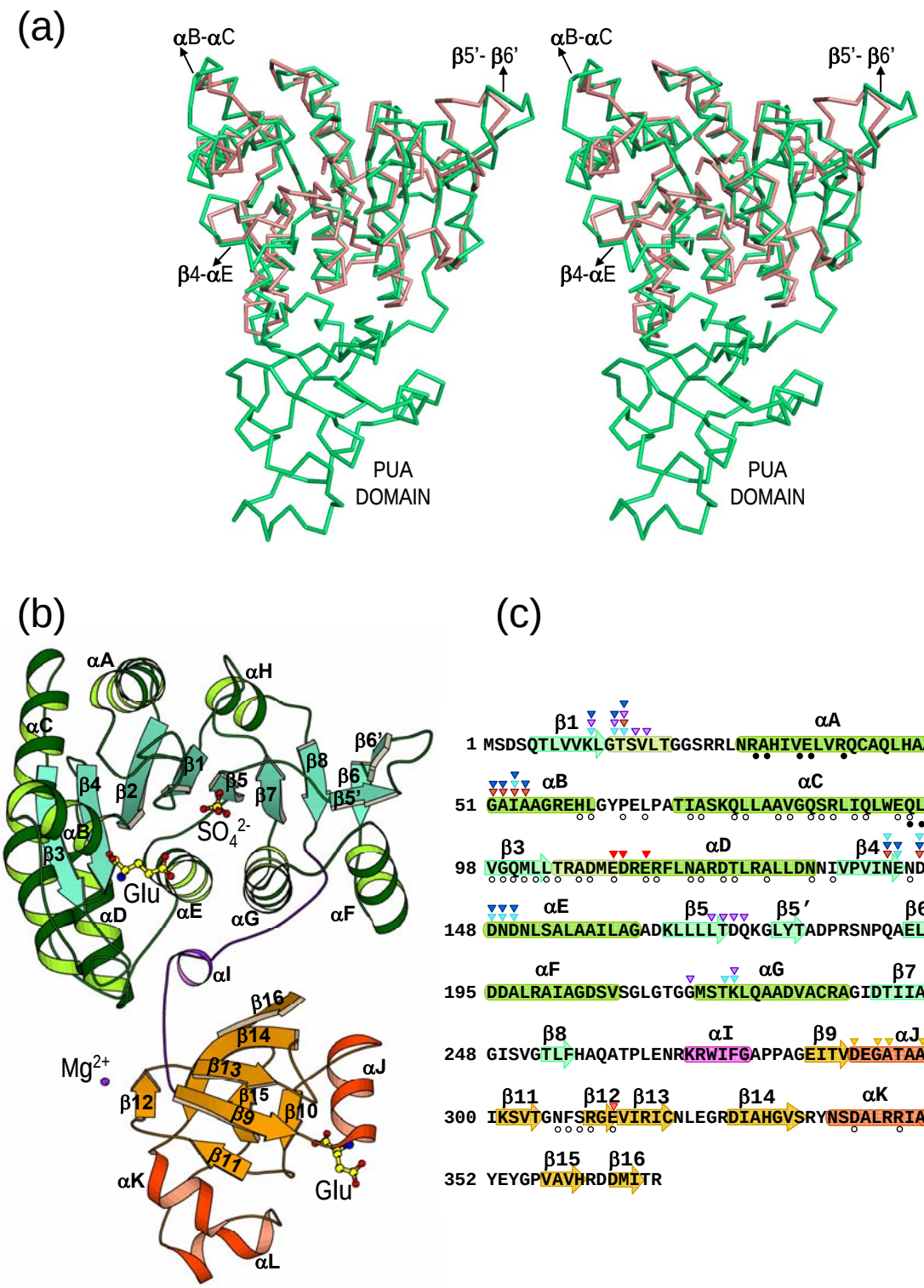

(a)
αB-αC
β5'- β6'
β4-αE
PUA
DOMAIN
αB-αC
β5'- β6'
β4-αE
PUA
DOMAIN
(b)
αA
αH
αC
β6'
β1
β8
β6
β5'
β4
β2
β5
β7
αB
β3
SO4 2-
Glu
αD
αE
αG
αF
αI
β16
β14
Mg2+
β13
αJ
β12
β15
β9
β10
β11
αK
Glu
αL
(c)
β1 αA β2
1 MSDSQTLVVKLGTSVLTGGSRRLNRAHIVELVRQCAQLHAAGHRIVIVTS
αB αC
51 GAIAAGREHLGYPELPATIASKQLLAAVGQSRLIQLWEQLFSIYGIH
β3 αD β4
98 VGQMLLTRADMEDRERFLNARDTLRALLDNNIVPVINENDAVATAEIKVG
αE β5 β5' β6' β6
148 DNDNLSALAAILAGADKLLLLTDQKGLYTADPRSNPQAELIKDVYGI
αF αG β7 αH
195 DDALRAIAGDSVSGLGTGGMSTKLQAADVACRAGIDTIIAAGSKPGVIGDVME
β8 αI β9 αJ β10
248 GISVGTLFHAQATPLENRKRWIFGAPPAGEITVDEGATAAILERGSSLLPKG
β11 β12 β13 β14 αK αL
300 IKSVTGNFSRGEVIRICNLEGRDIAHGVSRYNSDALRRIAGHHSQEIDAILG
β15 β16
352 YEYGPVAVHRDDMITR

Figure 3

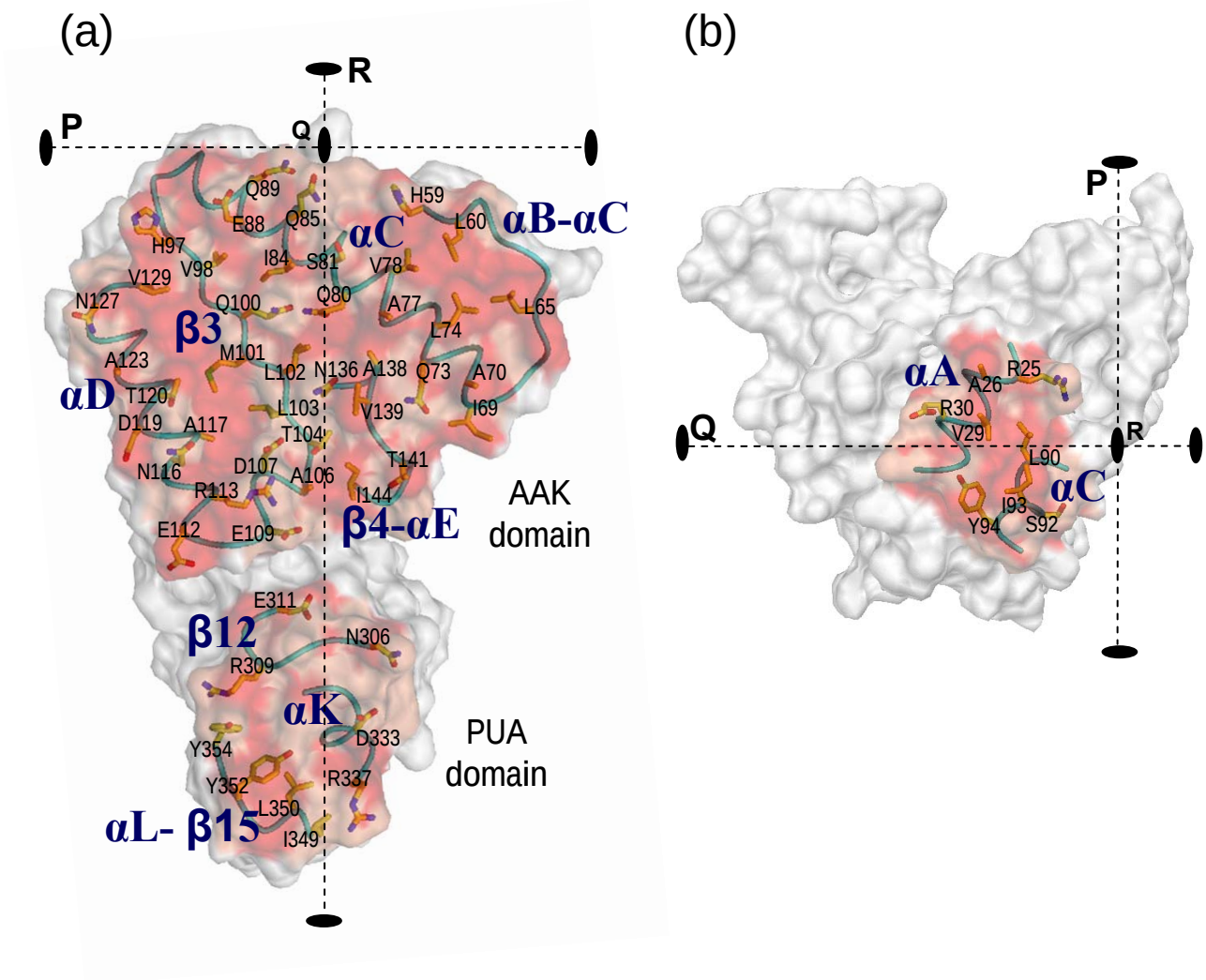

(a)
R
P
Q
Q89
H59
E88
Q85
αC
L60
αB-αC
H97
I84
S81
V78
V98
V129
Q80
N127
Q100
A77
L65
β3
L74
A123
M101
L102
N136
A138
Q73
A70
αD
T120
L103
V139
I69
D119
A117
T104
T141
N116
D107
A106
R113
I144
AAK
domain
β4-αE
E112
E109
E311
β12
N306
R309
αK
D333
PUA
domain
Y354
R337
Y352
L350
αL- β15
I349
(b)
P
αA
R25
A26
R30
Q
V29
R
L90
αC
I93
Y94
S92

Figure 4

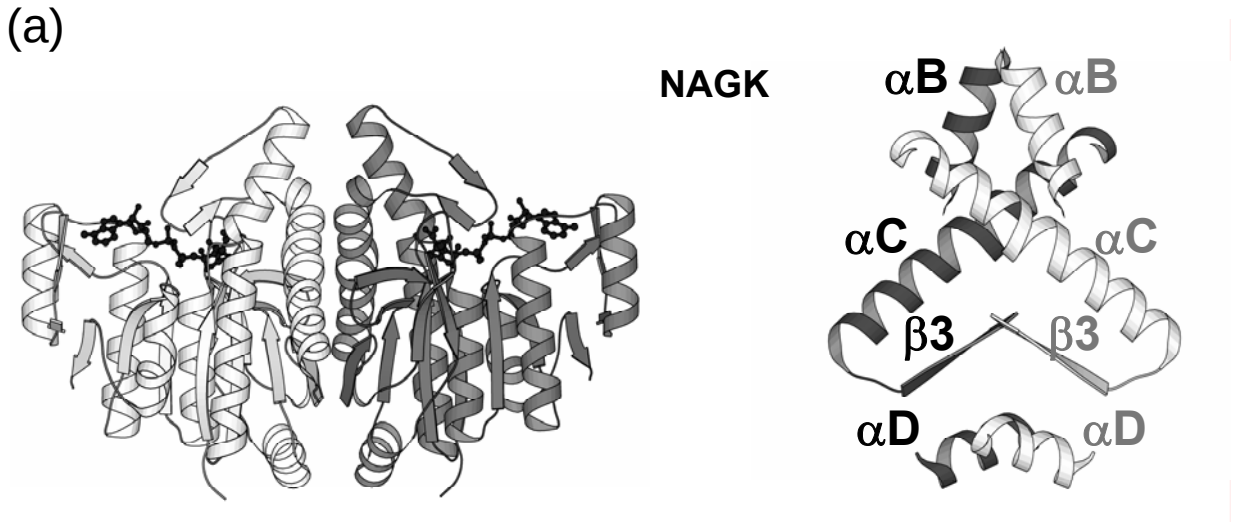

(a)
NAGK
αB
αB
αC
αC
β3
β3
αD
αD
(b)
UMPK
αB
αB
αC
αC
β3
β3
αD
αD
(c)
G5K
αC
αC
αB
αB
β3
β3
αD
αD

Figure 5

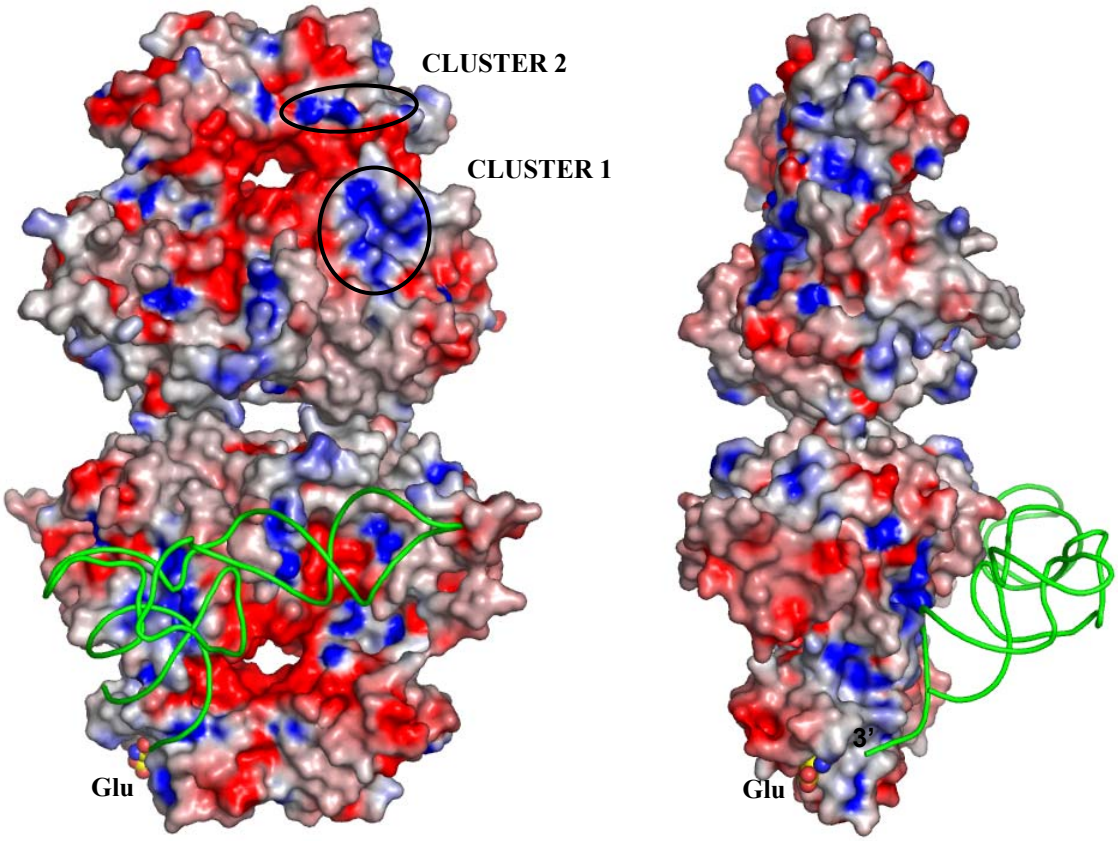


Figure 6

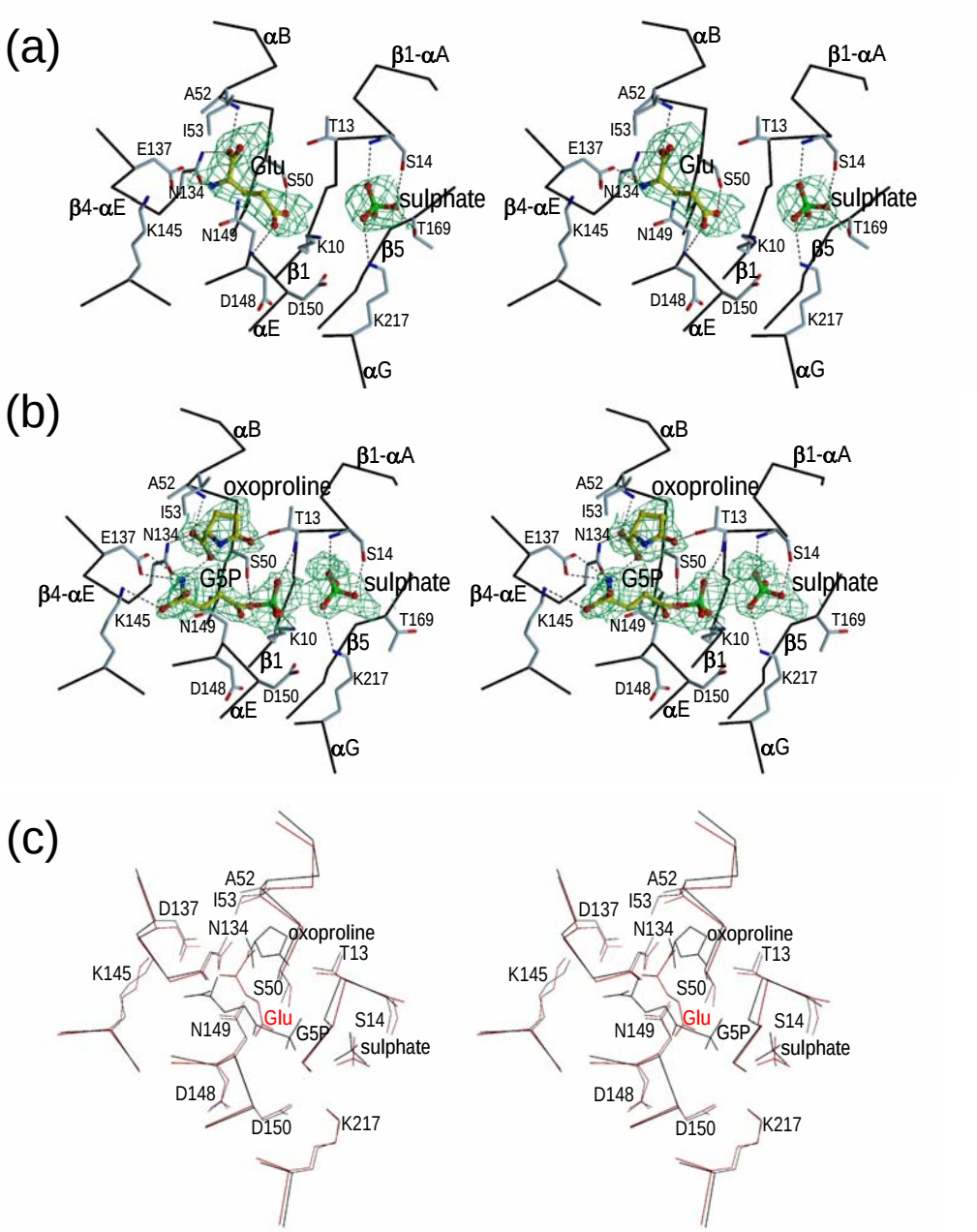
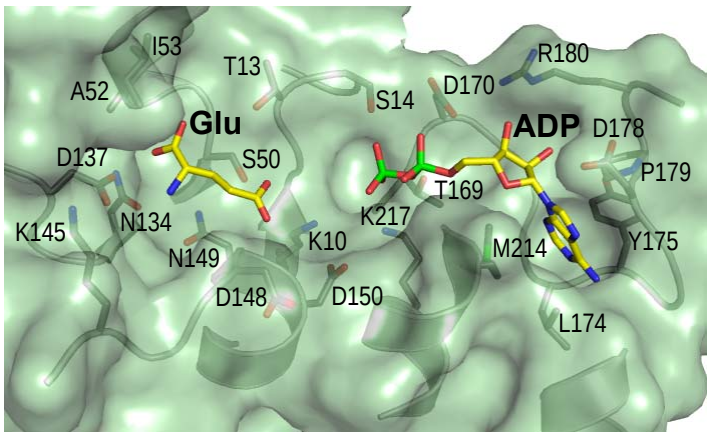

(a)
αB
β1-αA
A52
I53
T13
E137
Glu
S14
N134
S50
sulphate
β4-αE
K145
N149
T169
K10
β5
β1
D148
D150
K217
αE
αG
(b)
αB
β1-αA
A52
oxoproline
I53
T13
E137
N134
S14
S50
G5P
sulphate
β4-αE
K145
N149
T169
K10
β5
β1
D148
K217
D150
αE
αG
(c)
A52
D137
I53
N134
oxoproline
T13
K145
S50
Glu
N149
G5P
S14
sulphate
D148
D150
K217

Figure 7

(a)


I53
T13
A52
Glu
S14
D170
R180
ADP
D178
D137
S50
P179
T169
K145
N134
K217
K10
N149
M214
Y175
D148
D150
L174

(b)

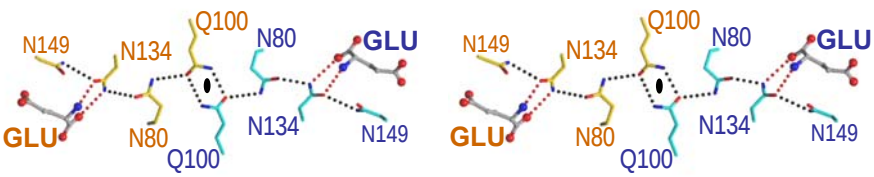

N149
N134
Q100
N80
GLU
GLU
N80
Q100
N134
N149
N149
N134
Q100
N80
GLU
GLU
N80
Q100
N134
N149

Figure 8

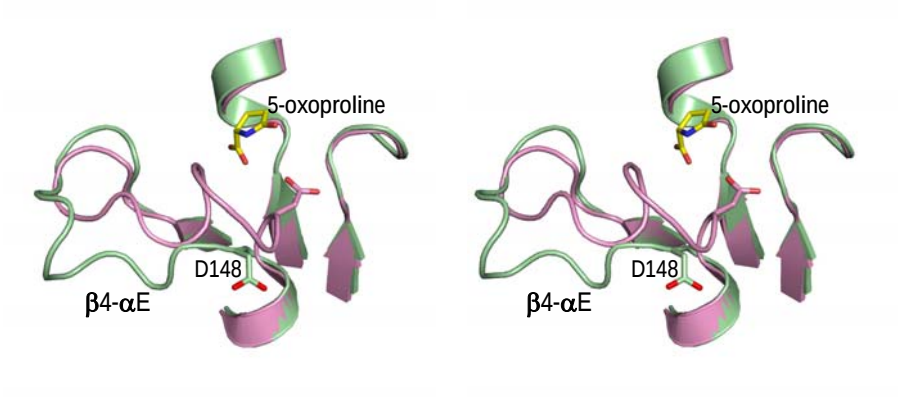


Figure 9

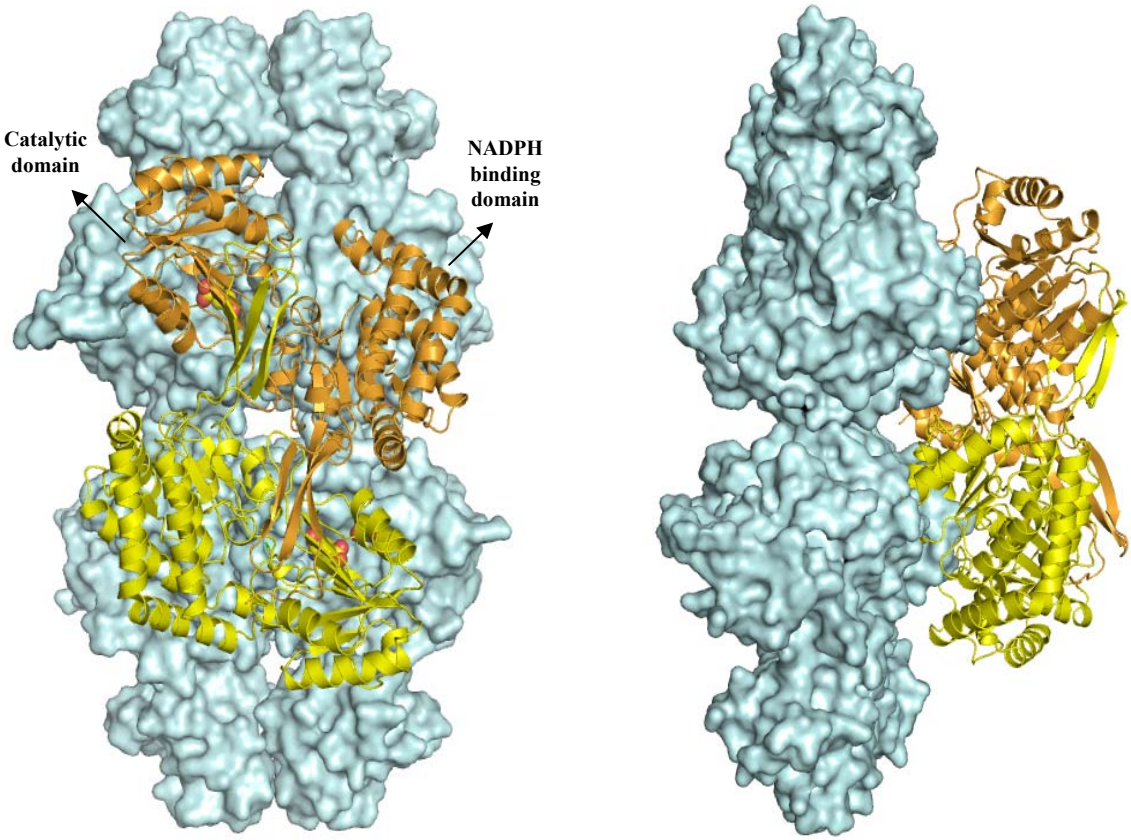


Figure 10

RESUMEN DE LOS RESULTADOS

RESUMEN DE LOS RESULTADOS

Los resultados de este trabajo se presentan en cuatro capítulos. Los tres primeros corresponden a tres artículos ya publicados y el cuarto, a otro en fase de evaluación editorial. A continuación se resume brevemente cada uno de ellos:

Capítulo 1

Site-directed mutagenesis of *Escherichia coli* acetylglutamate kinase and aspartokinase III probes the catalytic and substrate-binding mechanisms of these amino acid kinase family enzymes and allows three-dimensional modelling of aspartokinase.

Clara Marco-Marín, Santiago Ramón-Maiques, Sandra Tavárez y Vicente Rubio.

Enviado a publicar el 29 de Mayo de 2003 y aceptado el 19 de septiembre de 2003.
Publicado en *Journal of Molecular Biology* (2003) **334**, 459-476.

De este trabajo S. Ramón-Maiques y yo realizamos la clonación de AKIII, la producción de los mutantes de la N-acetil-L-glutamato quinasa (NAGK) y de la aspartoquinasa III (AKIII) de *E. coli*, la sobreexpresión y purificación de las formas silvestre y mutantes de la NAGK y de la AKIII, los ensayos enzimáticos con estos enzimas y el modelizado de la AKIII. El trabajo lo realizamos de forma conjunta y a partes iguales, sin que pueda separarse lo realizado por uno y otro. La producción del mutante delecionado de la AKIII y la caracterización de este mutante fue realizada por S. Tavárez.

La aspartoquinasa cataliza en microorganismos y plantas el primer paso en la síntesis de treonina, lisina, metionina e isoleucina. Este enzima, que no está presente en animales, es una diana potencial para el diseño de antimicrobianos y además tiene interés biotecnológico ya que es el punto de control en la síntesis de aminoácidos esenciales y al ser fuertemente inhibido por lisina en plantas, limita el contenido de este aminoácido en vegetales.

Como ya se ha indicado en la Introducción, intentamos determinar la estructura

de aspartoquinasa (AK), y entre las distintas variantes de este enzima seleccionamos la aspartoquinasa III (AKIII) de *E. coli* que es una forma simple de AK, compuesta por un sólo tipo de subunidad, de 449 aminoácidos, y que está implicada en la síntesis de lisina y es inhibida por este aminoácido. También como se ha indicado ya, a pesar de gran número de intentos no conseguimos obtener cristales adecuados de este enzima, por lo que decidimos modelizar su dominio aminoácido quinasa a partir de la estructura de N-acetil-L-glutamato quinasa (NAGK) de *E. coli*, que había sido determinada previamente en el laboratorio. Sin embargo, existiendo sólo 18 % de identidad de secuencia entre ambos enzimas, era razonable dudar de la calidad de los alineamientos, existiendo el riesgo de un modelizado erróneo. Para garantizar al máximo la calidad del alineamiento decidimos identificar residuos clave para la catálisis o para la unión de los sustratos que se correspondieran con residuos equivalentes de la NAGK y que guiaran así el alineamiento.

La estructura de la NAGK de *E. coli* había sido determinado en presencia de sus sustratos, y se habían propuesto papeles muy concretos para algunos residuos en la unión de los sustratos y en la catálisis. La estrategia seguida fue seleccionar alguno de estos residuos y realizar mutaciones en la NAGK de *E. coli* que representaran cambios leves para evitar inactivar el enzima, y así poder cuantificar el efecto producido sobre su actividad y sobre los parámetros cinéticos. Paralelamente se introdujeron mutaciones en AKIII, en los residuos correspondientes en el alineamiento, o en el sitio posible para el aspartato (propuesto por comparación con el sitio para acetilglutamato de la NAGK), y se estudió su efecto. El paralelismo de los resultados obtenidos para los distintos mutantes de la NAGK y la AKIII dio pie a suponer que el alineamiento era correcto.

Además, se predijo la estructura secundaria del dominio aminoácido quinasa de la AKIII utilizando varios programas de predicción, y se modificó el alineamiento haciendo corresponder esta predicción con los elementos correspondientes de la estructura de la NAGK. La fiabilidad de los distintos programas se corroboró prediciendo la estructura secundaria de la NAGK de *E. coli* y contrastando las predicciones con la información estructural experimental de este enzima, siendo excelente el grado de acuerdo obtenido cuando se utilizaron las predicciones en que había coincidencia con todos los programas de predicción usados.

El alineamiento mejorado se utilizó para modelizar la estructura del dominio aminoácido quinasa de la AKIII. En una primera fase se impusieron fuertes restricciones generadas a partir de la estructura modelo, y posteriormente se relajaron las restricciones para modelizar las zonas menos conservadas, principalmente algunos

lazos. Finalmente se ajustaron algunas cadenas laterales de modo manual para evitar contactos inadecuados o entornos poco favorables para algunos residuos.

Una vez construido el modelo, éste fue sometido a distintos programas de validación tal y como si se tratara de una estructura cristalográfica. Se comprobó que no habia contactos no permitidos y que el modelo presentaba buena estereoquímica, con el 87 y el 13 % de los residuos en la zona más favorable y aceptable (‘*additionally allowed*’) del diagrama de Ramachandran.

La AKIII además de presentar un dominio aminoácido quinasa, presenta una región carboxi-terminal que a juzgar por la secuencia, incluye dos dominios ACT (dominios relacionados con regulación). El modelizado de la AKIII no incluyó los dominios ACT, aunque consideramos que éstos presentan el plegamiento característico de estos dominios descrito en la treonina deaminasa y fenilalanina hidroxilasa.

Este trabajo incluye la preparación de un mutante delecionado de la AKIII que carece de los 151 residuos del extremo carboxi-terminal. Este mutante fue preparado por S. Tavárez, de nuestro laboratorio, pero los resultados obtenidos con él, se han tomado en consideración para derivar las conclusiones de este trabajo.

A continuación se incluye un resumen formal de los resultados y conclusiones obtenidos: En este trabajo hemos probado usando mutagénesis dirigida, predicciones basadas en la estructura cristalina de la N-acetil-L-glutamato quinasa de *E. coli*, el paradigma de la familia aminoácido quinasa, acerca del papel de determinados residuos en la unión de los sustratos y en la catálisis. Las mutaciones K8R y D162E disminuyeron $V_{[sustrato]=8}$ 100 y 1000 veces respectivamente, de acuerdo con las predicciones de que K8 cataliza la transferencia del grupo fosforilo y que D162 organiza los grupos catalíticos. R66K y N158Q incrementaron de forma selectiva la K_m^{Asp} de 3 a 4 órdenes de magnitud, de acuerdo con la unión de R66 y N158 a los sustituyentes del Ca del acetilglutamato. Mutagénesis en paralelo de la aspatoquinasa III (AKIII fosforila aspartato en vez de glutamato), otro miembro importante de la familia aminoácido quinasa de estructura desconocida permitió identificar en la AKIII dos residuos, K8 y D202, que parecen desempeñar papeles similares a los de K8 y D202 de NAGK, y apoya la implicación de E119 y R198, de forma similar a R66 y N158 de NAGK, en la unión del sustrato aminoacídico, aparentemente interaccionando respectivamente con el NH_3^+ y a-COO^- del aspartato. Estos resultados y un alineamiento mejorado de las secuencias de NAGK y AKIII nos han guiado en el modelizado del dominio aminoácido quinasa de AKIII usando NAGK como molde. La AKIII es homidimérica y es inhibida por lisina. Probablemente la lisina se una a un

dominio regulador, localizado en la región C-terminal de AKIII. Hemos hecho una deleción C-terminal en AKIII ($AKIII_t$) que muestra que la región C-terminal está implicada en las interacciones intersubunidad, ya que $AKIII_t$ es monomérica. Además, es inactiva tal como correspondería si el dímero fuera esencial para la catálisis. De acuerdo con estos resultados se proponen varios modelos para la arquitectura de AKIII.

Capítulo 2

First-time crystallization and preliminary X-ray crystallographic analysis of a bacterial-archaeal type UMP kinase, a key enzyme in microbial pyrimidine biosynthesis.

Clara Marco-Marín, Juan Manuel Escamilla-Honrubia y Vicente Rubio.

Enviado a publicar el 1 de octubre de 2004 y aceptado el 17 de noviembre de 2004.
Publicado en *Biochim. Biophys. Acta.* (2005) **1747**, 271-275.

La clonación del gen codificante para la UMP quinasa (UMPK) de *P. furiosus*, la sobreexpresión y la purificación de este enzima fueron realizadas por J. M Escamilla-Honrubia, así como los primeros experimentos de cristalización, realizados con mi asesoramiento. Luego realicé algunas variaciones en las condiciones de cristalización, tras las cuales conseguimos cristales de mayor tamaño y calidad y también llevé a cabo la difracción de los cristales y el procesado de los datos de difracción.

La UMPK cataliza en bacterias la fosforilación del UMP, un paso clave en la síntesis de pirimidinas ya que a partir del UDP se sintetizan todos los nucleotidos di y trifosfato de pirimidinas. En eucariotas esta reacción está catalizada por un enzima completamente diferente, la CMP/UMP quinasa y por esta razón, así como por su importancia funcional, la UMPK bacteriana es una diana potencial para el diseño de antimicrobianos.

El posible gen *pyrH* de *P. furiosus* fue amplificado a partir de DNA genómico de este microorganismo y clonado en el plásmido de expresión pET15b. El plásmido resultante fue introducido en células BL21(DE3) junto con el plásmido pSJS1240, que codifica para varios RNAs de transferencia poco frecuentes en *E. coli* y las células

transformadas se utilizaron para sobreexpresar la UMPK de *P. furiosus*. El enzima recombinante se purificó con un grado de pureza mayor del 90% en unos pocos pasos: rotura de las células mediante sonicación, centrifugación para clarificar el extracto, calentamiento corto para eliminar proteínas no termoresistentes y dos pasos de cromatografía de intercambio iónico.

Se realizaron ensayos enzimáticos mediante los cuales se determinó que la proteína purificada era realmente una UMPK altamente específica para UMP.

Se prepararon cristales de este enzima, que crecieron en 3-4 dias a 21 ºC usando como precipitante una solución 3.5-4 M de formato sódico. Estos cristales, de apariencia cúbica, se congelaron en una solución crioprotectora preparada añadiendo 10 % glicerol a la solución de cristalización, y se difractaron en el sincrotrón ESRF de Grenoble hasta resolución de 2.4 ? . Estos datos se utilizaron para resolver la estructura de este enzima, como se describe en el capitulo siguiente.

El procesado de los datos de difracción indicó que estos pertenecen al grupo espacial I23 con parámetros de celdilla a= b= c= 144.94 Å, a= ß= ?= 90º. A partir del volumen de celdilla y de la masa molecular de la proteína, se estimó la presencia de dos subunidades por unidad asimétrica, que podrían corresponder a un homodímero.

Los resultados de la función de rotación sugirieron la formación de hexámeros, lo cual era coincidente con los datos bioquímicos disponibles para la UMPK de *E.coli*. Se observó que en las diferentes secciones de la función de rotación, además de los picos correspondientes al grupo espacial I23, aparecía un pico correspondiente a un eje de simetría binaria, orientado perpendicularmente al eje ternario cristalográfico. Estos resultados eran compatibles con la formación de hexámeros organizados como trímeros de dímeros en los que el eje ternario coincidiría con el eje ternario cristalográfico, y las subunidades de cada uno de los dímeros estarían relacionadas a través del eje binario no cristalográfico. Como se verá en el capítulo siguiente la organización hexamérica de UMPK se corresponde con esta propuesta realizada a partir de los datos de difracción.

Capítulo 3

The crystal structure of *Pyrococcus furiosus* UMP kinase provides insight into catalysis and regulation in microbial pyrimidine nucleotide biosynthesis.

Clara Marco-Marín, Fernando Gil-Ortiz y Vicente Rubio.

Enviado a publicar el 28 de mayo de 2005 y aceptado el 14 de julio de 2005.
Publicado en *Journal of Molecular Biology* (2005) **352**, 438-454.

Soy responsable principal del contenido completo de este trabajo.

Al principio del resumen del capítulo anterior se indica la importancia de la UMPK, enzima (de *P. furiosus*) que había sido cristalizado y del que disponíamos datos de difracción de rayos X.

En este capitulo se describe la determinación y el análisis de la estructura de este enzima, que es hexamérico y que hemos caracterizado estructuralmente en su forma libre y como complejo con el análogo inerte del ATP, AMPPNP, y con dicho análogo y su sustrato UMP.

Con el fin de obtener las fases cristalográficas, se preparó el enzima con sustitución de las metioninas por selenometioninas, y para ello se transformaron células de la cepa de *E. coli* B834 (DE3) pLysS, auxótrofa para el aminoácido metionina, con el plásmido de expresión de la UMPK de *P. furiosus* cuya obtención se describe en el capítulo anterior y con el plásmido pSJS1240, que codifica para varios RNAs de transferencia poco frecuentes en *E. coli*. Se hizo crecer estas células en un medio en el que se sustituyó la metionina por selenometionina, y se sobreexpresó la proteína recombinante por inducción con IPTG. La purificación del enzima y su cristalización se realizaron siguiendo los mismos protocolos que se habían utilizado anteriormente para la proteína nativa. Los cristales fueron difractados en el sincrotrón BESSY de Berlín a tres longitudes de onda, calculándose las fases mediante la técnica de dispersión anómala múltiple (MAD).

La estructura de este enzima muestra que éste no se parece a otras nucleósido monofosfato quinasas, incluida la CMP/UMP quinasa presente en organismos eucarióticos. El enzima forma hexámeros que están organizados como trímeros de dímeros. Cada una de las subunidades presenta el mismo plegamiento básico y topología que los otros dos enzimas de la familia aminoácido quinasa previamente caracterizados, carbamato quinasa (CK) y acetilglutamato quinasa (NAGK).

Hemos determinado también la estructura de dos complejos de UMPK a partir de cristales crecidos en ausencia de sustratos, que fueron empapados en soluciones con AMPPNP, un análogo estructural del ATP, o con AMPPNP y UMP. La comparación de UMPK y NAGK revela que ambos enzimas unen sus sustratos en lugares equivalentes. Una de las principales diferencias entre ambos enzimas es la presencia de un lazo largo

en UMPK cuya conformación depende del contenido del centro activo del enzima, y que en presencia de AMPPNP y UMP cubre al UMP y al sitio de transferencia del grupo fosforilo. En UMPK hay dos iones magnesio en el centro activo, que participan en la unión del polifosfato del AMPPNP, y uno de éstos desempeña el mismo papel que una lisina de NAGK que estabiliza el estado de transición, y que está ausente en UMPK.

Al igual que la CK y la NAGK, la UMPK forma dímeros pero el modo de asociación de las subunidades en la UMPK es radicalmente distinto, ya que a pesar de que los elementos que participan en la superficie de dimerización son básicamente los mismos que en la CK y la NAGK, la orientación de las subunidades varía sustancialmente. Como se verá en el capítulo siguiente el cambio en la orientación de las subunidades que forman el dímero es todavía más extremo en la glutamato 5-quinasa.

La organización hexamérica de la UMPK es también un rasgo novedoso. Los hexámeros se forman por contacto entre dímeros adyacentes, que establecen distintas superficies de contacto entre sí. Paralelamente al trabajo realizado con UMPK, se ha determinado también en nuestro laboratorio la estructura de dos acetilglutamato quinasas inhibibles por arginina, que también presentan organización hexamérica. Sin embargo los hexameros de UMPK y NAGK son completamente diferentes, ya que los contactos entre subunidades de distintos dímeros están mediados por distintas superficies en ambos hexámeros. Los hexámeros de NAGK son trímeros de dímeros en los que cada dímero es equivalente al encontrado para la CK y para la NAGK no inhibible de *E. coli*. Los hexámeros de NAGK se forman gracias a la presencia de una hélice amino terminal extra presente en las subunidades de las NAGK inhibibles, que permite el engarce entre distintos dímeros. La UMPK carece de hélice amino terminal y las interacciones entre dímeros se establecen por contactos extensos entre distintas subunidades. De hecho estas superficies de contacto son comparables con la superficie de dimerización.

Capítulo 4

A novel two-domain architecture within the amino acid kinase enzyme family revealed by the crystal structure of *Escherichia coli* glutamate 5-kinase.

Clara Marco-Marín, Fernando Gil-Ortiz, Isabel Pérez-Arellano, Javier Cervera, Ignacio Fita y Vicente Rubio.

Enviado a *Journal of Molecular Biology* el 22 de Noviembre de 2006.

En este capitulo se incluye la determinación de la estructura tridimensional de la glutamato 5-quinasa (G5K) de *E. coli*, que se ha resuelto a partir de dos tipos de cristales de este enzima. Ambos tipos de cristales fueron producidos por I. Perez-Arellano y F. Gil-Ortiz. Yo he realizado el cálculo de las fases, la construcción y afinado de la estructura y el análisis de ésta.

La glutamato 5-quinasa (G5K) es el enzima que cataliza el primer paso en la sintesis de prolina, que es también el primer paso en la síntesis de arginina en mamíferos. Este enzima es el punto de regulación de la ruta, siendo inhibido por prolina en microorganismos y plantas y por ornitina en animales. La síntesis de prolina no es únicamente importante desde el punto de vista biosintético sino también por las propiedades osmoprotectoras de esta molécula. Además el estudio de G5K tiene interés clínico ya que ha sido identificado un déficit congénito en humanos que causa hiperamonemia y que se debe a una mutación en este enzima.

La estructura de la G5K de *E. coli* se resolvió a partir de dos tipos de cristales, que difractaron a 2.5 o 2.9 Å de resolución respectivamente. Las fases cristalográficas se calcularon mediante la técnica de reemplazo molecular utilizando como modelo de búsqueda el dímero de G5K de *C. jejuni*. La G5K de *C. jejuni* carece del dominio PUA, dominio que sí tienen la G5K de *E. coli* y de la mayoría de bacterias. Los dominios PUA también se encuentran en enzimas modificadoras de RNA, como la arqueosina tRNA-guanina transglicosilasa, a partir de cuya estructura se seleccionó la región correspondiente al dominio PUA, que se utilizó como molde para construir el dominio PUA de la G5K de *E. coli*.

La estructura de cada una de las subunidades de G5K revela la organización de los dominios AAK y PUA, que están unidos mediante una a hélice. El dominio aminoácido quinasa presenta el plegamiento básico y la topología del resto de enzimas de la familia aminoácido quinasa previamente caracterizados, carbamato quinasa (CK), acetilglutamato quinasa (NAGK) y UMP quinasa (UMPK), constituyendo al igual que UMPK una versión simplificada de este plegamiento. El dominio PUA está formado por un sandwich $ß_5ß_4$ y tres a hélices, una de las cuales está ausente en otros dominios PUA.

El dominio aminoácido quinasa provee de los elementos necesarios para la unión de los sustratos y para la catálisis. En cada una de las subunidades de ambos tipos de cristal se ha identificado un glutamato unido en el dominio aminoácido quinasa. En el cristal de mayor resolución, encontramos una región de densidad electrónica estrechamente unida al grupo ?-carboxilato del glutamato, indicando que en algunas

moléculas este sustrato está fosforilado.

La comparación de G5K con NAGK y UMPK revela que la unión del sustrato fosforilable (glutamato, acetilglutamato o UMP respectivamente), es equivalente en los diferentes enzimas, con diferencias derivadas de la distinta naturaleza de estos ligandos. El sitio de unión del glutamato es más amplio que el del acetilglutamato en la NAGK o el del UMP en la UMPK, y está totalmente expuesto al solvente, siendo posiblemente ésta la causa de que la G5K presente una K_m muy elevada para este sustrato.

Aunque no hemos obtenido cristales de glutamato 5-quinasa en presencia de prolina, los datos de mutagénesis dirigen la unión de este inhibidor a una región muy concreta de la proteína, cercana al sitio de unión del glutamato y también de la superficie de dimerización. En el cristal de mayor resolución, encontramos una densidad junto al sitio de unión del glutamato que posiblemente corresponda con una molécula de oxoprolina y cuya localización podría apuntar al sitio de unión del inhibidor.

El homodímero de G5K se forma por interacciones extensas y mutuas entre los dominios aminoácido quinasa y los dominios PUA de las dos subunidades. La superficie del dominio aminoácido quinasa de G5K que participa en la formación del homodímero, es equivalente a la superficie de dimerización de CK, NAGK y UMPK. Sin embargo, existen importantes diferencias en la organización de los distintos dímeros, debidas a cambios de orientación en las subunidades que lo forman.

La G5K es un tetrámero organizado como un dímero de dímeros en el que únicamente los dominios aminoácido quinasa participan en los contactos interdiméricos. El tetrámero de G5K es planar y elongado y en éste los cuatro dominios aminoácido quinasa quedan en posición centrada mientras que los dos pares de dominios PUA quedan en las posiciones distales. La organización de las subunidades en el tetrámero de la G5K permite proponer un modo de interacción entre este enzima y la glutamilfosfato reductasa, el enzima que cataliza el siguiente paso de la ruta. La asociación de estos dos enzimas que en animales forman parte de un mismo polipéptido, permitiría la canalización del glutamilfosfato entre los dos centros activos, impidiendo de este modo que esta molécula, altamente inestable se cicle antes de ser utilizada.

DISCUSIÓN GENERAL

DISCUSIÓN GENERAL

El conocimiento estructural adquirido representa a todos los enzimas de la familia menos a la acetilglutamato sintasa.

Si al comienzo de este trabajo de Tesis los conocimientos estructurales sobre el grupo de enzimas *E.C.* 2.7.2. se restringían, dentro de la familia aminoácido quinasa, a la carbamato quinasa y la acetilglutamato quinasa no inhibible por arginina (Marina et al., 1999; Ramón-Maiques et al., 2000; Ramón-Maiques et al., 2002), y, fuera de esta familia, a la fosfoglicerato quinasa (Watson et al., 1982), nuestro trabajo, junto con otros resultados de nuestro grupo y con datos de otros grupos, han permitido ampliar mucho nuestro conocimiento estructural de los miembros de la familia aminoácido quinasa y del grupo *E.C.* 2.7.2. Restringiéndonos a la familia aminoácido quinasa, nosotros hemos determinado las estructuras cristalinas de la UMP quinasa de *P. furiosus* y de la glutamato 5-quinasa de *E. coli*, y también hemos adelantado en el conocimiento de la aspartoquinasa, modelizando su dominio aminoácido quinasa. Otros componentes de nuestro grupo han determinado la estructura cristalina de la acetilglutamato quinasa inhibible por arginina (Ramón-Maiques et al., 2006), y otros grupos ajenos a nuestro laboratorio han publicado las estructuras de las aspartoquinasas de *Arabidopsis thaliana* (Mas-Droux et at., 2006) y III de *E. coli* (Kotaka et al., 2006), además de haber determinado, paralelamente a nuestra determinación, la estructura de la UMP quinasa de *E. coli* (Briozzo et al., 2005). Por otro lado, y aunque sin publicación ni análisis, se ha depositado también en el Protein Databank (PDB) la estructura de la glutamato 5-quinasa carente de dominio PUA de *Campylobacter jejuni*, como complejo con ADP, lo que nos ha permitido complementar nuestros datos obtenidos con el enzima equivalente de *E. coli*, éste sí con dominio PUA y con glutamato unido, pero carente de nucleótido en el centro activo. En resumen, sólo uno de los enzimas de la familia aminoácido quinasa, la acetilglutamato sintasa, no ha podido ser resuelto estructuralmente aún, aunque nuestro laboratorio ha propuesto una estructrura grosera para el mismo así como la identificación del sitio para arginina sobre la base de la estructura de la acetilglutamato quinasa sensible a arginina, enzima homólogo a la sintasa (Ramón-Maiques et al., 2006). En cuanto al resto de los componentes del grupo *E.C.* 2.7.2., su conocimiento se ha ampliado también considerablemente con la determinación de las estructuras de acetato quinasa, propionato quinasa y butirato quinasa, quedando por determinar únicamente la estructura de la formiato quinasa. Por tanto, nuestra información es ya suficiente para establecer generalizaciones que serán posiblemente

válidas para todos los enzimas de la familia aminoácido quinasa, y también para establecer comparaciones e identificar diferencias clave con los otros enzimas del grupo *E.C.* 2.7.2, y, en el caso de la UMP quinasa, con la UMP/CMP quinasa eucariótica.

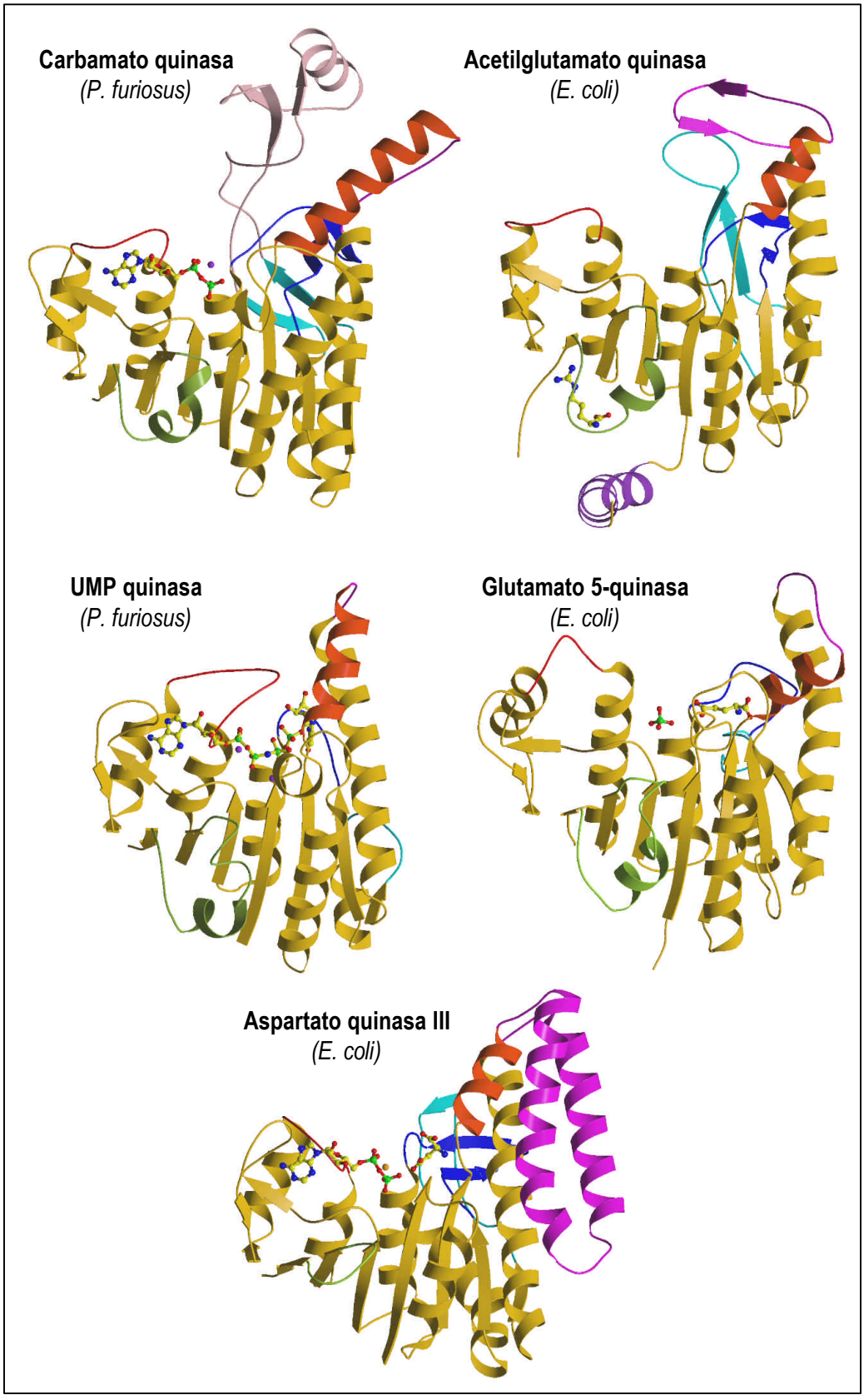


Figura 1. (página anterior) Representación en modelo de cintas del dominio aminoácido quinasa de la carbamato quinasa de *Pyrococcus furiosus* en presencia de MgADP (PDB 1E19), la acetilglutamato quinasa (NAGK) de *Thermotoga maritima* en presencia de arginina (PDB 2BTY), la UMP quinasa de *Pyrococcus furiosus* con AMPPNPMg_2 y UMP (PDB 2BMU), la glutamato 5-quinasa de *E. coli* en presencia de glutamato y sulfato (PDB 2J5T) y la aspartoquinasa III de *E. coli* con aspartato y ADP (PDB 2JOW). Los elementos comunes a todos estos enzimas están coloreados en amarillo y los elementos diferenciales en distintos colores: 1) morado para la hélice N-terminal presente en las NAGK inhibibles, 2) naranja para aB, 3) magenta para los elementos entre aB y aC, 4) cyan para la unión ß3-aD, 5) rosa para el subdominio lejano de carbamato quinasa, 6) azul para la unión ß4-aD, 7) rojo para el lazo aF-aG y 8) verde para la unión ß7-ß8, que incluye a aH en la mayoria de los enzimas.

Constancia del plegamiento básico.

La primera generalización concerniente a la familia aminoácido quinasa es la constatación de la constancia del plegamiento de la cadena polipeptídica que caracteriza al dominio aminoácido quinasa (Figura 1). En el caso de la acetilglutamato sintasa, el único miembro de estructura todavía por determinar, el alineamiento del dominio correspondiente de este enzima con la acetilglutamato quinasa de *Thermotoga maritima*, (Figura 2), es fuertemente indicativo de la identidad estructural de este dominio en la sintasa, con conservación probable de todos los elementos de la estructura de la quinasa, con sólo dos elementos diferenciales de escaso impacto en la estructura del dominio: una vuelta más en la hélice más larga del dominio, αC, y una longitud mayor del lazo αF-αG, lazo que presenta variabilidad importante entre los distintos enzimas de la familia, como refleja en forma extrema el lazo correspondiente elongado de la UMPK de *P. furiosus*.

```
ARGB THEMA    ----MRIDTVNVLLEALPYIKEFYGKTFVIKFGGSAMKQENAKKAFIQDIILLKYTGIKP 56
ARGA ECOLI    MVKERKTELVEGFRHSVPYINTHRGKTFVIMLGGEAIEHENFS-SIVNDIGLLHSLGIRL 59
                   : : *: : .::***: . ****** :**.*:::** . ::::** **:  **:

ARGB THEMA    IIVHGGGPAISQMMKDLGIEPVFKNGHRVTDEKTMEIVEMVLVG---KINKEIVMNLNLH 113
ARGA ECOLI    VVVYGARPQIDANLAAHHHEPLYHKNIRVTDAKTLELVKQAAGTLQLDITARLSMSLNNT 119
              ::*:*. * *.  :     **:::::. **** **:*:*: .      .*. .: *.**

ARGB THEMA    GGRAVGICGKDSKLIVAEKETKHG--DIGYVGKVKKVNPEILHALIENDYIPVIAPVGIG 171
ARGA ECOLI    PLQGAHINVVSGNFIIAQPLGVDDGVDYCHSGRIRRIDEDAIHRQLDSGAIVLMGPVAVS 179
                :.. *   ..::*:*:    ..  *  : *:::::: : :*  ::.. * ::.**.:.

ARGB THEMA    EDGHSYNINADTAAAEIAKSLMAEKLILLTDVDGVLKD-GKLISTLTPDEAEELIRDGTV 230
ARGA ECOLI    VTGESFNLTSEEIATQLAIKLKAEKMIGFCSSQGVTNDDGDIVSELFPNEAQARVEAQEE 239
                *.*:*:.::  *:::* .* **:* : . :** :* *.::* * *:**:  :.

ARGB THEMA    TG----GMIPKVECAVSAVRGGVGAVHIINGGLEHAILLEIFSRKGIGTMIKELEG 282
ARGA ECOLI    KGDYNSGTVRFLRGAVKACRSGVRRCHLISYQEDGALLQELFSRDGIGTQIVMESA 295
              .*    * :  :. **.* *.**   *:*.   : *:* *:***.**** *   ..
```

Figura 2. Alineamiento (CLUSTALW, Thompson et al., 1994) de la acetilglutamato quinasa de *Thermotoga maritima* (NAGKTM) (ArgB, Swissprot Q9X2A4) con el dominio aminoácido quinasa de la acetilglutamato sintasa de *E. coli* (ArgA; Swissprot P0A6C5). Los elementos de estructura secundaria en NAGKTM (PDB 2BTY) se indican como flechas (hebras ß) y cilindros (a hélices). Figura modificada a partir de Ramon-Maiques *et al.* (2006) *J.Mol.Biol.* 356, 695-713, con permiso de los autores.

A partir de esta comparación de todos los enzimas de la familia puede concluirse con seguridad que la estructura básica del dominio aminoácido quinasa (Figura 3) está configurada por un sandwich αβα, en el que la hoja β central abierta se compone de ocho hebras β, seis de ellas paralelas y dos, localizadas en un extremo de la hoja, paralelas entre sí y antiparalelas a las anteriores. Esta hoja β está emparedada entre dos capas de hélices α, una con dos elementos constantes (hélices A y C) y la otra con cuatro elementos (D, E, G y F). Además, y muy característicamente, el plegamiento incluye una α hélice (αB) que prolonga la hoja β en su borde C-terminal, de la que retorna el polipéptido a la hoja principal a través de αC, hélice que presenta una extensión por el lado correspondiente al borde C-terminal de la hoja β que hace plantearse si dicha extensión fue en el pasado una hélice independiente, ya que en carbamato quinasa y acetilglutamato quinasa αC presenta una vuelta de hélice anormalmente amplia en la unión entre las dos porciones (la que sobresale y la que es paralela a la hoja β) de esta hélice. Otros elementos de la estructura que son característicos y constantes son amplios lazos que conectan β4 y αE así como β5 y β6, incluyendo en este último caso una horquilla β (β5'-β6'); y el lazo que une las hélices antiparalelas entre sí αF y αG, lazo de extensión variable, pero que limita el sitio del nucleótido y que en la UMPK de *P. furiosus* cubre casi enteramente el UMP unido al enzima. Finalmente, en todos los enzimas de la familia excepto la aspartoquinasa existe entre β7 y β8 una α-hélice corta (αH) que en aspartoquinasa se ha acortado hasta el extremo de haberse convertido en un lazo no helicoidal.

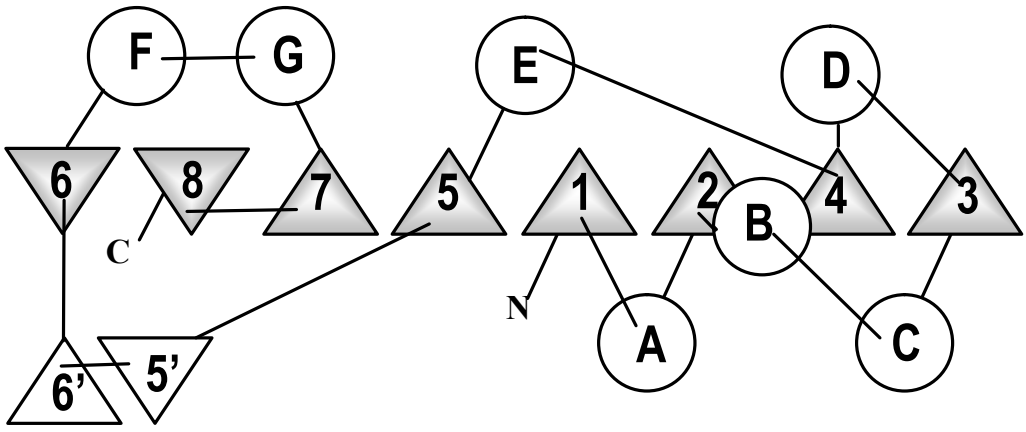


Figura 3. Topologia del dominio aminoácido quinasa común a las carbamato quinasas, acetilglutamato quinasas, aspartoquinasas, glutamato 5-quinasas y UMP quinasas de estructura conocida. Los círculos representan las a hélices y los triangulos los elementos ß. Los elementos ß sombreados son los que forman la hoja ß central.

Existencia de dos dominios.

Una característica estructural conservada en todos los miembros de la familia, y que resulta en parte de la aposición en mitad de la hoja β de dos hebras paralelas pero no adyacentes, (β1 y β5), es la posibilidad de considerar que la estructura está compuesta por dos dominios, uno N-terminal, nucleado por los elementos β1 al β4, y que incluye las α hélices A-E, y otro C-terminal, que incluye el resto de los elementos de la subunidad. La consideración de estas porciones de la estructura como dominios separados se sustenta en la unión por el dominio N-terminal del sustrato a fosforilar y por el C-terminal de la mayor parte del nucleótido (en el caso de la acetilglutamato sintasa el dominio N-terminal acomodaría el glutamato y el C-terminal la porción ADP del acetilCoA). Sustenta también la existencia de estos dos dominios el hecho de que puedan moverse en bloque uno con respecto al otro, como se ha demostrado con la acetilglutamato quinasa inhibible por arginina (Ramón-Maiques et al., 2006). En todo caso, esta organización bidominio, con protuberancias a ambos extremos del borde C-terminal de la hoja β central causadas por αB y αC en un extremo y por los lazos β5-β6 y αF-αG en el otro extremo, confieren constantemente al dominio aminoácido quinasa un aspecto en nido en el que la cavidad central del mismo se ubica en el punto de encuentro de los dos dominios, siendo en esta cavidad donde tiene lugar la transferencia del fosforilo (Figura 4).

Elementos de diversidad estructural.

Aunque compartiendo esta estructura básica común, hay diferencias entre los distintos enzimas de la familia en los tamaños de los elementos que componen dicha estructura básica (por ejemplo, hay importantes diferencias en los tamaños de αB y de β3) y, sobre todo, se dan las siguientes diferencias principales adicionales (Figura 1):

1) Una α-hélice extra N-terminal en la acetilglutamato quinasa sensible a arginina, que se usa para formar el hexámero. Con casi completa seguridad (dado el alineamiento de secuencias) está presente también en acetilglutamato sintasa.

2) La hélice B, aunque constante, difiere en su longitud y orientación, lo que afecta a la configuración y exposición del sitio para el sustrato fosforilable y a la arquitectura del dímero, e incluso puede tener implicaciones para la catálisis, como en el caso de la acetilglutamato quinasa, en la que el extremo positivo del dipolo de la hélice B colabora a neutralizar la carga negativa en el grupo fosforilo que se transfiere.

3) El lazo αB y αC es largo y forma una horquilla β en acetilglutamato quinasa, horquilla que recubre el sitio para acetilglutamato. Una horquilla similar posiblemente esté presente en acetilglutamato sintasa. En aspartoquinasa la conexión entre αB y αC es mucho más larga y forma una hebra de dos α-hélices largas, antiparalelas entre sí, que se orientan hacia la parte externa de la subunidad, formando con las hélices B y C un haz de cuatro α-hélices que cubre las superficies de αB y αC. Hay que hacer notar que la superficie cubierta por estas dos nuevas α hélices participa en otros miembros de la familia en la formación de dímeros. De este modo, las dos hélices nuevas de aspartoquinasa cierran dicha superficie y abolen la formación de dímeros por esta zona.

4) La unión β3-αD es larga en acetilglutamato quinasa, carbamato quinasa y aspartoquinasa, formando en las dos primeras (y probablemente también en acetilglutamato sintasa) una horquilla β. En el caso de la carbamato quinasa hay una inserción en el centro de la horquilla que se proyecta hacia afuera constituyendo un subdominio de cuatro elementos (tres elementos β y una hélice α), denominado subdominio lejano, que actúa como tapadera sobre el centro activo y que posiblemente se mueva y aporte en carbamato quinasa una lisina catalítica.

5) El lazo β4-αE es extenso en todos los enzimas de la familia, formando en carbamato quinasa, acetilglutamato quinasa (y probablemente acetilglutamato sintasa) y aspartoquinasa largas horquillas β, que interaccionan con el lazo entre β3 y αD formando con la horquilla β de este último lazo de carbamato quinasa y acetilglutamato quinasa una hoja β extra de 4 elementos.

6) Como ya se indicó más arriba, es variable el grado de desarrollo y la posición del lazo que une las hélices αF y αG, siendo dicho desarrollo máximo en el caso de la UMP quinasa de *Pyrococcus furiosus*.

6) La posición y/o el tamaño de la hélice H son también variables, hasta el extremo de su desaparición en aspartoquinasa.

7) Como ya se indicó en la introducción y se esquematizó en la figura 7 de la misma, algunos enzimas de esta familia, incluyendo glutamato 5-quinasa, aspartoquinasa y ciertas formas de acetilglutamato quinasa tienen el extremo C-terminal del dominio aminoácido quinasa conectado a otro dominio adicional.

Muchos de los lazos en los que se dan diferencias importantes entre enzimas de la familia pertenecen al dominio N-terminal, y están relacionadas con el sitio de unión del sustrato fosforilable, lo que resulta comprensible porque éste es el sustrato distinto para cada enzima. Con frecuencia estos lazos variables están implicados en la formación directa del sitio para el sustrato, o proveen elementos estructurales que soportan dicho sitio, o que lo cubren.

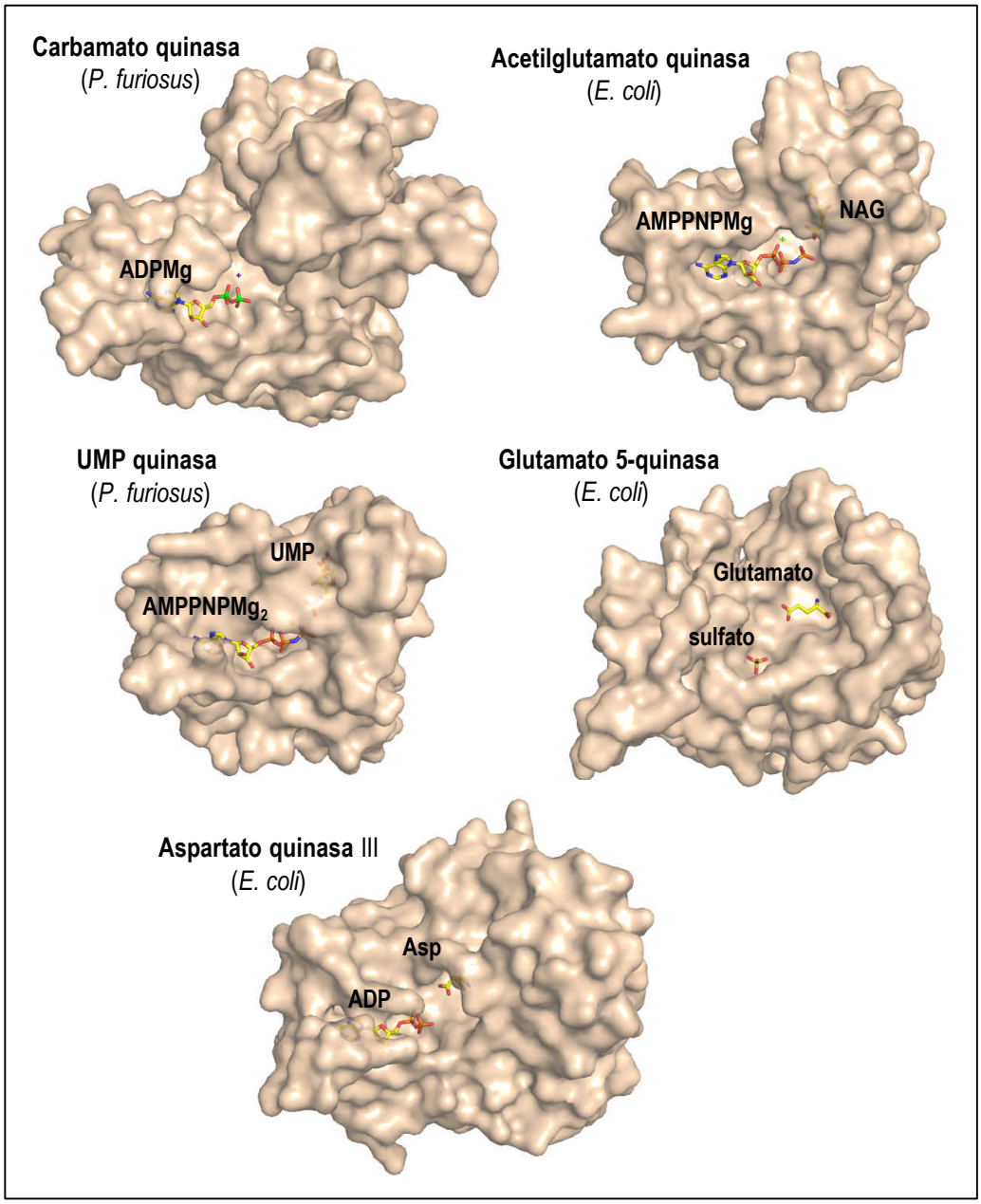


Figura 4- Representación en superficie de la subunidad de la carbamato quinasa de *Pyrococcus furiosus* en presencia de MgADP (PDB 1E19), la acetilglutamato quinasa (NAGK) de *E. coli* en presencia de AMPPNPMg y N-acetil-L-glutamato (NAG) (PDB 1GS5), la UMP quinasa de *Pyrococcus furiosus* con $AMPPNPMg_2$ y UMP (PDB 2BMU) y los dominios aminoácido quinasa de la glutamato 5-quinasa de *E. coli* en presencia de glutamato y sulfato y la aspartoquinasa de *E. coli* con aspartato y ADP (PDB 2JOW).

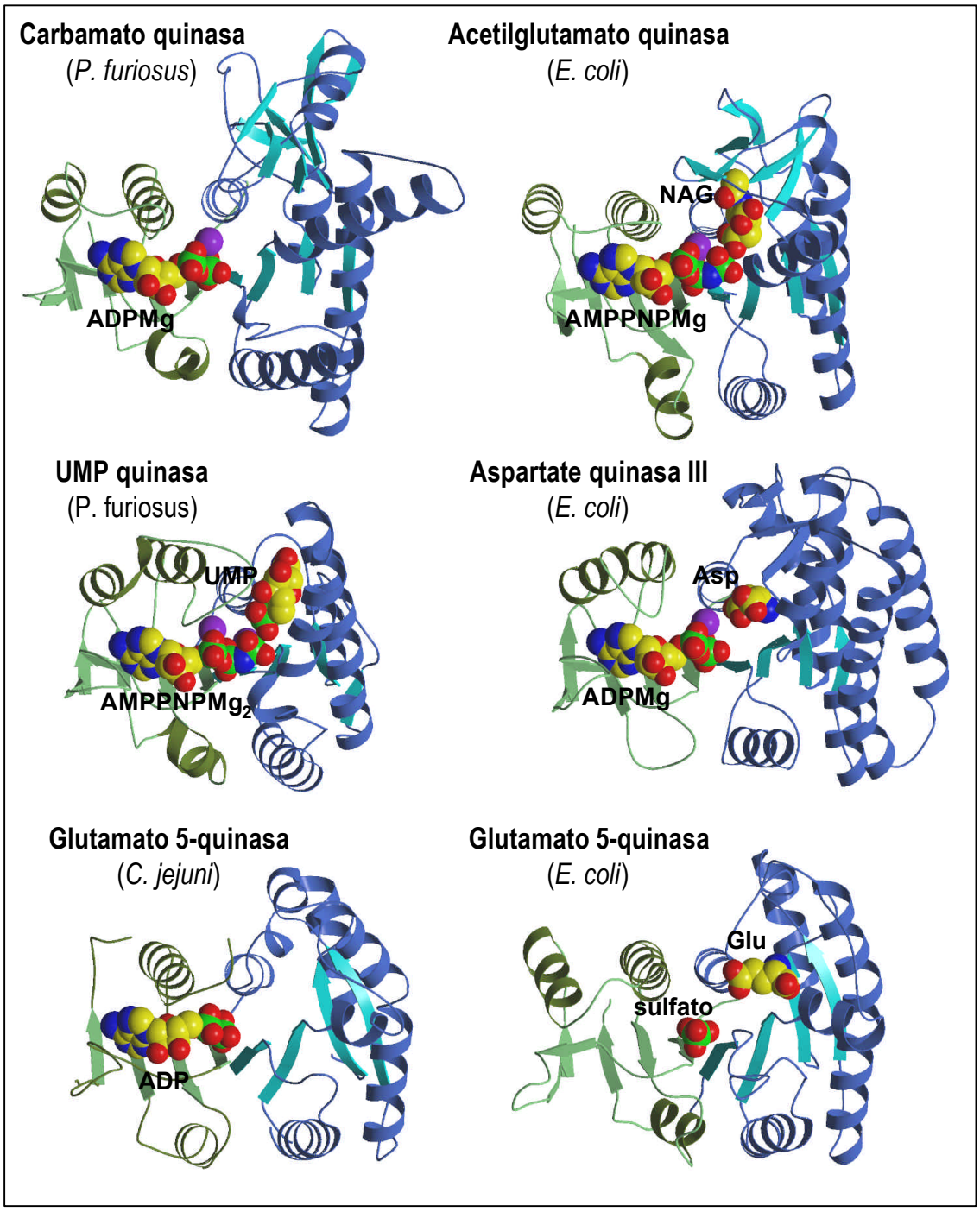


Figura 5- Representación en modelo de cintas de la subunidad de la carbamato quinasa de *Pyrococcus furiosus* en presencia de MgADP (PDB 1E19), la acetilglutamato quinasa (NAGK) de *E. coli* en presencia de AMPPNPMg y N-acetil-L_glutamato (NAG) (PDB 1GS5), la UMP quinasa de *Pyrococcus furiosus* con $AMPPNPMg_2$ y UMP (PDB 2BMU), el dominio aminoácido quinasa de la aspartoquinasa III de *E. coli* con aspartato y ADP (PDB 2JOW).la glutamato 5-quinasa de *C .jejuni* en presencia de ADP (PDB 2AKO) y el dominio aminoácido quinasa de la glutamato 5-quinasa de *E. coli* en presencia de glutamato y sulfato. Los elementos del dominio N-terminal o C-terminal en cada uno de los enzimas se colorean azul o verde respectivamente.

Constancia en el modo de unión de los sustratos y elementos comunes de la catálisis.

Si el plegamiento de la cadena o "fold" es esencialmente constante en todos los enzimas de la familia, también parece serlo el modo de unión de los sustratos (Figura 5). Aún existiendo diferencias significativas en los detalles, y aunque no se ha determinado el complejo con ATP o con un análogo inerte del ATP en todos los casos, sí se han determinado complejos al menos con ADP para todos los enzimas para los que existe estructura cristalina. Los datos estructurales indican que el nucleótido se une en un surco formado entre el borde C-terminal de la hoja β central y la capa de cuatro hélices, con la adenina hacia la porción C-terminal, y con el polifosfato hacia el dominio N-terminal, estando, en los casos en que se consiguió cristalizar el complejo bisustrato (acetilglutamato quinasa y UMP quinasa), el γ-P muy próximo al oxígeno atacante. No se han cristalizado complejos de carbamato quinasa con carbamato o con carbamil fosfato, pero en todos los demás casos en que disponemos de estructura cristalina se han cristalizado complejos con el sustrato fosforilable, y se ha podido comprobar que dicho sustrato se une siempre en el dominio N-terminal, en un sitio encuadrado por los lazos de dicho dominio. La ubicación del sitio es tan monótona que, cuando sólo se disponía de la estructura de la acetilglutamato quinasa, predijimos correctamente los residuos que en aspartoquinasa III de *E. coli* unen los grupos α-amino y α-carboxilato del aspartato, a pesar de que tan sólo existe 18 % de identidad de secuencia entre acetilglutamato quinasa y aspartoquinasa III.

La similitud entre los sitios para los sustratos se extiende también a algunos componentes de la maquinaria catalítica, que incluye en todos los casos una lisina que emerge generalmente de β1, aunque en carbamato quinasa está reemplazada por una lisina mucho más distal en la secuencia y en UMP quinasa arqueal por un Mg^{2+} extra, y que incluye también a un aspartato que establece interacciones con la cadena lateral de dicha lisina (o Mg^{2+}) catalítica, teniendo dicho aspartato un papel clave como organizador del centro activo. Otros elementos generalmente involucrados en la unión del fosfato que se transfiere son las conexiones β1-αA y β2-αB, que incluyen residuos pequeños, generalmente con predominio de glicinas, y que proveen algún puente de hidrógeno donador al fosforilo que se transfiere. Es constante la ubicación del carboxilato o fosfato atacante entre la conexión β2-αB y el lazo β4-αE, formando algún puente de hidrógeno con este último lazo. Un elemento catalítico adicional identificado en la acetilglutamato quinasa, pero que los estudios con UMP quinasa y con aspartoquinasa III favorecen también, es la existencia en los enzimas de esta familia de

estress compresivo, la inducción de una conformación del enzima en el complejo bisustrato en el que las distancias entre los grupos reaccionantes son anormalmente cortas, sugiriendo que el enzima estabiliza de este modo, uniendo con máxima afinidad, el estado de transición o intermedio pentavalente del fosfato (Ramón-Maiques et al., 2002). Los datos con UMP quinasa y aspartoquinasa apoyan de forma clara un mecanismo asociativo de transferencia del fosfato en los enzimas de esta familia.

Claramente, en la evolución desde la quinasa ancestral predecesora última de los enzimas de esta familia, se ha mantenido esencialmente incambiado durante todo el proceso evolutivo el sitio para ATP y la ubicación del sitio para el sustrato a fosforilar así como los mecanismos de fijación de los grupos reaccionantes en posición para que tenga lugar la reacción y algunos elementos catalíticos clave, introduciendo sólo pequeñas variaciones que han permitido conferir especificidad diferencial para el sustrato fosforilable, así como algunos cambios en los elementos catalíticos, cuya motivación no entendemos, aunque podrían resultar en diferencias en la velocidad de catálisis entre los diferentes enzimas. Un caso particular lo plantea la acetilglutamato sintasa. Las similitudes entre este enzima y la acetilglutamato quinasa indican que ambos enzimas derivan de una duplicación reciente. La sintasa ha incorporado un dominio extra transacetilasa, y nuestro grupo ha propuesto que la unión de los sustratos y la catálisis en la acetilglutamato sintasa requiere de la colaboración de ambos dominios del componente aminoácido quinasa con el dominio transacetilasa de la subunidad contigua.

Diferencias con otros enzimas que catalizan reacciones análogas.

La comparación con los enzimas de la familia E.C. 2.7.2. manifiesta a las claras la existencia en esta familia de tres grupos estructurales bien diferenciados, entre los que, por ahora, no parece fácil establecer relaciones evolutivas, dadas las muy importantes diferencias en el plegamiento de la cadena polipeptídica en unos y otros. Frente al plegamiento característico aminoácido quinasa descrito aquí, acetato quinasa, butirato quinasa, propionato quinasa y muy probablemente formiato quinasa (dada su homología franca con los enzimas anteriores) presentan el característico plegamiento de la superfamilia ASKHA (acetato y azucar quinasas/Hsc70/actina), plegamiento caracterizado por la presencia de dos dominios similares que posiblemente hayan derivado de una duplicación de un dominio original. Cada uno de estos dominios presenta un plegamiento aß, formado por una hoja ß (de 8 elementos en el dominio amino terminal y de 7 en el carboxi terminal) a uno de cuyos lados se dispone una capa

de a hélices. Los sustratos se unen en la hendidura entre los dos dominios, sobre la superficie del dominio carboxiterminal (Figura 6).

El otro plegamiento del grupo *E.C.* 2.7.2, el de la fosfoglicerato quinasa, no se asemeja al del dominio aminoácido quinasa ni tampoco al de la acetato quinasa pero presenta igual que este último, dos dominios claramente diferenciados que unen cada uno de los sustratos (Figura 6). Cada dominio está formado por una hoja ß de 5 elementos paralelos que forma un sandwich entre dos capas de dos a hélices cada una. El dominio N-terminal, que une al fosfoglicerato, presenta además una horquilla ß, mientras que el dominio C-terminal, responsable de la unión del nucleótido, presenta una hoja ß de 3 elementos antiparalelos y 4 hélices adicionales. Se conoce la estructura de distintos complejos de este enzima y se ha visto que durante el ciclo catalítico se producen movimientos importantes de los dominios, que permiten el acercamiento de ambos sustratos y permiten la transferencia del grupo fosforilo entre ambos.

La comparación de los centros activos en los distintos enzimas muestra que fosfoglicerato quinasa y NAGK presentan elementos comunes como la presencia de dos lazos ricos en glicinas y varios residuos con carga positiva en el sitio de transferencia del grupo fosforilo. Tambien en ambos enzimas hay dos hélices que orientan el extremo N-terminal hacia el sitio de transferencia del grupo fosforilo, pudiendo asistir en la catalisis. El centro activo de acetato quinasa presenta mayores diferencias y el el destaca la presencia de varias arginas y una histidina que participan en la catalisis.

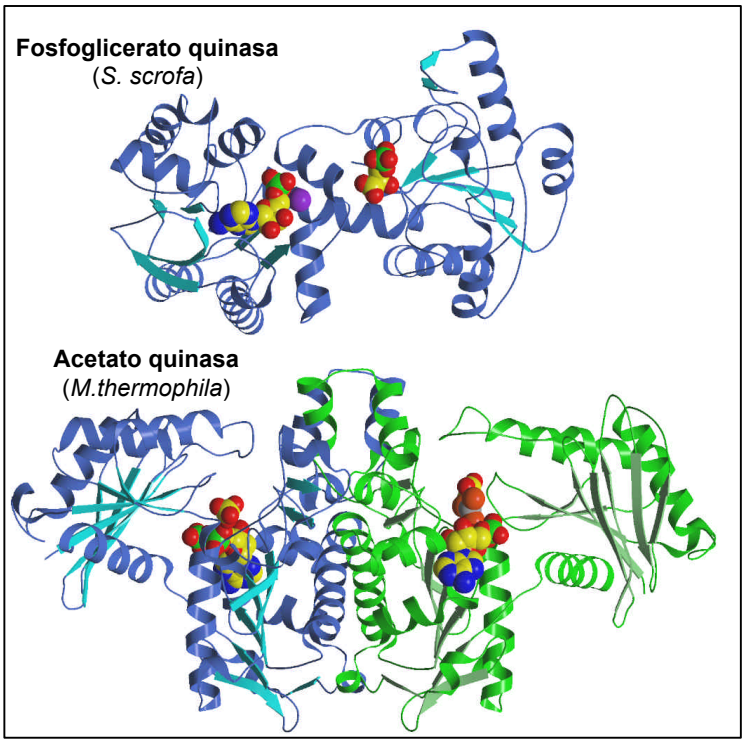


Figura 6- Representación en modelo de cintas de la fosfoglicerato quinasa de *S. scrofa* en presencia de 3-fosfoglicerato y MgAMP (PDB 1HDI) y del dímero de la acetato quinasa de *Methanosarcina thermophila* en presencia de ADP, AlF_3 y acetato (PDB 1TUY). Los sustratos se representan en modelo de bolas y cada una de las subunidades de la acetato quinasa en un color diferente.

Regulación feed-back: una cuestión sólo parcialmente resuelta.

Otra propiedad clave de la familia aminoácido quinasa es su susceptibilidad a la inhibición feed-back, que hace que cinco de los seis enzimas de esta familia sean reguladores cruciales de sus rutas metabólicas, incluyendo en el caso de dos de ellos, acetilglutamato quinasa y UMPK, no sólo retroinhibición, sino también activación, por la proteína P_{II} y por GTP, respectivamente. Los hallazgos estructurales han dado ya una respuesta razonable para la regulación en los casos de acetilglutamato quinasa y aspartoquinasa. La inhibición de la acetilglutamato quinasa por arginina, aclarada por otros miembros de nuestro grupo que han cristalizado formas sin arginina y con arginina, parece deberse a la expansión del anillo del hexámero enzimático causada por la unión de la arginina flanqueando las uniones interdiméricas, expansión que distancia los dos dominios de cada subunidad, previniendo el estrés compresivo que es un importante elemento en la catálisis (Ramón-Maiques et al., 2006). La unión del acetilglutamato favorece el acercamiento de ambos dominios y explica el carácter competitivo de la inhibición por la arginina respecto al acetilglutamato. Este mismo mecanismo mediado a través de elementos diferentes explica las acciones activadoras de la proteína P_{II} sobre la acetilglutamato quinasa de organismos fotosintéticos (resultados aún no publicados de nuestro grupo). Dada la probable similitud estructural entre acetilglutamato quinasa y acetilglutamato sintasa, y la predicción en este último enzima de sitios para arginina equivalentes a los de acetilglutamato quinasa, es esperable que en la sintasa la inhibición por arginina obedezca a mecanismos similares a los propuestos para la quinasa.

El mecanismo de inhibición de la aspartoquinasa, aclarado muy recientemente al determinar la estructura de las formas R y T del enzima III de *E. coli* (Kotaka et al., 2006), parece depender de la unión del efector en un dominio extra regulador externo al dominio aminoácido quinasa, el dominio ACT, y de cambios en el sitio para el ATP puestos en marcha por dicha unión del efector, cambios que aparentemente se ven estabilizados por la formación de tetrámeros cuando el efector está presente. Excepto en el punto de que la organización oligomérica desempeña un papel importante en el proceso de inhibición, no hay aspecto alguno común entre las inhibiciones de acetilglutamato quinasa por arginina y de aspartoquinasa por lisina: los sitios para el efector están en lugares diferentes y los mecanimos de inhibición parecen totalmente distintos.

A la misma conclusión de diversidad de mecanismos efectores, se llega cuando examinamos cuanto sabemos sobre bases estructurales de la regulación en los otros dos

enzimas de la familia estudiados aquí, UMP quinasa y glutamato 5-quinasa. En ambos enzimas la regulación se ejerce a través del dominio aminoácido quinasa, y no a través de un dominio especializado como en aspartoquinasa. Desconocemos cómo tiene lugar la activación por GTP en UMP quinasa, pero la inhibición de este enzima por UTP parece deberse a la unión del efector en el sitio del UMP, solapando con el sitio para el polifosfato del ATP, a juzgar por la estructura cristalina del complejo con UTP publicada para el enzima de *E. coli* (Briozzo et al., 2005), y también obtenida por nosotros (Figura 7; resultados no publicados) con el enzima de *P. furiosus*. Por tanto, en el caso de la UMPK, la inhibición por UTP no puede considerarse alostérica, al tratarse de una pura competición con los sustratos. La inhibición por prolina o por ornitina de la glutamato 5-quinasa se encuentra envuelta aún en mayor oscuridad, ya que existen datos cinéticos y datos derivados de estudios de mutagénesis dirigida del enzima de *E. coli* que apoyan la unión de la prolina y del glutamato a sitios solapantes (Pérez-Arellano et al., 2006), y nuestros resultados estructurales revelan la presencia de oxoprolina, un análogo de la prolina, en un sitio en la vecindad del glutamato. Será esencial establecer la estructura del complejo con prolina para entender la inhibición de la G5K por este efector, pero entre tanto todo apunta a que este efector se une en un sitio próximo al sitio para el glutamato y ejerce sus acciones dificultando la unión del glutamato.

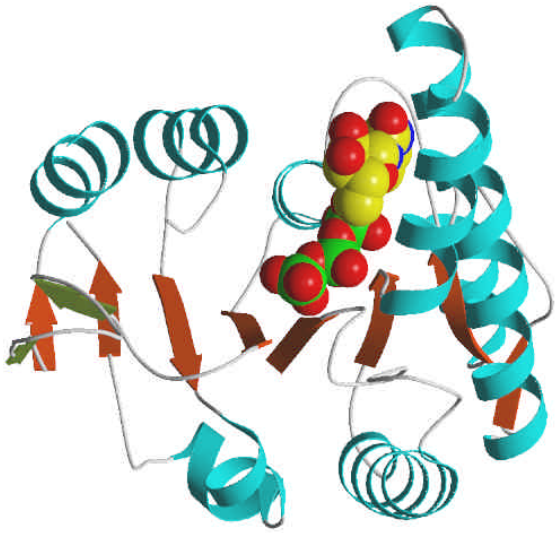


Figura 7- Representación en modelo de cintas de la UMP quinasa de *P. furiosus* en presencia de UTP. Las a hélices se colorean en azul y las hebras ß en naranja si forman parte de la hoja ß central o en verde si forman la horquilla ß. El UTP se representa en modelo de bolas.

Por tanto, los mecanimos de retroinhibición son diversos en los diversos enzimas de la familia, a pesar de que el plegamiento es monótono y la unión de los sustratos y la catálisis también lo son: los resultados indican que el control feed-back es una incorporación tardía, aparentemente desarrollada varias veces mediante mecanismos diversos y agregada a la maquinaria catalítica común. Sólo en el caso de acetilglutamato sintasa y acetilglutamato quinasa parece existir una diversificación de ambos enzimas posterior a la introduccción del mecanismo de control alostérico por arginina, aunque esto deberá ser corroborado mediante la determinación de la estructura de la acetilglutamato sintasa y la demostración de las bases estructurales de su mecanismo de inhibición. Es cierto que en el caso de la aspartoquinasa la regulación se ejerce a través de un dominio especializado en esta función, el dominio ACT. Sin embargo, es sorprendente que las dos recientísimas publicaciones en que se describe la estructura de dos aspartoquinasas resaltan la existencia de diferencias notables entre los dominios ACT de estos enzimas y los de otros enzimas que también lo poseen, incluyendo entre las diferencias el modo de unión de los inhibidores. Por tanto, incluso en el caso de este dominio especializado en regulación no parece existir un patrón constante de ubicación de lugar para efector similar a la ubicación constante de los lugares para sustratos en el dominio aminoácido quinasa, ni tampoco es evidente que los mecanismos por los que este dominio media la regulación estén conservados entre los distintos enzimas que portan tal dominio.

La aspartoquinasa III proporciona la oportunidad de colegir el valor de las técnicas de modelizado.

Nuestro modelizado del dominio aminoácido quinasa de la aspartoquinasa III ha utilizado todos los recursos posibles para alcanzar la máxima calidad: se ha guíado por la mutagénesis de residuos estimados clave y por la predicción de estructura secundaria a partir de varios programas diferentes, utilizando una estructura establecida, la de la acetilglutamato quinasa de *E. coli*, para estimar la calidad de la predicción. El modelizado se ha afinado mediante el uso de tests estrictos de calidad estereoquímica, liberando la estructura de restricciones basadas en la identidad de secuencia en las últimas fases de afinamiento, con la finalidad de alcanzar una estructura estereoquímicamente óptima. El hecho de que muy recientemente se haya publicado la estructura experimental determinada mediante cristalografía de rayos X de la aspartoquinasa III de *E. coli* a 2.8 Å de resolución (Kotaka et al., 2006), nos ofrece la oportunidad de cotejar nuestros resultados del modelizado con los resultados

experimentales de determinación estructural. Tal comparación tiene valor, pues dado el elevado nivel de exigencia de nuestra técnica de modelizado, nuestros resultados pueden considerarse razonables para secuencias con grados de indentidad del orden del 18% (la identidad entre aspartoquinasa III y acetilglutamato quinasa, ambas de *E. coli*). Naturalmente, no fue posible predecir los dos elementos α interpuestos entre αB y αC, para los que no hay correspondencia en acetilglutamatoquinasa. La Figura 8 superpone el trazado de la cadena de C_α para los elementos predichos del modelo y para sus correlatos experimentales en la estructura cristalina. Además, se representa la superposición de cada uno de los dos subdominios del dominio aminoácido quinasa. Como puede verse la superposición de los elementos de estructura secundaria del dominio N-terminal es correcta en su mayor parte, siendo la peor predicción la de los elementos αD y β4. También se aprecian diferencias importantes en los lazos predichos y en las orientaciones relativas de αB y αC. Como consecuencia, la superposición de los C_α de este dominio da un valor de rmsd de 1.915 Å para 124 C_α equivalentes, o, si sólo se utilizan los C_α de las α hélices y de los elementos β de la hoja principal del dominio, un valor algo mejor, de 1.823 Å para 98 C_a de primer subdominio.

La predicción es mucho peor para el subdominio C-terminal. La imposibilidad de predecir correctamente el largo lazo que une β5 a β6 ha resultado en predicciones completamente erróneas de β6 y αF, y parcialmente errónea de αG. Sin embargo, predijimos correctamente la ausencia de αH. Si superponemos todos los elementos de este dominio el valor de rmsd es de 2.139 para 41 átomos C_α, mientras que si se superponen los elementos bien predichos, el valor de rsmd es de 1.244 para y 19 C_α.

La superposición de los dominios AAK completos, experimental y predicho (sin los elementos que no intentamos predecir), se observa un acuerdo bastante bueno de la hoja β central y mucho menos bueno de las α hélices flanqueantes. El resumen de todas estas observaciones es que, al menos para identidades de secuencia bajas, como las que se dan entre aspartoquinasa III y acetilglutamato quinasa, la predicción está todavía lejos de proporcionar resultados consistentemente fidedignos, aunque sea capaz ya de predecir una fracción importante de la estructura y de predecir con éxito el papel de determinados residuos del centro activo en la catálisis y en la unión de los sustratos.

En cuanto a la unión de sustratos y la catálisis, nuestras prediccciones basadas en el modelizado han sido razonablemente buenas. El aspartato se une a la AKIII en la cavidad que nosotros habiamos propuesto por homologia con acetilglutamato quinasa, aunque la localización del sustrato es algo diferente. La cadena lateral de Phe184, que en nuestro modelo quedaba lejos del sitio de unión del aspartato, participa en el sitio de

unión de este sustrato, con el que establece interacciones de tipo hidrofóbico. El residuo E119 está implicado en la unión de grupo a-amino del aspartato y el R198, a pesar de ocupar una posición en la estructura distante a la que habiamos predicho, une el grupo a-carboxilato del glutamato de acuerdo con nuestra propuesta. Además, los papeles de K8 y D202 en la catálisis y la organización del centro activo se corroboran en la estructura experimental. Sin embargo, aunque el nucleotido se une en el mismo sitio que en el resto de enzimas de la familia, el modelizado de los residuos que conforman este sitio es incorrecto como consecuencia de nuestra incapacidad para predecir correctamente la horquilla ß5'-ß6', β6, aF, aF-aG y en menor grado aG. Como consecuencia la adenina no reposa sobre la Met251, sino sobre Val258, y no está cubierta exteriormente por la Tyr227, sino, como en UMPK, por los residuos del lazo que incluye la horquilla β5'-β6'.

Una forma de plegamiento básico, varias arquitecturas oligoméricas.

El trabajo realizado en esta tesis, el trabajo realizado en paralelo por otros miembros del laboratorio y los trabajos recientes sobre aspartoquinasa muestran que muchos de los enzimas de la familia aminoácido quinasa forman homo-oligómeros de orden superior al dímero. La glutamato 5-quinasa de *Escherichia coli* es tetramérica (Pérez-Arellano et al., 2005), y la aspartoquinasa III de este microorganismo forma tetrámeros en presencia de lisina. Las NAGKs inhibibles por arginina y las UMPKs son hexaméricas, aunque la organización de ambos tipos de hexámeros es completamente diferente (Ramón-Maiques et al., 2006). Cabe plantearse cuáles son las claves de la formación de esas distintas formas oligoméricas a partir de una misma estructura básica del dominio predominante y cuál es el valor funcional de los oligómeros. A la primera cuestión se ha respondido en realidad ya en nuestros trabajos sobre UMP quinasa y G5K así como en nuestra discusión sobre aspartoquinasa. No hay duda que en todos los casos se utiliza esencialmente la misma superficie de dimerización del dominio aminoácido quinasa observada en carbamato quinasa, para establecer interacciones mutuas entre dominios aminoácido quinasa en los diferentes enzimas de la familia, excepto en aspartoquinasa. Sin embargo, como ya se ha comentado, la diferente orientación mutua de los elementos de interacción, particularmente entre las hélices C de las subunidades que interaccionan, determina el tipo de dímero básico formado. En el caso de la aspartoquinasa se ha desarrollado una horquilla de hélices que mimetizan a la otra subunidad y previenen la interacción entre dominios aminoácido quinasa, en el dímero básico.

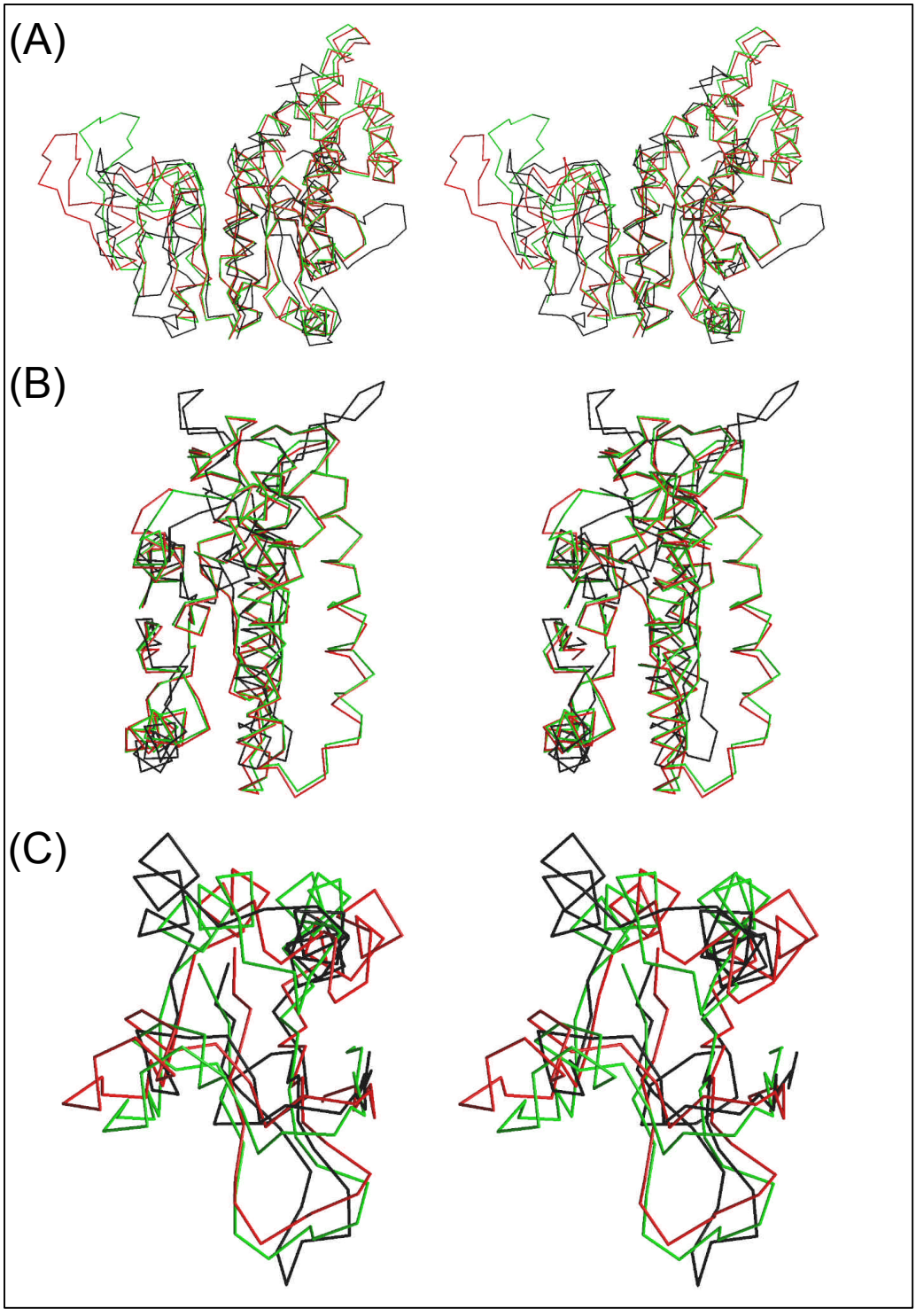


Figura 8- Comparación del modelizado y la estructura experimental de la aspartoqinasa III de *E. coli*. (A)Vista estereocópica de la superposición del dominio aminoácido quinasa de la aspartoquinasa III de *E. coli* modelizada por nosotros (PDB 2OHI) (en negro) con la estructura experimental de este enzima determinada en presencia de aspartato y ADP (PDB 2JOW) (en verde) o aspartato y lisina (PDB 2JOX) (en rojo). (B) y (C) vista correspondiente a la rotación de 90 º en el eje horizontal y paralelo al papel, respecto a A) en la que se representa bien la superposición del dominio N-terminal (B) o del C-terminal (C).

El distinto tipo de dímero formado condiciona la formación de hexámeros diferentes en acetilglutamato quinasa y UMPK, ya que estos hexámeros son el resultado de la interacción entre caras distintas del sandwich αβα. Aspartoquinasa y glutamato 5-quinasa representan casos especiales, ya que la existencia de un dominio C-terminal extra condiciona o incluso determina en gran parte las caracterísitcas del dímero básico formado. El mensaje básico derivado de estas observaciones es evidente: la similitud en el plegamiento básico no necesariamente se corresponde con similitud en el tipo de oligómero formado.

En cuanto a la cuestión del valor de formar oligómeros superiores, las evidencias obtenidas de los diferentes enzimas sugieren que el oligómero formado puede ser un determinante esencial en la incorporación de la inhibición feed-back, como parece ser el caso de la acetilglutamato quinasa, y, en menor grado, en el caso de la aspartoquinasa. Además, como sucede en la UMP quinasa, puede ser una condición clave para la incorporación de nuevas funciones, tales como la regulación de la expresión génica mediada por las interacciones con PepA. Quizá las necesidades de la canalización condicionen también el carácter oligomérico de glutamato 5-quinasa. Sin embargo, el valor del oligómero en la funcionalidad del enzima debe someterse a prueba experimental mediante modificaciones en la proteína enzimática que alteren la arquitectura oligomérica. Están en marcha en nuestro grupo experimentos en esta dirección con los diferentes enzimas de la familia.

CONCLUSIONES

CONCLUSIONES

1- Hemos demostrado usando mutagénesis dirigida, que la lisina 8 y el aspartato 162 de la acetilglutamato quinasa de *E. coli*, desempeñan papeles clave en la catálisis y que la arginina 66 y la asparagina 158 están implicadas directamente en la unión del acetilglutamato. Con ello probamos experimentalmente predicciones basadas en la estructura de este enzima.

2- La mutagénesis dirigida ha demostrado que la lisina 8 y el aspartato 202 de la aspartoquinasa III de *E. coli* tienen papeles catalíticos análogos a los de la lisina 8 y el aspartato 162 de la acetilglutamato quinasa. Esta misma aproximación ha demostrado la implicación directa del glutamato 119 y de la arginina 198 de la aspartoquinasa III de *E. coli* en la unión del aspartato.

3- Hemos generado mediante modelizado a partir de la acetilglutamato quinasa de *E. coli* la estructura tridimensional del dominio aminoácido quinasa de la aspartoquinasa III de *E. coli.* La comparación de este modelo con la estrcutura determinada para este enzima recientemente por otros es globalmente excelente (1.9 Å para 180 Ca's).

4- Hemos determinado la estructura tridimensional a nivel de resolución de 2.4-3 Å, de la UMP quinasa de *Pyrococcus furiosus* sóla o en complejo con el análogo inerte del ATP, AMPPNP o con AMPPNP y UMP. El enzima es un hexámero esférico organizado como un trímero de dimeros hueco en su núcleo. Hemos aclarado su arquitectura, la unión de los sustratos y la catálisis, y hemos predicho el modo de unión a la proteina PepA, con la que regula la expresión de los genes de la carbamil fosfato sintetasa.

5- Hemos determinado la estructura de la glutamato 5-quinasa de *Escherichia coli* a nivel de resolución de 2.5-2.9 Å en presencia de glutamato y sulfato o en presencia de glutamilfosfato, sulfato y oxoprolina. El enzima es un tetrámero planar y elongado organizado como un dímero de dímeros. La estructura aclara el modo de unión de los sustratos y la catálisis y ayuda a proponer modelos de interacción con el enzima glutamilfosfato reductasa que hagan posible la canalización del glutamilfosfato entre ambos enzimas. Los resultados demuestran el modo de integración de un dominio extra de la familia PUA en la arquitectura del tetrámero.

BIBLIOGRAFIA

CORRESPONDIENTE A LA INTRODUCCIÓN GENERAL, RESUMEN DE LOS RESULTADOS Y DISCUSIÓN GENERAL.

BIBLIOGRAFÍA CORRESPONDIENTE A LA INTRODUCCIÓN GENERAL, RESUMEN DE LOS RESULTADOS Y DISCUSIÓN GENERAL.

Abadjieva, A., Pauwels, K., Hilven, P. y Crabeel, M. (2001). A new yeast metabolon involving at least the two first enzymes of arginine biosynthesis: acetylglutamate synthase activity requires complex formation with acetylglutamate kinase. *J. Biol. Chem.* **276,** 42869-42880.

Aral, B. y Kamoun, P. (1997).The proline biosynthesis in living organisms, *Amino Acids* **13**, 189-217.

Aravind, L. y Koonin E. V. (1999).Novel predicted RNA-binding domains associated with the translation machinery. *J. Mol. Evol.* **48**, 291-302.

Arcondeguy, T., Jack, R. y Merrick, M. (2001). P_{II} signal transduction proteins, pivotal players in microbial nitrogen control. *Microbiol. Mol. Biol. Rev.* **65,** 80-105.

Baumgartner, M. R., Hu, C. A., Almashanu, S., Steel, G., Obie, C., Aral, B., Rabier, D., Kamoun, P., Saudubray J. M. y Valle, D. (2000). Hyperammonemia with reduced ornithine, citrulline, arginine and proline: a new inborn error caused by a mutation in the gene encoding Δ^1-pyrroline-5-carboxylate synthase. *Hum. Mol. Genet.* **9,** 2853-2858

Briozzo, P., Evrin, C., Meyer, P., Assairi, L., Joly, N., Barzu, O. y Gilles, A. M. (2005). Structure of *Escherichia coli* UMP kinase differs from that of other nucleoside monophosphate kinases and sheds new light on enzyme regulation. *J. Biol. Chem*. **280**, 25533-25540.

Burillo, S., Luque, I., Fuentes, I. y Contreras, A. (2004). Interactions between the nitrogen signal transduction protein PII and N-acetyl glutamate kinase in organisms that perform oxygenic photosynthesis. *J. Bacteriol.* **186,** 3346-3354.

Buss, K. A., Cooper, D. R., Ingram-Smith, C., Ferry, J. G., Sanders, D. A. y Hasson, M. S. (2001). Urkinase: structure of acetate kinase, a member of the ASKHA superfamily of phosphotransferases. *J. Bacteriol*. **183**, 680-686.

Cassan, M., Parsot, C., Cohen, G. N. & Patte, J. C. (1986). Nucleotide sequence of *lys*C gene encoding the lysine-sensitive aspartokinase III of *Escherichia coli* K12. Evolutionary pathway leading to three isofunctional enzymes. *J. Biol. Chem.* **261**, 1052-1057.

Chipman, D. M. y Shaanan, B. (2001). The ACT domain family. *Curr. Opin. Struct. Biol.* **11**, 694-700.

Cunin, R., Glansdorff, N., Piérard, A. y Stalon, V. (1986). Biosynthesis and metabolism of arginine in bacteria. *Microbiol. Rev.* **50,** 314-352.

Fernandez-Murga M. L., Gil-Ortiz F., Llacer J. L. y Rubio V. (2004) Arginine biosynthesis in *Thermotoga maritima*: characterization of the arginine-sensitive N-acetyl-L-glutamate kinase. *J Bacteriol*. **186**, 6142-6149.

Fujita, T., Maggio, A., Garcia-Rios, M., Stauffacher, C., Bressan R. A. y Csonka L. N. (2003). Identification of regions of the tomato γ-glutamyl kinase that are involved in allosteric regulation by proline. *J. Biol. Chem.* **278**, 14203-14210.

Gagyi, C., Bucurenci, N., Sîrbu, O., Labessse, G., Ionescu, M., Ofiteru, A., Liliane, A., Stephanie, L., Danchin, A., Bârzu, O., y Gilles, A. M. (2003). UMP kinase from the Gram-positive bacterium *Bacillus subtilis* is strongly dependent on GTP for optimal activity. *Eur. J. Biochem.* **270**, 3196-3204.

Gil-Ortiz, F., Ramón-Maiques, S., Fita, I. y Rubio, V. (2003). The course of phosphorus in the reaction of N-acetyl-L-glutamate kinase, determined from the structures of crystalline complexes, including a complex with an AlF_4^- transition state mimic. *J. Mol. Biol.* **331,** 231-244.

Haas, D. y Leisinger, T. (1975). N-acetylglutamate 5-phosphotransferase of *Pseudomonas aeruginosa*. Catalytic and regulatory properties. *Eur. J. Biochem.* **52,** 377-383.

Hall, D. A., Zhu, H., Zhu, X., Royce, T., Gerstein, M. y Snyder, M. (2004). Regulation of gene expression by a metabolic enzyme. *Science* **306,** 482-484.

Heinrich, A., Maheswaran, M., Ruppert, U. y Forchhammer, K. (2004). The *Synechococcus elongatus* P_{II} signal transduction protein controls arginine synthesis by complex formation with N-acetyl-L-glutamate kinase. *Mol. Microbiol.* **52,** 1303-1314.

Hu, C. A. Lin, W. W., Obie C. y Valle, D. (1999). Molecular enzymology of mammalian Δ^1-pyrroline-5-carboxylate synthase. Alternative splice donor utilization generates isoforms with different sensitivity to ornithine inhibition. *J. Biol. Chem.* **274,** 6754–6762

Ishitani, R., Nureki, O., Fukai, S., Kijimoto, T., Nameki, N., Watanabe M., Kondo H., Sekine, M., Okada, N., Nishimura S. y Yokoyama, S. (2002). Crystal structure of archaeosine tRNA-guanine transglycosylase. *J. Mol. Biol.* **318**, 665-677.

Kholti, A., Charlier, D., Gigot, D., Huysveld, N., Roovers, M. y Glansdorff, N. (1998). *pyrH*-encoded UMP-kinase directly participates in pyrimidine-specific modulation of promoter activity in *Escherichia coli*. *J. Mol. Biol*. **280**, 571-582.

Kobashi, N., Nishiyama, M. y Tanokura, M. (1999). Kinetic and mutation analyses of aspartate kinase from *Thermus flavus*. *J. Biosci. Bioeng.* **87**, 739-745.

Kotaka, M., Ren, J., Lockyer, M., Hawkins, A. R. y Stammers, D. K. (2006). Structures of R- and T-state *Escherichia coli* aspartokinase III: Mechanisms of the allosteric transition and inhibition by lysine. *J Biol Chem.* **281**, 31544-31552.

Leisinger, T., y Haas, D. (1975). N-Acetylglutamate synthase of *Escherichia coli* regulation of synthesis and activity by arginine. *J. Biol. Chem.* **250**, 1690-1693.

Leisinger, T. (1996). Biosynthesis of proline. In: F.C. Neidhardt, Editor, *Escherichia coli and Salmonella: Cellular and Molecular Biology*, ASM Press, Washington, DC, pp. 434–441.

Marina, A., Alzari, P. M., Bravo, J., Uriarte, M., Barcelona, B., Fita, I. y Rubio V. (1999). Carbamate kinase: New structural machinery for making carbamoyl phosphate, the common precursor of pyrimidines and arginine. *Protein Sci*. **8,** 934-940.

Mas-Droux, C., Curien, G., Robert-Genthon, M., Laurencin, M., Ferrer, J. L y Dumas, R. (2006). A Novel organization of ACT domains in allosteric enzymes revealed by the crystal structure of *Arabidopsis* aspartate kinase. *Plant Cell* **18**, 1681-1692.

Michal, G. (1999) *Biochemical Pathways: An Atlas on Biochemistry and Molecular Biology*, John Wiley and Sons Inc., NY

Neuhard, J. y Kelln, R. A. (1996). Biosynthesis and conversions of pyrimidines. In *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd. ed. (Neidhardt, F. C., ed.), vol I, pp. 580-599, ASM Press, Washington D.C.

Ogura, M., Kawata-Mukai, M., Itaya, M., Takio, K. y Tanaka, T. (1994) Multiple copies of the proB gene enhance degS-dependent extracellular protease production in *Bacillus subtilis*. *J. Bacteriol*. **176**, 5673-5680.

Ogura, M. y Tanaka, T. (1996). Transcription of *Bacillus subtilis degR* is sigma D dependent and suppressed by multicopy *proB* through sigma D. *J. Bacteriol.* **178**, 216-222.

Pan, H., Agarwalla, S., Moustakas, D. T, Finer-Moore J. y Stroud R. M. (2003). Structure of tRNA pseudouridine synthase TruB and its RNA complex: RNA recognition through a combination of rigid docking and induced fit, *Proc. Natl. Acad. Sci. USA* **100**,12648-12653.

Pauwels, K., Abadjieva, A., Hilven, P., Stankiewicz, A. y Crabeel, M. (2003). The N-acetylglutamate synthase/N-acetylglutamate kinase metabolon of *Saccharomyces cerevisiae* allows co-ordinated feedback regulation of the first two steps in arginine biosynthesis. *Eur. J. Biochem.* **270,** 1014-1024.

Pérez-Arellano, I., Gil-Ortiz, F. Cervera , J. y Rubio, V. (2004). Glutamate-5-kinase from *Escherichia coli*: gene cloning, overexpression, purification and crystallization of the recombinant enzyme and preliminary X-ray studies. *Acta Crystallogr. D* **60,** 2091-2094.

Perez-Arellano, I., Rubio, V. y Cervera, J. (2006). Mapping active site residues in glutamate-5-kinase. The substrate glutamate and the feed-back inhibitor proline bind at overlapping sites. *FEBS Lett*. **580**, 6247-6253.

Ramón-Maiques, S., Marina, A., Uriarte, M., Fita, I. y Rubio, V. (2000). The 1.5 Å resolution crystal structure of the carbamate kinase-like carbamoyl phosphate synthetase from the hyperthermophilic archaeon *Pyrococcus furiosus*, bound to ADP, confirms that this thermostable enzyme is a carbamate kinase, and provides insight into substrate binding and stability in carbamate kinases. *J. Mol. Biol.* **299,** 463-476.

Ramón-Maiques, S., Marina, A., Gil-Ortiz, F., Fita, I. y Rubio, V. (2002). Structure of acetylglutamate kinase, a key enzyme for arginine biosynthesis and a prototype for the amino acid kinase enzyme family, during catalysis. *Structure*, **10,** 329-342.

Ramón-Maiques, S., Fernandez-Murga, M. L., Gil-Ortiz, F., Vagin, A., Fita, I. y Rubio, V. (2006). Structural Bases of Feed-back Control of Arginine Biosynthesis, Revealed by the Structures of Two Hexameric N-Acetylglutamate Kinases, from *Thermotoga maritima* and *Pseudomonas aeruginosa*. *J. Mol. Biol*. **356**, 695-713.

Saygin, O. (1984). Photochemical carbamylphosphate formation and metal ion catalysed transphosphorilations between carbamylphosphate and adenine nucleotides or carboxyl groups. *Origins of life* **14**, 131-137.

Seddon, A. P., Zhao, K. Y. y Meister, A. (1989). Activation of glutamate by γ-glutamate kinase: formation of γ-cis-cycloglutamyl phosphate, an analog of γ-glutamyl phosphate. *J. Biol. Chem.* **264**, 11326-11335.

Serina, L., Blondin, C., Krin, E., Sismeiro, O., Danchin, A., Sakamoto, H. Gilles, A. M. y Barzu, O. (1995). *Escherichia coli* UMP-kinase, a member of the aspartokinase family, is a hexamer regulated by guanine nucleotides and UTP. *Biochemistry*, **18,** 5066-5074.

Smith, L. T. (1985). Characterization of a γ-glutamyl kinase from *Escherichia coli* that confers proline overproduction and osmotic tolerance. *J. Bacteriol.* **164** , 1088-1093.

Strizhov, N., Abraham, E., Okresz, L., Blickling, S., Zilberstein, A., Schell, J., Koncz, C. y Szabados, L. (1997). Differential expression of two P5CS genes controlling proline accumulation during salt-stress requires ABA and is regulated by ABA1, ABI1 andAXR2 in *Arabidopsis. Plant J.* **12** , 557-569.

Sugiyama, K., Hayakawa, T., Kudo, T., Ito, T. y Yamaya, T. (2004). Interaction of N-acetylglutamate kinase with a P_{II}-like protein in rice. *Plant Cell Physiol.* **45**, 1768-1778.

Thompson, J. D., Higgins, D. G. y Gibson, T. J. (1994). CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucl. Adics Res*. **22**, 4673-4680.

Truffa-Bachi, P. (1973). Microbial aspartokinases. In *The Enzymes* (Boyer, P., ed.), pp. 509-553, Academic Press, New York and London.

Uriarte, M., Marina, A., Ramon-Maiques, S., Fita, I. y Rubio V. (1999). The carbamoyl-phosphate synthetase of *Pyrococcus furiosus* is enzymologically and structurally a carbamate kinase. *J. Biol. Chem.* **274**, 16295-16303.

Watson, H. C., Walker, N. P., Shaw, P. J., Bryant, T. N., Wendell, P. L., Fothergill, L. A., Perkins, R. E., Conroy, S. C., Dobson, M. J., Tuite, M. F., et al. (1982). Sequence and structure of yeast phosphoglycerate kinase. *EMBO J.* **1**, 1635-1640.

Zrenner, R. Pyrimidine nucleotide kinases and primary plant metabolism. International Conference on Arginine and Pyrimidines. 17-20 August, 2006. Lund, Sweden.