UNIVERSITAT DE BARCELONA

Molecular studies of two methylerythritol 4-phosphate pathway enzymes of isoprenoid biosynthesis: the 4-diphosphocytidyl-2C-methyl-D-erythritol kinase and the 1-deoxy-D-xylulose 5-phosphate synthase

Estudios moleculares de dos enzimas de la ruta del metileritritol 4-fosfato de biosíntesis de isoprenoides: la 4-difosfocitidil-2C-metil-D-eritritol quinasa y la 1-dexosi-D-xilulosa 5-fosfato sintasa

Víctor Giménez Oya

Molecular studies of two methylerythritol 4-phosphate pathway enzymes of isoprenoid biosynthesis: the 4-diphosphocytidyl-2C-methyl-D-erythritol kinase and the 1-deoxy-D-xylulose 5-phosphate synthase

Estudios moleculares de dos enzimas de la ruta del metileritritol 4-fosfato de biosíntesis de isoprenoides: la 4-difosfocitidil-2C-metil-D-eritritol quinasa y la 1-desoxi-D-xilulosa 5-fosfato sintasa

Víctor Giménez Oya

Barcelona, 2009

Molecular studies of two methylerythritol 4-phosphate pathway enzymes of isoprenoid biosynthesis: the 4-diphosphocytidyl-2C-methyl-D-erythritol kinase and the 1-deoxy-D-xylulose 5-phosphate synthase

Estudios moleculares de dos enzimas de la ruta del metileritritol 4-fosfato de biosíntesis de isoprenoides: la 4-difosfocitidil-2C-metil-D-eritritol quinasa y la 1-desoxi-D-xilulosa 5-fosfato sintasa

Memoria presentada por **Víctor Giménez Oya**, licenciado en Biología por la Universidad de Girona, para optar al grado de doctor por la Universidad de Barcelona.

Los estudios de tercer ciclo necesarios para tal efecto se han enmarcado en el programa de doctorado de Biotecnología, bienio **2004-2006**, de la Universidad de Barcelona; y el proyecto de tesis doctoral se encuentra inscrito en el **Departamento de Bioquímica y Biología Molecular** de la facultad de Biología de la Universidad de Barcelona. El trabajo experimental y la redacción de la memoria que se presentan han sido dirigidos por el doctor **Santiago Imperial Ródenas**.

Barcelona, septiembre de 2009

El director de la tesis: El doctorando:

Dr. Santiago Imperial Ródenas **Víctor Giménez Oya**

A mis padres y a mi hermano,

A mi compañera por, entre otras muchas razones, querer asemejarme a lo que veo en sus ojos cuando me mira,

A todo aquel que se haya sentido implicado en este viaje,

Y, por supuesto, para los que no.

Avenida Rivadavia intersección con calle Ayacucho

Buenos Aires

Argentina

CONTENIDO

1. INTRODUCCIÓN GENERAL 3

1.1 Los isoprenoides 3
1.1.1La ruta del mevalonato o ruta clásica 6
1.1.2 La ruta del metileritritol 4-fosfato (MEP) o ruta alternativa 7
1.1.3 Distribución de las rutas del MVA y del MEP 8
1.1.4 Origen evolutivo de las rutas 9
1.1.5 La ruta del MEP representa una atractiva diana para la búsqueda de antibióticos, antimaláricos y herbicidas 10
1.2 Búsqueda de nuevos compuestos. Introducción al diseño racional de nuevos fármacos 11
1.2.1 Mecánica molecular 11
1.2.1.1 Parametrización de la energía: Campos de fuerza 11
1.2.1.2 Métodos de optimización 12
1.2.1.3 Dinámica molecular 12
1.2.1.3.1 Solvatación del sistema 13
1.2.1.3.2 Contraiones 13
1.2.1.3.3 Cálculo de propiedades 13
1.2.2 Búsqueda en bases de datos 3D 13
1.2.2.1 Hipótesis farmacofórica 14
1.2.2.2 Anclaje de los compuestos (o Docking) 14
1.2.2.3 Ordenación de los compuestos 15
1.2.2.4 Regla del 5 de Lipinski 15
1.2.3 Uso de péptidos 15
1.2.4 Técnicas empíricas de caracterización de la interacción ligando-receptor 16
1.2.4.1 Disminución de la intensidad por laser de deonización/ionización asistido por matriz-tiempo de vuelo (Intensity-fading MALDI-TOF MS) 16
1.2.4.2 Resonancia del plasmón en superficie 16
1.3 La 4-difosfocitidil-2C-metil-D-eritritol quinasa (CMK) 17
1.3.1 La superfamilia GHMP 18
1.4 La 1-desoxi-D-xilulosa 5-fosfato sintasa (DXS) 19
1.4.1 Información estructural de la DXS 20
1.5 Proteólisis 21
1.5.1 Estabilidad del enlace peptídico 21
1.5.2 Mecanismos de autoproteólisis intermolecular 22
1.5.3 Mecanismos de autoproteólisis intramolecular 22
1.5.3.1 Reorganización de los grupos acilo N-O o N-S 23
1.5.3.1.1 Enzimas piruvoil 24
1.5.3.1.2 Ntn hidrolasas, nucleoporinas y proteínas con el dominio SEA 25
1.5.3.1.3 Proteínas Hedgehog 27
1.5.3.1.4 Splicing proteico 28
1.5.3.1.4.1 Bacterial intein-like's (BIL) 29
1.5.3.2 Formación de succinimidas (imidas cíclicas) 30
1.5.3.2.1 La deamidación y la formación de isoaspartatos en proteínas 30

2. OBJETIVOS 35

INFORME FACTOR DE IMPACTO DE LOS ARTÍCULOS PUBLICADOS 37

3. RESUMEN GENERAL: RESULTADOS; DISCUSIÓN Y CONCLUSIONES

3.1 RESULTADOS 41
3.1.1 Imitando interacciones directas proteína-proteína e interacciones mediadas por solvente en el homodímero CDP-ME quinasa: Una aproximación de búsqueda virtual dirigida a través de farmacóforo. 42
3.1.2 Diseño de una molécula pequeña y un péptido cíclico natural como ligandos de la CDP-metileritritol quinasa imitando interacciones proteína-proteína e interacciones mediadas por solvente. 43
3.1.3 Análisis del estado oligomérico de la CDP-metileritritol quinasa: la formación del

homodímero no es necesaria para la actividad enzimática. 45
3.1.4 Clonación molecular y caracterización de la 1-desoxi-D-xilulosa 5-fosfato de la bacteria termofílica *Thermotoga maritima* 46
3.1.5 Procesamiento autoproteolítico de la 1-desoxi-D-xilulosa 5-fosfato sintasa 47
3.2 DISCUSIÓN 50
3.3 CONCLUSIONES 57

4. BIBLIOGRAFÍA 61

5. PUBLICACIONES 67

Capítulo 1

Imitando interacciones directas proteína-proteína e interacciones mediadas por solvente en el homodímero CDP-metileritritol quinasa: Una aproximación de búsqueda virtual dirigida a través de farmacóforo. 69

Mimicking direct protein-protein and solvent-mediated interactions in the CDP-methylerythritol kinase homodimer : A pharmacophore-directed virtual screening approach. 71

Capítulo 2

Diseño de una molécula pequeña y un péptido cíclico natural como ligandos de la CDP-metileritritol quinasa imitando interacciones proteína-proteína e interacciones mediadas por solvente. 91

Design of a small molecule and a natural cyclic peptide as ligands of CDP-methylerythritol kinase by mimicking direct protein-protein and solvent-mediated interactions 93

Capítulo 3

Análisis del estado oligomérico de la CDP-ME quinasa de *E.coli*: la formación del homodímero no es necesaria para la actividad enzimática 115

Analysis of the oligomeric state of *E.coli* CDP-ME kinase: Homodimer formation is not necessary for enzyme activity 117

Capítulo 4

Clonación molecular y caracterización de la 1-desoxi-D-xilulosa 5-fosfato de la bacteria termofílica *Thermotoga maritima* 135

Molecular cloning and Characterization of 1-deoxy-D-xylulose 5-phosphate synthase from the thermophilic bacterium *Thermotoga maritima* 137

Capítulo 5

Procesamiento autoproteolítico de la 1-desoxi-D-xilulosa 5-fosfato sintasa 153

Autoproteolytic processing of 1-deoxy-D-xylulose 5-phosphate synthase 155

ANEXO1 173

Moléculas y efecto sobre la actividad enzimática de CMK 175
Hipótesis farmacofórica de 5 puntos 175
Hipótesis farmacofórica de 7 puntos 179

ANEXO2 181

INTRODUCCIÓN GENERAL

1. INTRODUCCIÓN GENERAL

1.1 Los isoprenoides

Los isoprenoides son una de las familias de compuestos más extensas que se conocen en la naturaleza con más de 30000 compuestos naturales descritos, con una muy diversa estructura y función (Sacchettini and Poulter 1997). Los isoprenoides están presentes en todos los reinos de la vida y participan en un amplio abanico de funciones que van desde procesos indispensables para las células, hasta productos del metabolismo secundario, éstos últimos de gran importancia en las plantas. Es en este grupo precisamente, las plantas, en el que los isoprenoides alcanzan una mayor complejidad estructural y funcional (Chapman and Hill 2000). A pesar de su gran diversidad de funciones y estructuras todos los isoprenoides se sintetizan a partir de dos precursores comunes, de 5 átomos de C, el isopentenil difosfato (IPP) y su isómero estructural el dimetilalil difosfato (DMAPP)(Fig.1).

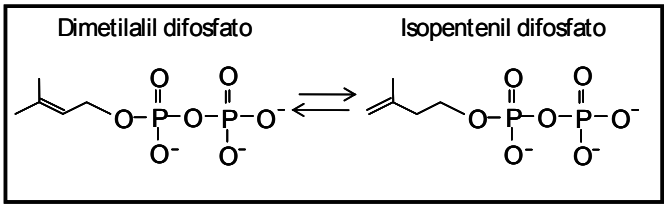


Fig. 1. Imagen de los isómeros isopentenil difosfato (IPP) y dimetilalil difosfato (DMAPP), precursores universales de los isoprenoides.

La síntesis de la mayor parte de los isoprenoides se realiza mediante la condensación secuencial de moléculas de IPP sobre una molécula de su isómero, el DMAPP (Fig.2). Como resultado se obtienen compuestos isoprenoides con un número creciente de átomos de carbono y múltiplo de 5 llamados prenildifosfatos. A continuación se llevan a cabo otras reacciones para obtener isoprenoides de mayor tamaño como la condensación de dos moléculas de farnesil difosfato para formar escualeno (C_{30}) o de dos moléculas de geranilgeranil difosfato para formar fitoeno (C_{40}) catalizadas respectivamente por las enzimas escualeno sintasa y fitoeno sintasa. Seguidamente los prenildifosfatos obtenidos experimentan una serie de reacciones (hidroxilación, metilación, isomerización, desmetilación, reducción…) para dar lugar al amplio abanico de compuestos isoprenoides presentes en los distintos organismos.

Los isoprenoides se han clasificado históricamente según el número de átomos que contienen y en base a los monoterpenos (C_{10}) que fueron los primeros isoprenoides identificados. Según esta nomenclatura tenemos hemiterpenos (C_5), monoterpenos (C_{10}), sesquiterpenos (C_{15}), diterpenos (C_{20}), sesterpenos (C_{25}), triterpenos (C_{30}), tetraterpenos (C_{40}) y politerpenos.

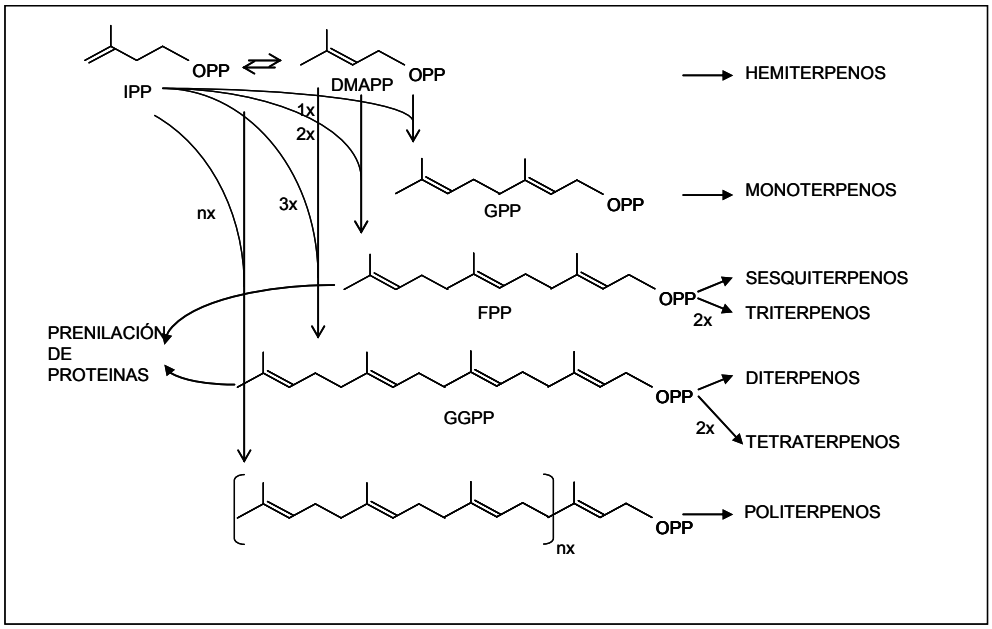


Fig. 2. Esquema del proceso de condensación secuencial de moléculas de IPP sobre una molécula de DMAPP en la síntesis de los compuestos isoprenoides.

Los isoprenoides incluyen compuestos bacterianos interesantes como los hopanos, componentes importantes para la pared bacteriana, las ubiquinonas que participan en la cadena de transporte electrónico y las menaquinonas, vitaminas liposolubles (vitamina K2) normalmente sintetizadas por la flora bacteriana y que no pueden ser sintetizadas por el ser humano. En las plantas se encuentran compuestos isoprenoides esenciales como la ubiquinona y la plastoquinona que intervienen en la cadena de transporte electrónico mitocondrial y fotosintética, respectivamente así como la cadena lateral de las clorofilas. También contienen moléculas como las citoquininas, los brasinosteroides, los esteroles, el ácido abcísico y las giberelinas implicadas en la regulación del crecimiento y del desarrollo; las proteínas preniladas que son importantes para la señalización y la regulación del ciclo celular y los carotenoides que actúan como protectores de la maquinaria fotosintética (Chappell 1995; McGarvey and Croteau 1995; Lange, Wildung et al. 1998; Harker and Bramley 1999). En los animals cabe destacar también las quinonas, los dolicoles, los esteroides y los retinoides (McGarvey and Croteau 1995) (véase ejemplos de compuestos isoprenoides en la tabla 1).

Terpenos		Ejemplos	Funciones	Distribución en los organismos
C_5	Hemiterpenos	isopentenil-tRNA	traducción proteica	todos los organismos
		citoquininas	hormona vegetal	plantas superiores
C_{10}	Monoterpenos	esencias de flores/plantas (mentol)	comunicación entre plantas/actividad antimicrobiana	plantas superiores y algas
			aromas, perfumes	
C_{15}	Sesquiterpenos	aceites	micotoxinas/actividad antimicrobiana	plantas superiores, hepàticas, musgos, hongos, algas
		fitoalexinas (capsidiol)		
C_{20}	Diterpenos	fitol, grupo lateral de la clorofila	cadenas de transporte electrónico	plantas superiores y algas
		grupo lateral del tocoferol y filoquinonas		
		fitoalexinas	actividad antimicrobiana	plantas superiores,hongos
		giberelinas	hormona vegetal	
		taxol	antitumoral	
C_{30}	Triterpenos	esteroles	fluideza de las membranas	todos los organismos
		brasinoesteroides	hormona vegetal	plantas superiores
		otros triterpenos (saponina)	interacción con el entorno	
C_{40}	Tetraterpenos	carotenoides	protectores del aparato fotosintético	plantas superiores
			pigmentos, vitamina A	
$C_{>40}$	Politerpenos	ubiquinona	cadenas de transporte electrónico	todos los organismos
		plastoquinona		
		dolicol	glicosilación de proteinas	

Tabla 1. Tabla con ejemplos de compuestos isoprenoides y su distribución, clasificados según el número de átomos de carbono que contienen.

Actualmente se conocen dos rutas metabólicas para la biosíntesis de IPP y DMAPP: la ruta del mevalonato (MVA) y la ruta de metileritritol 4-fosfato (MEP). En la década de los 50, se identificó la ruta del MVA de síntesis de IPP en levaduras y animales (Rodríguez-Concepción and Boronat 2002) y se extrapoló su presencia a todos los organismos. No obstante, con el tiempo, se observó que la síntesis de algunos terpenos no se podía explicar considerando únicamente esta ruta metabólica (Lichtenthaler 1997; Eisenreich, Schwarz et al. 1998; Lange, Wildung et al. 1998; Lichtenthaler 1999; Rohmer 1999; Rodríguez-Concepción and Boronat 2002). En la década de los 90, a partir de experimentos de marcaje llevados a cabo en plantas y bacterias se pudo identificar una ruta alternativa de síntesis de IPP y DMAPP, la cual se denominó como ruta alternativa y, posteriormente, ruta del MEP.

1.1.1La ruta del mevalonato o ruta clásica

La ruta del MVA de síntesis de IPP fue la primera descubierta, por lo que también se denomina ruta clásica. Bloch y Rittemberg establecieron las bases experimentales para el descubrimiento de la ruta al confirmar la formación de esteroles a partir de acetil-CoA (Rawn 1989). La ruta del mevalonato empieza con la condensación secuencial de tres moléculas de acetil-coA en dos pasos enzimáticos para formar 3-hidroxi-3-metilglutaril-coA (HMG-CoA). En la etapa siguiente la enzima HMG-CoA reductasa, cataliza la reducción de HMG-CoA a mevalonato en la primera reacción irreversible de la ruta. A esta enzima se le ha atribuido el papel de catalizar la etapa limitante de la ruta. Por ello, y por ser el mevalonato el primer compuesto específico de la ruta, se la denominó ruta del mevalonato. A partir del mevalonato y mediante dos fosforilaciones, dependientes de ATP, seguidas de una descarboxilación se forma el IPP (Fig. 3). Estas etapas están catalizadas respectivamente por las enzimas mevalonato quinasa (MVK), fosfomevalonato quinasa (PMK) y fosfomevalonato decarboxilasa (PMD o MDD). Es interesante comentar que las dos primeras pertenecen a la superfamilia Galactoquinasa, Homoserina quinasa, Mevalonato quinasa y Fosfomevalonato quinasa (o GHMP, véase apartado 1.3.1). Finalmente, el IPP es transformado a DMAPP por acción de la enzima isopentenil difosfato isomerasa.

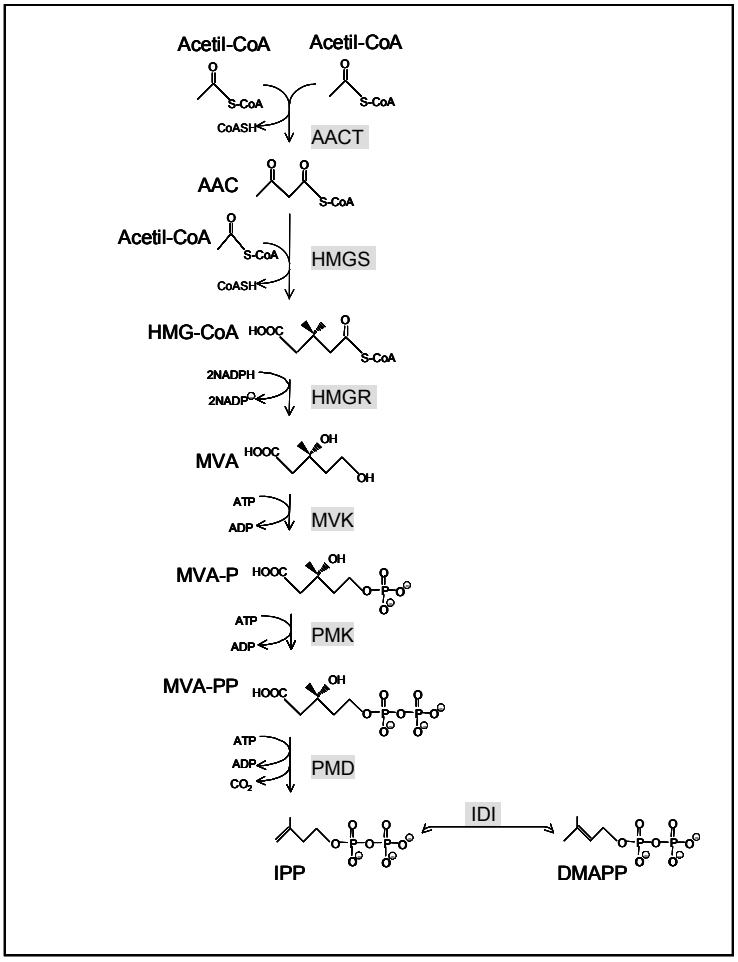


Fig. 3. Esquema de la ruta clásica o del MVA de síntesis de IPP y DMAPP. **AAC**, acetoacetil-CoA, **HMG-CoA**, 3-hidroxi-3-metilglutaril-CoA, **MVA**, mevalonato, **MVA-P**, MVA 5-fosfato, **MVA-PP**, MVA 5-difosfato, **IPP**, isopentenil difosfato, **DMAPP**, dimetilalil difosfato, **AACT**, ACC tiolasa, **HMGS**, HMG-CoA sintasa, **HMGR**, HMG-CoA reductasa, **MVK**, MVA quinasa, **PMK**, MVA-P quinasa, **PMD**, MVA-PP descarboxilasa, **IDI**, IPP-DMAPP isomerasa

1.1.2 La ruta del metileritritol 4-fosfato (MEP) o ruta alternativa

Los experimentos de Rohmer y colaboradores en los que estudiaron la incorporación de distintos precursores marcados con ^{13}C (Glucosa, acetato, eritrosa y piruvato) en hopanoides de diferentes eubacterias permitió proponer un ruta de síntesis de IPP a partir de triosas fosfato (Rohmer 1999). Estos experimentos se basaron en observaciones anteriores (Pandian 1981) según las cuales el marcaje radiométrico en las cadenas laterales de la ubiquinona en distintos organismos procariotas no se ajustaba al patrón esperable si la biosíntesis tuviera lugar mediante la ruta del mevalonato. La hipótesis de Rohmer fue corroborada posteriormente mediante experimentos realizados con cepas de *E.coli* defectivas en enzimas de la ruta glucolítica, consiguiéndo definir exactamente el origen de todos los átomos de C presentes en la molécula de IPP.

Se propuso que la primera reacción de la ruta debía ser similar a la descrita para las transcetolasas, donde la condensación de la hidroxietiltiamina, derivada del piruvato, con el grupo aldehído del gliceraldehído 3-fosfato (Rohmer 1999) formaría desoxixilulosa 5-fosfato (DXP). Esta hipótesis corroboraba el marcaje preferencial de los carbonos 1 y 5 en los experimentos de marcaje de IPP y que resultaba incompatible con la hipótesis de la ruta del MVA según la cual se esperaría un marcaje preferencial de los carbonos 2, 3 y 4. A la enzima que cataliza esta reacción se la denominó como 1-desoxi-D-xilulosa 5-fosfato sintasa (DXS) (Sprenger, Schorken et al. 1997; Lois, Campos et al. 1998). En la etapa siguiente se produce un reordenamiento intramolecular de la pentosa fosfato (DXP), a través de un proceso dependiente de NADPH para formar 2-C metileritritol 4-fosfato (MEP). Este metabolito proviene de una reacción irreversible y es el primer compuesto específico de la ruta independiente de MVA de síntesis de IPP, por lo que por analogía con la ruta del MVA se la pasó a denominar la ruta del MEP. La reacción está catalizada por la enzima 1-desoxi-D-xilulosa 5-fosfato reductoisomerasa (DXR) (Kuzuyama, Shimizu et al. 1998; Kuzuyama, Takahashi et al. 2000). En la etapa siguiente el MEP es transformado en 4-difosfocitidil-2C-metil-D-eritritol (CDP-ME) en una reacción dependiente de citidin trifosfato (CTP) que cataliza la enzima 4-difosfocitidil-2C-metil-D-eritritol sintasa (Rohdich, Wungsintaweekul et al. 1999). A continuación, el CDP-ME es fosforilado por la enzima 4-difosfocitidil-2C-metil-D-eritritol quinasa (Rohdich, Wungsintaweekul et al. 2000). Esta quinasa pertenece a la superfamilia de quinasas GHMP. A continuación el CDP-MEP se transforma en 2-C-metil-D-eritritol 2,4-ciclodifosfato y citidin monofosfato (CMP) por acción de la enzima 2-C-metil-D-eritritol 2,4-ciclodifosfato sintasa (MCS). En el siguiente paso enzimático se produce la apertura del anillo del 2-C-metil-D-eritritol 2,4-ciclodifosfato en un paso de reducción produciéndose 1-hidroxi-2-metil-2(E)butenil difosfato mediante una reacción catalizada por la enzima 1-hidroxi-2-metil-2(E)butenil difosfato sintasa (HDS) (Querol, Rodriguez-Concepcion et al. 2001; Seemann, Campos et al. 2002). Por último, se obtienen los productos finales de la ruta del MEP: IPP y DMAPP en una proporción de 5:1 (Jomaa, Wiesner et al. 1999; Adam, Hecht et al. 2002; Rohdich, Hecht et al. 2002) por acción de la enzima 1-hidroxi-2-metil-2(E)butenil difosfato reductasa (HDR) (Fig.4).

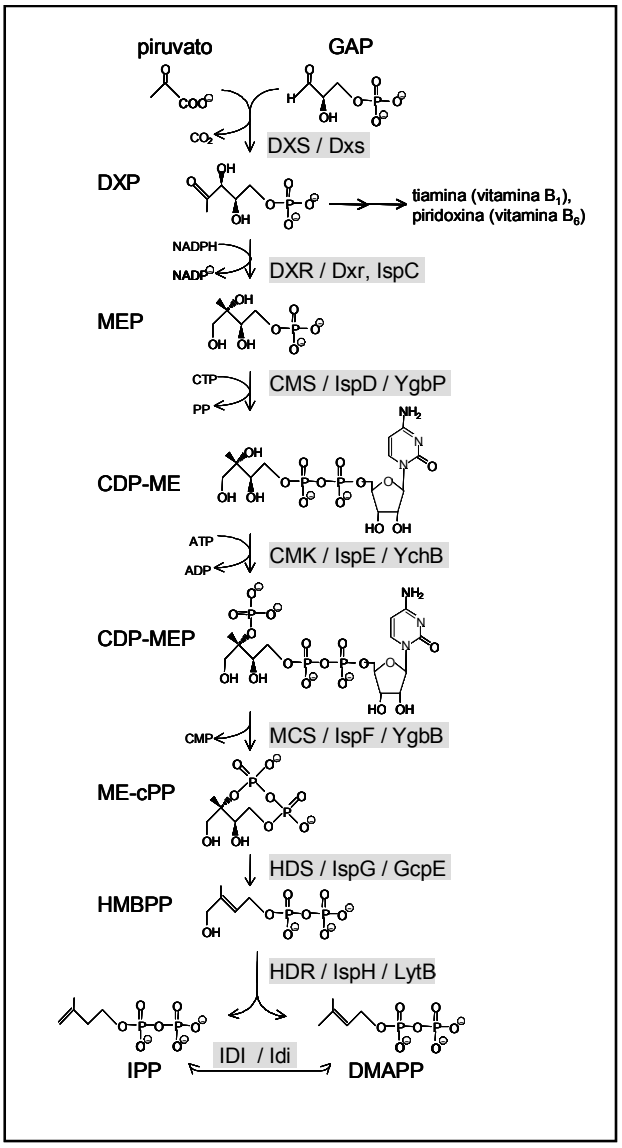


Fig. 4. Esquema de la ruta alternativa o del MEP de síntesis de IPP y DMAPP. **GAP**, gliceraldehído 3-fosfato, **DXP**, 1-desoxi-D-xilulosa 5-fosfato, **MEP**, 2-C-metil-D-eritritol 4-fosfato, **CDP-ME**, 4-difosfocitidil-2-C-metil-D-eritritol ,**CDP-MEP**, CDP-ME 2-fosfato, **ME-cPP**, 2-C-metil-D-eritritol 2,4-ciclodifosfato, **HMBPP**, hidroximetilbutenil 4-difosfato, **IPP**, isopentenil difosfato, **DMAPP**, dimetilalil difosfato, **DXS**, DXP sintasa, **DXR**, DXP reductoisomerasa, **CMS**, CDP-ME sintasa, **CMK**, CDP-ME quinasa, **MCS**, ME-cPP sintasa, **HDS**, HMBPP sintasa, **HDR**, HMBPP reductasa. También se incluyen las nomenclaturas bacterianas y de *E.coli* para las enzimas.

1.1.3 Distribución de las rutas del MVA y del MEP

El descubrimiento de la ruta del MEP obligó a revisar la biosíntesis de los isoprenoides en todos los organismos. Los esfuerzos de secuenciación genómica que se han ido produciendo en estos últimos tiempos han permitido la localización de las diferentes rutas de síntesis de IPP. Actualmente se sabe que la ruta del MVA está presente en exclusividad en arqueobacterias, en algunos protozoos, en hongos y en animales. La mayoría de las eubacterias, por el contrario, sintetiza el IPP y DMAPP a partir de la ruta del MEP, aunque algunas de ellas los sintetizan a partir de la ruta del MVA. Ejemplo de eubacterias con la ruta del MEP son *E. coli* y las cianobacterias, los

antecesores de los plastos (Lichtenthaler 1997; Lichtenthaler 1999; Lichtenthaler 2000; Eisenreich, Rohdich et al. 2001). La ruta del MEP también está presente en las algas verdes y en protozoos como el parásito de la malaria *Plasmodium falciparum* (Lichtenthaler 1997; Lichtenthaler 1999; Lichtenthaler 2000; Eisenreich, Rohdich et al. 2001).

Finalmente, cabe destacar como muy interesante el hecho de que algunas algas, musgos, helechos y plantas superiores compartan ambas rutas de síntesis de los precursores de isoprenoides. Esta coexistencia se produce en compartimentos celulares separados (Lichtenthaler 1997; Lichtenthaler, Rohmer et al. 1997; Lange, Wildung et al. 1998; Lichtenthaler 2000; Eisenreich, Rohdich et al. 2001), ya que mientras la ruta del MVA se encuentra en el citosol y el retículo endoplasmático, la ruta del MEP se localiza en los plastos (Lange and Croteau 1999) (Fig.5). La ruta del MVA acaba produciendo los compuestos isoprenoides citosólicos y proporciona IPP para la biosíntesis de compuestos isoprenoides mitocondriales; mientras que la ruta del MEP aporta el IPP necesario para la biosíntesis de los compuestos isoprenoides de origen plastídico. En los últimos años se ha puesto mucho interés en intentar determinar la existencia de un posible trasvase de los precursores entre ambas rutas así como de una posible comunicación entre ellas.

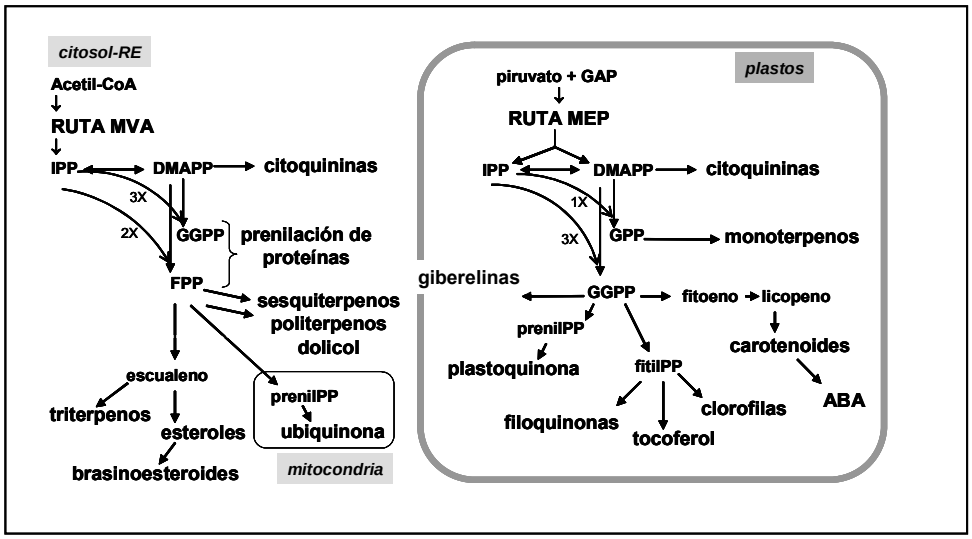


Fig. 5. Esquema de la síntesis de isoprenoides derivados de la ruta del MVA y de la ruta del MEP.

1.1.4 Origen evolutivo de las rutas del MVA y del MEP

Debido a que los compuestos isoprenoides son uno de los productos naturales más antiguos conocidos, además de tener una interesante complejidad generada a partir de una molécula común, el IPP, ha hecho del estudio de su historia evolutiva un punto importante. Se ha sugerido que las plantas superiores habrían adquirido la ruta del MEP a partir de ancestros procariotas fotosintéticos, de acuerdo con la teoría endosimbiótica (Margulis 1975; Strassburger 1988). Este hecho viene sugerido por la presencia de la ruta del MEP en cianobacterias. Paralelamente, la célula huésped eucariota mantendría la ruta del MVA presente en animales y levaduras, que muy probablemente, provendría de las arqueobacterias. La diversidad observada en la presencia de las rutas en eubacterias se explicaría por procesos de transferencia horizontal de genes (Boucher and Doolittle 2000).

1.1.5 La ruta del MEP representa una atractiva diana para la búsqueda de antibióticos, antimaláricos y herbicidas

La ruta del MEP está presente en muchos patógenos incluidos los agentes causales de algunas meningitis bacterianas, de la lepra, de trastornos gastrointestinales, de enfermedades de transmisión sexual, de la tuberculosis y de ciertos tipos de neumonía, e incluso en el protozoo causante de la malaria, *Plasmodium falciparum.* Por otro lado, la ruta del MEP está ausente en animales, y dado el carácter esencial de los compuestos isoprenoides, la ruta del MEP se presenta como una diana para la búsqueda de nuevos compuestos antibióticos, antimaláricos y herbicidas (Kuzuyama, Shimizu et al. 1998; Jomaa, Wiesner et al. 1999; Rohmer 1999; Lichtenthaler 2000).

Desde el descubrimiento de los primeros antibióticos y su uso masificado, el problema de la aparición de la resistencia a múltiples drogas (MDR, siglas en inglés de multidrug resistance) en organismos patógenos ha obligado a la búsqueda incesante de nuevas dianas terapéuticas. Las características ya comentadas de la ruta del MEP la hacen una prometedora diana para desarrollar nuevos compuestos que permitan combatir el problema de la MDR.

La fosmidomicina es un ejemplo paradigmático de esta nueva clase de compuestos terapéuticos. La fosmidomicina es un antibiótico natural extraído de la actinobacteria *Streptomyces lavendulae*. Este compuesto inhibe la enzima que cataliza la segunda etapa de la ruta del MEP, la desoxixilulosa 5-fosfato reductoisomerasa (DXR), tiene una estructura similar al substrato y actúa como un inhibidor competitivo. Se ha comprobado que la fosmidomicina y sus derivados inhiben el crecimiento bacteriano y son eficaces en el tratamiento de las infecciones causadas por *Plasmodium* (Jomaa, Wiesner et al. 1999; Borrmann, Issifou et al. 2004). Una terapia combinada de fosmidomicina y clindamicina, un compuesto químico que inhibe la maquinaria de traducción proteica procariota en los apicoplastos de *Plasmodium falciparum*, se ha mostrado eficaz contra la malaria en un tratamiento de 3 días con una tasa de éxito del 95% (Borrmann, Issifou et al. 2004).

Otro ejemplo de inhibidor es el clomazón. Este herbicida conocido desde los años 80, actúa de forma indirecta, ya que son sus productos de degradación los que inhiben la desoxixilulosa 5-fosfato sintasa (DXS) (Jomaa, Wiesner et al. 1999; Mueller, Schwender et al. 2000), y dan lugar a un bloqueo en la acumulación de los isoprenoides plastídicos (Lange and Croteau 1999).

1.2 Búsqueda de nuevos compuestos. Introducción al diseño racional de nuevos fármacos

Históricamente, el descubrimiento de nuevos compuestos activos se realizaba a través de la observación fenotípica de sus efectos. Con el descubrimiento y avance de la ciencia, se potenciaron procesos de criba de infinidad de productos naturales sobre una determinada diana terapéutica. Y aunque estos métodos han permitido el descubrimiento de innumerables compuestos terapéuticos hasta el día de hoy, son procesos largos y costosos.
La determinación de la estructura atómica de las proteínas mediante técnicas de difracción de rayos X, cuyo primer exponente fue la mioglobina, junto con avances informáticos que mediante complejos métodos computacionales unidos a la química teórica han permitido estudios teóricos y el tratamiento de grandes cantidades de datos han conducido al desarrollo de una nueva disciplina, la modelización molecular. La modelización molecular se erige en una eficaz, práctica y útil herramienta para el descubrimiento de nuevos fármacos, ya que, en parte, disminuye los costosos procesos de búsqueda.
La modelización molecular se podría definir como una disciplina que utiliza técnicas computacionales basadas en química-física teórica (química cuántica, mecánica molecular), resultados experimentales (estructuras de cristalización, experimentos de cálculo de constantes de disociación…) y algoritmos rápidos de visualización de estructuras para estudiar estructuras moleculares e interacciones de los complejos formados. Una de las aplicaciones más utilizada de la modelización molecular es el diseño de fármacos asistido por ordenador. Esta estrategia, en condiciones óptimas, utiliza la información estructural obtenida previamente de un complejo receptor-fármaco, aunque en la mayoría de los casos se parte sólo de información estructural del receptor. En estos últimos casos hay que tener presente que la conformación que se obtenga puede que no sea la conformación bioactiva. Finalmente, también se pueden llegar a hacer modelos teóricos sin tener información estructural ni del receptor ni del compuesto.

1.2.1 Mecánica molecular

Debido al gran número de átomos que contienen la mayoría de las biomoléculas, la aproximación viable al estudio de este tipo de complejos es por mecánica molecular. La mecánica molecular parametriza la energía potencial del sistema, permitiendo un cálculo rápido y la realización de minimizaciones de energía y dinámicas. La mecánica molecular tiene la restricción de no poder estudiar reacciones ni cualquier proceso en que haya un cambio de estado electrónico debido a que se basa en las Leyes de Newton. Aún así, para estos estudios, no se descarta el uso de las parametrizaciones de pequeñas moléculas orgánicas (sustratos, cofactores…) por química cuántica con métodos semiempíricos.

1.2.1.1 Parametrización de la energía: Campos de fuerza

La función energía potencial parametrizada en función de las coordenadas de los átomos de la biomolécula se conoce como campo de fuerza. Existen varios campos de fuerza, pero los utilizados en este trabajo han sido los parametrizados bajo el paquete de programas AMBER (AMBER7) (Case, Cheatham et al. 2005). El campo de

fuerza es un potencial clásico que corresponde al sumatorio de diferentes propiedades estructurales y físicas. El campo de fuerza se divide en potenciales de enlace (V_R), de ángulo (V_{ang}), de diedro (V_Φ), impropios (V_ω), de van der Waals (V_{vdW}) y electroestático (V_{el}).

$$V = \Sigma V_r + \Sigma V_{ang} + \Sigma V_\Phi + \Sigma V_\omega + \Sigma V_{vdW} + \Sigma V_{el}$$

Los potenciales de enlace y de ángulo son funciones parabólicas dependientes de la distancia interatómica obtenidas a partir de datos estructurales. La energía aumenta rápidamente al aumentar la distancia entre los átomos, por lo que no se pueden estudiar sistemas en los que este enlace se rompa (la distancia se haga lo suficientemente grande). Para el potencial de ángulo el tratamiento es similar pero con tres átomos para definirlo. El potencial de diedro intenta describir todo el perfil energético que se produce por la rotación de los diedros. El potencial de diedros impropios se utiliza para los casos en los que los cuatro átomos no formen físicamente un ángulo diedro, siendo muy útil en la parametrización de moléculas orgánicas. El potencial de van der Waals intenta reproducir la repulsión electrónica a distancias muy cortas y las interacciones atractivas de London cuando los átomos se alejan. Finalmente, cuando la distancia se hace grande ambas fuerzas se acercan a 0. El potencial de energía electroestática se describe mediante el potencial de Coulomb, siendo interacciones muy fuertes. Los potenciales de van der Waals y de energía electroestática no se aplican a pares de átomos separados por uno o dos enlaces, y por tres se aplican con restricciones.

1.2.1.2 Métodos de optimización

Existen varios métodos de optimización estructural que dependen de su precisión y su coste computacional. Los pasos previos de optimización son necesarios tanto para estudios estructurales como para posteriores procesos de dinámica molecular. Existe la posibilidad de que coordenadas de ciertos átomos no estén determinadas o incluso, como en el caso de difracción de rayos X, la técnica empírica no es capaz de determinar la posición de los átomos de hidrogeno. En estos casos, se tiene que completar el sistema con su consecuente tensionamiento, y su alejamiento del mínimo energético. Estos métodos libran de tensión al sistema, partiendo de un grado de minimización energética, obteniéndose estructuras más coherentes con los datos experimentales.

1.2.1.3 Dinámica molecular

Una vez minimizado el sistema, la mecánica molecular permite el cálculo de trayectorias dinámicas ya que se dispone de la energía potencial y las coordenadas iniciales de los átomos. La información que se obtiene es de gran utilidad, debido a que la información empírica, proveniente de difracción de rayos X, da poca información sobre la dinámica a temperatura ambiente. El proceso de obtención cristalográfica se produce a temperaturas bajas por lo que se obtiene una información de un sistema aparentemente congelado, en cambio la dinámica molecular permite añadir baños térmicos para obtener información a la temperatura deseada. Paralelamente, la dinámica molecular permite calcular la trayectoria en un sistema acuoso, obteniendo información más parecida al sistema natural de la proteína. El estudio de las interacciones de un complejo a lo largo del tiempo permite ver la estabilidad de los mismos de forma cuantitativa. Otra de las ventajas de la dinámica molecular es que permite

discernir entre interacciones producidas por el empaquetamiento cristalino. Para las dinámicas moleculares se suele utilizar como paso de integración 1/10 del tiempo de vibración más rápido, que correspondería al enlace C-H con 10fs. Como se ha comentado, la dinámica molecular se puede llevar a cabo teniendo un equilibrio térmico con el entorno añadiendo un baño térmico a través del algoritmo de Berendsen (Berendsen, Postma et al. 1984).

1.2.1.3.1 Solvatación del sistema

El agua posee un rol esencial en todos los procesos biológicos y evidentemente su presencia determina la estructura de las proteínas mediante su relación con los grupos hidrofílicos e hidrofóbicos. Normalmente en los estudios de dinámica molecular se solvata el sistema. En los estudios realizados se ha utilizado una caja de aguas TIP3P (Jorgensen, Chandrasekhar et al. 1983), que simula el agua pura a temperatura ambiente, lo suficientemente grande para que solvate toda la proteína. Posteriormente, con el paquete de programas AMBER v7 (Case, Cheatam et al. 2005) se duplicaron las cajas para la simulación mejor del solvente e impedir artefactos debidos a la finalización abrupta de átomos.

Las simulaciones se pueden realizar a volumen constante o a presión constante. En los primeros pasos de la dinámica se aumenta la temperatura hasta la deseada a volumen constante, después se ajusta la presión a la deseada y se realizan simulaciones a volumen y temperatura constante. Una vez llegado al equilibrio se empieza a quitar restricciones. Finalmente, a partir de un mínimo de energía convergido se realiza el tiempo de producción dinámica que se usa para el estudio de las propiedades del sistema.

1.2.1.3.2 Contraiones

Las biomoléculas pueden tener una carga neta por lo que en ocasiones se puede necesitar la inclusión de contraiones al sistema para poder realizar las dinámicas moleculares estables. Estos iones pueden ser Na^+ y Cl^-.

1.2.1.3.3 Cálculo de propiedades

Una vez realizada la producción dinámica se pueden calcular las interacciones no enlazantes y las energías libres de formación de los complejos, que son especialmente útiles en el diseño de fármacos.

1.2.2 Búsqueda en bases de datos 3D

Cada vez existen más bases de datos de organizaciones públicas y privadas en las que se muestran sus compuestos químicos. Estas bases de datos proveen la estructura tridimensional de los compuestos y normalmente contienen rápidos algoritmos para la búsqueda de la mejor conformación. A partir de estas bases de datos se permite hacer un cribado masivo de compuestos que cumplan los requisitos conformacionales y las restricciones a los grupos químicos impuestos por nuestra hipótesis farmacofórica. Durante nuestro estudio se ha utilizado la base de datos Catalyst™ (Accelrys Inc.) cribando los compuestos de las siguientes instituciones: Available Chemical Directory (ACD), Derwen World Drug Index, Nat diverse, IBS 2004, Aurora, National Cancer Institute, Maybridge, SPECS y ChemDiv New Chemistry.

1.2.2.1 Hipótesis farmacofórica

A partir de los datos geométricos obtenidos en la producción dinámica, se puede diseñar la hipótesis farmacofórica. La definición del farmacóforo seria un conjunto de características estéricas y electrónicas que son necesarias para asegurar las interacciones supramoleculares con una diana biológica específica y que permiten o bloquean su respuesta biológica. Esta hipótesis farmacofórica, una vez escogida la diana, contiene unos ciertos tipos de átomo o grupos con características químicas (como grupos ácidos y básicos, aromáticos, hidrofóbicos, aceptores y donadores de protones…) a los que se denomina puntos farmacofóricos. Los farmacóforos también se definen por las distancias, ángulos y planos entre los puntos farmacofóricos, los cuales, para obtener una mayor precisión, se obtienen a partir de la dinámica molecular (véase como ejemplo la fig. 6).

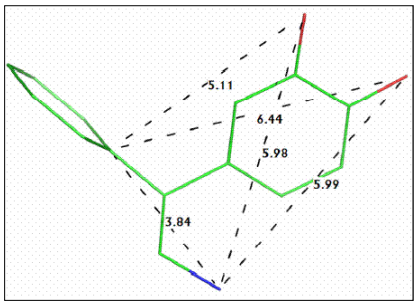


Fig.6. Farmacóforo de antagonistas del receptor de dopamina (Ursu 2006).

1.2.2.2 Anclaje de los compuestos (o *Docking*)

Una vez encontradas conformaciones de moléculas en las bases de datos, se procede a realizar el anclaje al receptor. Existen diferentes metodologías para realizar este anclaje basadas en información de dinámica molecular, algoritmos genéticos o basados en información de la flexibilidad del receptor. En nuestros estudios el programa utilizado fue Dock-dyn (Rubio-Martinez, Pinto et al. 2005). En un primer paso, este programa recomprueba los compuestos que satisfacen la hipótesis farmacofórica de referencia. Los puntos farmacofóricos son definidos como hidrofóbicos, carga positiva, carga negativa, aceptor de protones, donador de protones y anillo aromático. Estos puntos están determinados en un intervalo de coordenadas con un máximo y un mínimo extraído durante la dinámica molecular. El programa busca todas las posibles combinaciones de las conformaciones con la hipótesis farmacofórica y calcula las distancias. Para las combinaciones que cumplan las distancias, finalmente, realiza una superposición de los puntos de la molécula con los de la hipótesis farmacofórica extraída de la dinámica molecular y se determina el RMSD (desviación de la raíz cuadrada de las medias). En un segundo paso, las conformaciones que tengan un RMSD por debajo de un valor estimado se superponen al farmacóforo de referencia con el receptor. En este paso se evalúa el reconocimiento molecular y se descartan moléculas que conlleven algún impedimento estérico. El descarte se hace en función de que la distancia entre átomos de la molécula y el receptor sea menor a un porcentaje del radio de van der Waals de la molécula y del receptor.

1.2.2.3 Ordenación de los compuestos

Una vez las diferentes conformaciones son aceptadas en el proceso de anclaje, los compuestos pueden ser ordenados mediante diferentes programas. En este estudio utilizamos tanto el XSCORE (Wang and Lai, et al. 2002) como el mismo RMSD para ordenar las diferentes conformaciones obtenidas. El programa XSCORE calcula la energía libre de unión del complejo ligando-receptor, permitiendo un cálculo rápido debido a su parametrización. Una vez ordenados los compuestos, se hace una última criba visual, en dónde se escogen moléculas de familias diferentes de compuestos, que no tengan grupos químicos problemáticos… Una norma comúnmente aceptada en la búsqueda de nuevos fármacos es la regla del 5 de Lipinski. Una vez escogidos los compuestos se realizan los ensayos empíricos para comprobar la interacción ligando-receptor y calcular las constantes de disociación del complejo.

1.2.2.4 Regla del 5 de Lipinski

Lipinski a través del tratamiento estadístico de drogas comerciales describió una norma sencilla (Lipinski, Lombardo et al. 2001) que tendrían que cumplir las moléculas para, a priori, poder tener una mejor penetración a través de la bicapa lipídica y tener una mejor absorción y/o permeabilidad celular. Esta regla se basa en: a) un peso menor de 500 ua b) menos de 5 donadores de protones c) menos de 10 (5x2) aceptores de protones y d) tener un logP (octanol/agua) menor de 5.

Antes o una vez obtenidos resultados empíricos positivos, se puede continuar con estudios de dinámica molecular para caracterizar las interacciones del complejo, estudiar la energía libre de formación del complejo e incluso mejorar la unión mediante modificaciones en la molécula dirigidas por los estudios de la interacción.

1.2.3 Uso de péptidos

La utilización de pequeños péptidos, que ya bien por interacción con la diana escogida, o por mimetismo con la misma proteína, puede activar o bloquear el proceso biológico es otra estrategia para la búsqueda de compuestos con capacidades terapéuticas. Incluso, los hace interesantes para el estudio de interacciones proteína-proteína y en la búsqueda de interferir estas interacciones. Los péptidos pueden sufrir proteólisis y para solucionar este problema, e incluso para reducir la libertad conformacional en solución del péptido, se utilizan péptidos cíclicos. Esta mayor rigidez puede reducir su entropía perdida en la unión obteniéndose una ganancia en los valores de afinidad del péptido. Esta ganancia de afinidad de los péptidos ciclados en referencia a los respectivos péptidos no ciclados está bien descrita en la literatura (Bach II, Eyermann et al. 1994; Morgan, Holland et al. 1994; Zang, Yu et al. 1998; Dumez, Snaith et al. 2002; Andrews, McInnes et al. 2004; Nam, Ye et al. 2004; Chang and Chu 2005).

1.2.4 Técnicas empíricas de caracterización de la interacción ligando-receptor

Con el avance de la ciencia, y de la proteómica en particular, se han ido desarrollando cada vez más técnicas de caracterización de interacciones proteína-proteína, proteína-ADN, proteína-ligando, etc. Dentro de esta gran amalgama de técnicas se incluyen desde técnicas cualitativas como la difracción de rayos X, hasta métodos cuantitativos que permiten la determinación de las constantes del complejo como la calorimetría de titración isotérmica, ensayos de immunoabsorción ligado a enzimas (ELISA), ensayos radioimmunes (RIA), ensayos basados en espectrometría de masas, resonancia del plasmón en superficie, etc.
Durante nuestro estudio, la interacción ligando-receptor de las moléculas obtenidas durante la criba guiada por farmacóforo se estudió mediante las técnicas de disminución de la intensidad con láser de desorción/ionización asistido por matriz – tiempo de vuelo (*Intensity-fading* MALDI-TOF MS) y de resonancia del plasmón en superficie.

1.2.4.1 Disminución de la intensidad por láser de desorción/ionización asistido por matriz-tiempo de vuelo (*Intensity-fading* MALDI-TOF MS)

Esta técnica, descrita por Yanes y Villanueva, se basa en el uso de las intensidades iónicas de la técnica MALDI de espectrometría de masas, para detectar complejos unidos por interacciones no covalentes entre moléculas (proteína-proteína, proteína-ADN, proteína-molécula pequeña) no inmovilizadas (Villanueva, Yanes et al. 2003; Yanes, Nazabal et al. 2006). En esta técnica una de las moléculas tiene que ser de naturaleza proteica. El complejo es detectado por la disminución de la intensidad iónica del compuesto al añadirle el interactor en diferentes concentraciones a la muestra. La disminución de la intensidad iónica de la molécula observada es directamente asociada a la formación del complejo. Esta técnica puede aplicarse a nivel cualitativo o semicuantitativo añadiendo un patrón interno con el cual comparar la disminución de la intensidad iónica.

1.2.4.2 Resonancia del plasmón en superficie

La tecnología de la resonancia del plasmón en superficie se ha mostrado como una herramienta eficaz en el análisis cualitativo y cuantitativo de las interacciones de un complejo. Esta tecnología permite la parametrización en tiempo real de la formación del complejo sin necesidad de utilizar añadidos u otros interaccionadores. Se basa en la inmovilización del receptor a un chip y se mide la unión directa del ligando. La formación del complejo se estudia por cambios en la resonancia del plasmón en superficie, que están directamente asociados a cambios en el índice de refracción de la solución próxima al receptor inmovilizado. El índice de refracción está directamente asociado a la concentración en la capa de la superficie del chip. Por tanto, este índice de refracción aumenta cuando el complejo entre el ligando y el receptor se forma.

1.3 La 4-difosfocitidil-2C-metil-D-eritritol quinasa (CMK)

La 4-difosfocitidil-2C-metil-D-eritritol quinasa (CMK) es la enzima que cataliza el cuarto paso enzimático de la ruta del MEP. En este paso se fosforila el hidroxilo del C2 del CDP-ME para dar CDP-MEP en una reacción dependiente de ATP y magnesio (Kuzuyama, Takahashi et al. 2000; Luttgen, Rohdich et al. 2000; Rohdich, Wungsintaweekul et al. 2000). Hasta la fecha se ha determinado la estructura cristalina de la CMK de *Escherichia coli* (Miallau, Alphey et al. 2003), *Thermus thermophilus* (Wada, Kuzuyama et al. 2003) y *Aquifex aeolicus* (Sgraja, Alphey et al. 2008). La estructura cristalina de la CMK de *E.coli* (Miallau, Alphey et al. 2003) muestra una organización dimérica, en la que la unidad asimétrica consta de dos subunidades, A y B, con una masa total aproximada de 62 kDa, las cuales se ensamblan en una simetría C2. Esta estructura dimérica fue corroborada mediante experimentos de exclusión molecular, espectrometría de masas (Miallau, Alphey et al. 2003) y ultracentrifugación diferencial (Gabrielsen, Bond et al. 2004). Es interesante destacar que tan solo el 4% de la superficie de la proteína interviene en la formación del dímero (Fig.7). La estructura cristalina muestra que la formación del dímero se sustenta mediante interacciones electroestáticas entre el residuo Arg^{21} y el Asp^{80} del otro monómero, puentes de hidrógeno entre la amida del residuo Gly^{87} y el carbonilo del residuo Ala^{22}, interacciones hidrofóbicas entre las cadenas laterales de la Tyr^{25} y la Lys^{76} del otro monómero y una serie de puentes de hidrogeno mediados por solvente (Miallau, Alphey et al. 2003).

En la primera parte de la tesis se realiza un estudio estructural por dinámica molecular utilizando como base empírica la estructura cristalina de *E.coli* (PDB=1oj4). El estudio se centrará en la poca superficie de la proteína implicada en las interacciones importantes para la formación del dímero.

Esta pequeña quinasa pertenece a la superfamilia GHMP (llamada así por las enzimas galactoquinasa, homoserina quinasa, mevalonato quinasa y fosfomevalonato quinasa) (Bork, Sander et al. 1993; Lange and Croteau 1999; Bonanno, Edo et al. 2001; Wada, Kuzuyama et al. 2003).

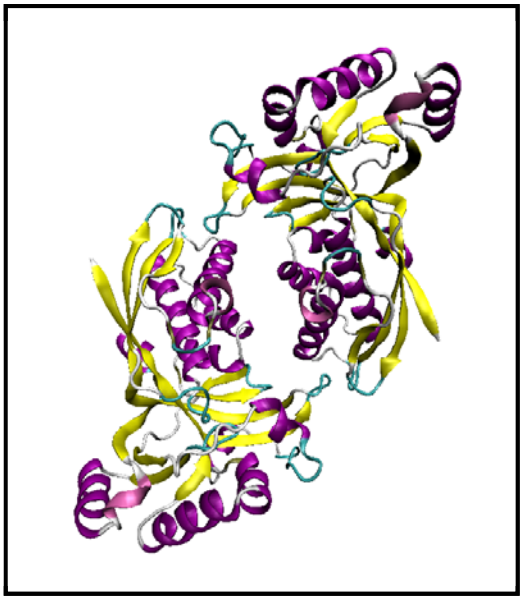


Fig. 7. Representación de la estructura de la CMK de *E.coli* (PDB=1oj4, (Miallau, Alphey et al. 2003)) realizada con el programa de visualización VMD (Humphrey, Dalke et al. 1996).

1.3.1 La superfamilia GHMP

La CMK forma parte de la superfamilia GHMP de quinasas. A esta familia pertenecen también tres enzimas de la ruta del MVA como son la MVK, la PMK y la MDD. Las enzimas de esta familia catalizan la transferencia de un grupo γ-fosforilo a un receptor particular, con la excepción de la proteína no enzimática Xol-1 que interviene en procesos de regulación del desarrollo (Luz, Hassig et al. 2003). Aunque la similitud de secuencia entre las proteínas de esta superfamilia es relativamente baja (10-20%), todas ellas conservan una estructura muy parecida (un intervalo de RMSD entre 2.6 y 4 Å entre los Cα). Esta superfamilia presenta un plegamiento característico de dos dominios α+β, con el centro catalítico inmerso en una hendidura entre ellos (Cheek, Zhang et al. 2002). La parte aminoterminal de estas proteínas se caracteriza por tener dos láminas β y cuatro hélices α, mientras que la parte carboxiterminal de estas proteínas tiene un centro parecido a ferredoxina y cuatro hélices α adicionales. Todas las proteínas de esta superfamília también contienen tres motivos conservados. El motivo I se encuentra localizado en el extremo aminoterminal de la proteína, con siete aminoácidos muy conservados. El motivo II (llamado rizo P) tiene la secuencia conservada PXXXGSSAA y se caracteriza por ser el motivo de unión al nucleótido. Este motivo es estructuralmente diferente al motivo clásico de unión a nucleótidos llamado motivo "Walker A" disponiendo a la molécula de ATP en una orientación diferente. Finalmente, el motivo III es un rizo rico en glicinas encargado de la unión al sustrato que tiene que ser fosforilado (Andreassi and Leyh 2004). Este elevado contenido en glicinas se considera importante debido a la alta libertad de rotación del aminoácido, lo que permitiría un mejor ensamblaje de la proteína al sustrato.

El hecho de que todas las proteínas de esta superfamilia tengan un esqueleto muy parecido parece indicar un mismo origen evolutivo, proviniendo de una pequeña proteína con actividad quinasa (véase fig 8).

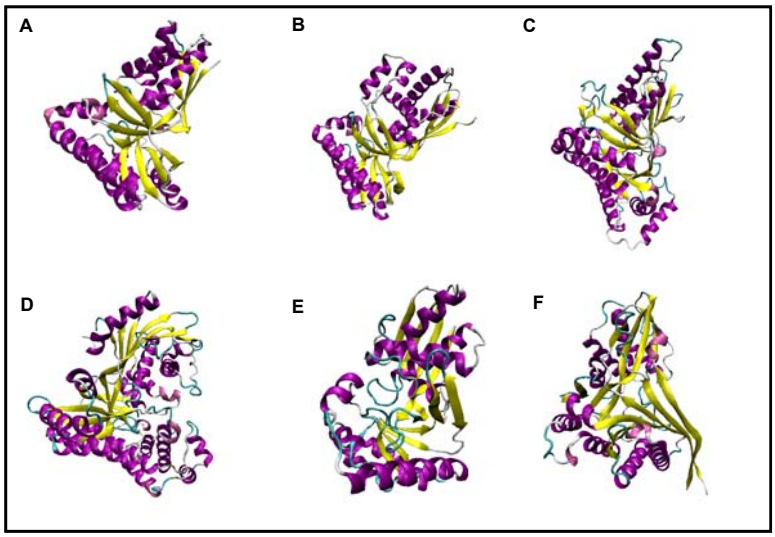


Fig. 8. Representación de la estructura monomérica de diferentes enzimas de la superfamilia galactoquinasa, mevalonato quinasa, homoserina quinasa y fosfomevalonato quinasa realizadas con el programa de visualización VMD (Humphrey, Dalke et al. 1996). A. Mevalonato quinasa de *Streptococcus pneumoniae* (2oi2). B. Fosfomevalonato quinasa de *Streptococcus pneumoniae* (1k47). C. Difosfomevalonato decarboxilasa de *Sacharomyces cerevisiae* (1fi4). D. Galactoquinasa de *Saccharomyces cerevisiae* (2aj4). E. Homoserina quinasa de *Methanococcus jannaschii* (1h72). F. Xol-1 de *Caenorhabditis elegans* (1mg7).

1.4 La 1-desoxi-D-xilulosa 5-fosfato sintasa (DXS)

La segunda parte de esta tesis se ha centrado en el estudio de la primera enzima de la ruta del MEP. La enzima DXS cataliza la condensación de piruvato y gliceraldehido 3-fosfato para formar 1-desoxi-D-xilulosa 5-fosfato (DXP) (Lange and Croteau 1999). Esta reacción del tipo transcetolasa es dependiente de pirofosfato de tiamina (TPP). La DXP es un precursor de la ruta del MEP de síntesis de IPP y DMAPP y se ha caracterizado a la DXS como una de las enzimas reguladoras de la ruta del MEP (Estevez, Cantero et al. 2001). Es interesante señalar que, la DXP también es el precursor de las coenzimas tiamina (Sprenger 1997) y del piridoxol en un pequeño grupo de γ-proteobacterias (Fig. 9). Ambas coenzimas son muy importantes ya que son necesarias para la actividad catalítica de muchas enzimas. La DXP entra en la ruta de síntesis de TPP en la formación del anillo de hidroxietiltiazol (Frank, Leeper et al. 2007). Es interesante remarcar el hecho de que la DXS sea a su vez dependiente de TPP, ya que podría sugerir un tipo de regulación de retroalimentación positiva dentro de una ruta con una regulación de por si muy interesante y complicada (Frank, Leeper et al. 2007). También se descubrió que *E.coli* condensaba 4-(fosfohidroxi)-L-threonina y DXP para formar la vitamina B_6 (piridoxal 5-fosfato) (Cane D.E 1998), y en un principio se postuló como la ruta de síntesis de la vitamina B_6. Posteriormente se descubrió que existe una ruta de síntesis independiente de DXP en *Saccharomyces cerevisiae* (Bean, Dvorachek et al. 2001), *Cercospora nicotinidae* (Ehrenshaft, Bilski et al. 1999; Osmani, May et al. 1999) y *Arabidopsis thaliana* (Titiz, Tambasco-Studart et al. 2006; Wagner, Bernhardt et al. 2006); y que la anterior sólo está presente en un pequeño grupo de γ-proteobacterias.

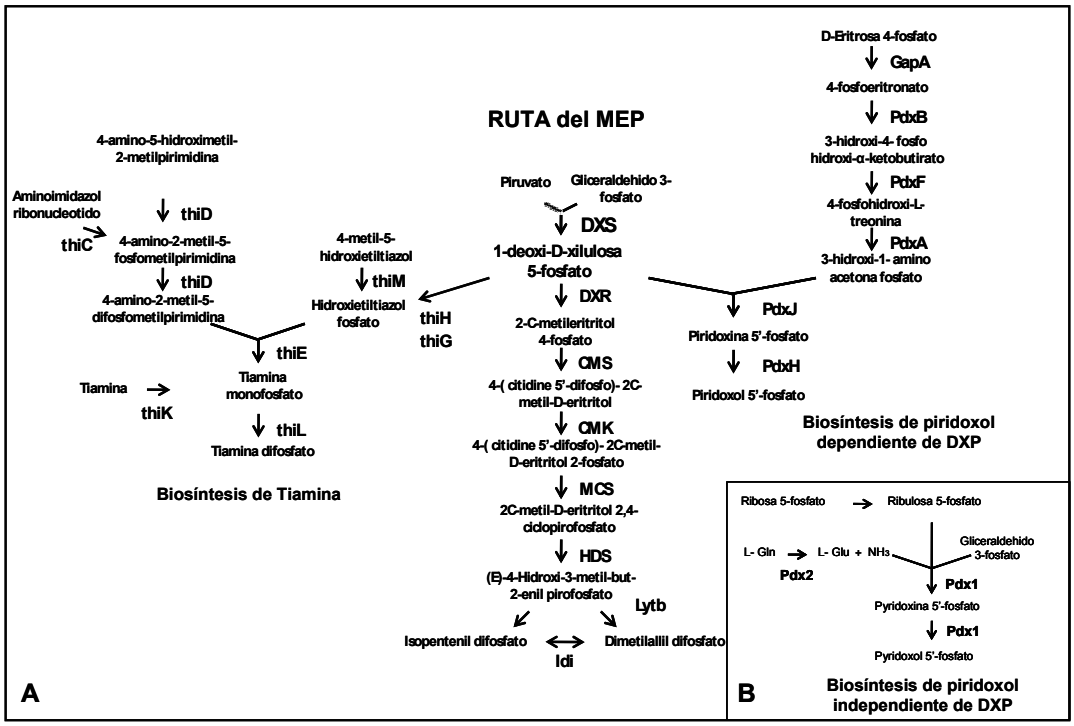


Fig. 9. A. Representación esquemática de las rutas de síntesis de la ruta del MEP y las rutas de síntesis de tiamina y de piridoxol dependiente de DXP. B. Representación esquemática de la ruta de síntesis de piridoxol independiente de DXP.

Hasta la fecha, se han caracterizado homólogos de la DXS a partir un gran número de organismos presentándose, incluso, como una familia multigénica dentro del grupo de las plantas. A pesar del gran número de DXS estudiadas, no se ha descrito hasta el momento ninguna caracterización de una DXS de un organismo termofílico.

Durante largo tiempo, la búsqueda de la estructura de la DXS mediante técnicas de cristalografía de proteínas y difracción de rayos X había sido un hito para muchos laboratorios, entre ellos el nuestro, debido al carácter crucial de la enzima. En el año 2006, Xiang y sus colaboradores pudieron finalmente presentar la estructura de las DXS de *Escherichia coli* y *Deinococcus radiodurans* (Xiang, Usunow et al. 2007). No obstante, para dar idea de la complejidad del estudio es importante señalar que la estructura descrita corresponde no a la enzima completa sino a una forma truncada. Esta forma truncada se había obtenido mediante proteólisis debida a la contaminación por un hongo de las soluciones utilizadas en la búsqueda de condiciones de cristalización (Xiang, Usunow et al. 2007).

1.4.1 Información estructural de la DXS

La DXS aparece con una organización homodimérica tanto en los análisis cristalográficos (Xiang, Usunow et al. 2007) (Fig.10) como en experimentos de dispersión de luz y electroforesis en geles nativos de poliacrilamida (Kuzuyama, Takahashi et al. 2000; Hahn, Eubanks et al. 2001). La interacción del dímero presenta un carácter fuertemente hidrofóbico y los monómeros se presentan con una distribución paralela, en la que cada dominio de un monómero está en contacto con el mismo dominio del otro monómero. Cada monómero presenta tres dominios con un plegamiento típico α/β. El dominio I (aminoácidos del 1 al 319) contiene 5 láminas β paralelas, el dominio II (aminoácidos del 320 al 495) presenta 6 láminas β paralelas y el dominio III (residuos del 496 al 629) contiene 5 láminas β paralelas y una antiparalela. Todas las láminas β se presentan rodeadas por hélices α. Cabe destacar que debido al proceso de proteólisis de origen fúngico, en la estructura de la DXS de *E.coli* las zonas correspondientes a los aminoácidos 183–238 y 292–317 no presentaban densidad electrónica (Xiang, Usunow et al. 2007). La primera zona está cerca del centro catalítico y la segunda se ubicaría en la zona de unión entre el dominio I y el dominio II. Durante el estudio de cristalización consiguieron rellenar estos vacíos partiendo de la estructura de la DXS de *D. radiodurans*, la cual no sufría el proceso de proteólisis. Cuando la estructuras obtenidas se compararon con las depositadas en las bases de datos se observó que eran similares a las descritas para las transcetolasas, la subunidad E1 de la piruvato deshidrogenasa y la 2-oxoisovalerato deshidrogenasa, como ya se había predicho en función de la actividad de la enzima (Rohmer 1999). A diferencia de éstas, no obstante, el dominio I de un monómero está situado por encima de los dominios II y III del mismo monómero, lo que da lugar a la formación del centro catalítico dentro de una misma cadena polipeptídica.

Fig. 10. Representación de la estructura de la DXS de *E.coli* (PDB=2o1s, (Xiang, Usunow et al. 2007)) realizada con el programa de visualización VMD (Humphrey, Dalke et al. 1996).

1.5 Proteólisis

1.5.1 Estabilidad del enlace peptídico

El enlace amida formado entre el grupo α-carboxilo de un aminoácido y el grupo α-amino de otro aminoácido, denominado enlace peptídico, se caracteriza por ser un enlace metaestable (Mathews 2003). La metaestabilidad es debida a que, aunque la reacción está favorecida termodinámicamente hacia la hidrólisis del enlace (cambio de energía libre >10kJ mol^{-1}), esta reacción es extremadamente lenta a pH y temperatura fisiológicos. La reacción de hidrólisis se puede acelerar mediante catalizadores o condiciones extremas de temperatura y pH.

Entre los catalizadores biológicos más estudiados se encuentran las enzimas proteolíticas o proteasas. Las enzimas proteolíticas las podríamos definir como proteínas capaces de catalizar la hidrólisis de un enlace peptídico de otra molécula proteíca. Estas proteasas se pueden tipificar a través de su mecanismo de actividad (metalo-, cistein-, serin- proteasas, etc). Las proteasas intervienen en infinidad de procesos celulares, que van desde sistemas de reciclaje (como la ubiquitinación o el sistema lisosómico), apoptosis celular (caspasas), angiogénesis, remodelación de la matriz, iniciación de cascadas de señalización hasta la regulación enzimática de rutas metabólicas. La importancia de estas enzimas proteolíticas se refleja en el hecho de que el genoma humano codifica para más de 500 proteasas diferentes (Puente, Sanchez et al. 2005).

A pesar de todo ello, cada vez se están identificando más procesos en los que la hidrólisis del enlace peptídico se produce sin la mediación de una proteína externa (o proteasa). Estos procesos tienen funciones biológicas importantes, ya sea desde el procesamiento de la preproteína hasta su forma funcional, la señalización celular o procesos de transferencia genética horizontal (Paulus 2000).

Existen mecanismos de autoproteólisis intermolecular y procesos de autoproteólisis intramolecular. La principal

diferencia entre ellos es la presencia de la enzima catalizadora de la hidrólisis del enlace peptídico en el primer mecanismo y su ausencia en la segundo. En una primera aproximación podriamos definir como mecanismos en *trans* los primeros y en *cis* los segundos. Aunque, como se describirá a continuación, se han descrito algunas excepciones de mecanismos autoproteolíticos intramoleculares en *trans* (véase sección 1.5.3.4, *splicing* proteico).

1.5.2 Mecanismos de autoproteólisis intermolecular

Los mecanismos de autoproteólisis intermolecular se caracterizan por el hecho que una molécula de la proteína ejerce de catalizador de la segmentación del enlace peptídico a otra molécula de su misma especie. Los procesos intermoleculares se caracterizan por que la aparición de los productos de la proteólisis está directamente relacionada en segundo grado con la cantidad de proteína, es decir, a mayor concentración de proteína inicial se produce un incremento exponencial de la lisis (Rosenblum and Blobel 1999; Ingram, Orth et al. 2006). Esto es debido a que la proteína a la vez es sustrato y catalizador. Esta característica difiere de los mecanismos intramoleculares en los que la proteólisis es directamente proporcional en primer grado en un amplio espectro de concentraciones.

Un ejemplo de este tipo de mecanismos es el caso de TACE o ADAM17, una metaloproteasa que procesa el factor necrótico tumoral α para que la célula sea capaz de liberarlo (Utsumi, Levitan et al. 1993) y, a su vez, sufre un proceso de autoproteólisis intermolecular en el cual el enlace peptídico entre los residuos Tyr^{352} y Val^{353} es hidrolizado (Ingram, Orth et al. 2006).

1.5.3 Mecanismos de autoproteólisis intramolecular

Los procesos de autoproteólisis intramolecular se caracterizan por dar lugar a una modificación post-traduccional de la proteína que concluye con la segmentación del enlace peptídico. Para poder producirse la segmentación del enlace peptídico, éste necesita modificarse previamente a un enlace con menor estabilidad, o producirse una reorganización de la cadena lateral del aminoácido anterior para, finalmente, romperse mediante diferentes mecanismos. La estructura terciaria de la proteína proporciona las condiciones necesarias para que se produzcan las reorganizaciones de los enlaces. Para esta introducción se han clasificado los procesos según esta reorganización inicial y nos centraremos en dos: mecanismos que parten de una reorganización de los grupos acilo N-O o N-S y de mecanismos relacionados con el enlace Asx(Asp y Asn)-Aa.

1.5.3.1 Reorganización de los grupos acilo N-O o N-S

Existen varios procesos autoproteolíticos que tienen en común un primer paso en el que se produce la reorganización de grupos acilo N-O o N-S, cuya consecuencia es el cambio de un enlace peptídico por un enlace (tio)éster entre los aminoácidos. Para que se produzca esta reorganización es necesaria la presencia de residuos aminoacídicos de Ser, Cys o Thr en la parte carboxiterminal del enlace. Las cadenas laterales de estos residuos aminoacídicos llevan a cabo un ataque nucleofílico sobre el carbono del carboxilo del enlace peptídico precedente, para formar un enlace de tipo éster (en caso de que el aminoácido sea una serina o treonina) o tioéster (en el caso de que sea una cisteína) (Paulus 2000). Se ha postulado que estos ataques nucleofílicos se producen a través de la formación previa de un intermediario oxioxazolidino u oxitiozolidino (Fig.11). El enlace éster formado entre los aminoácidos es menos estable que el enlace amida y puede reaccionar frente a un ataque nucleofílico produciéndose la segmentación. La reorganización de los grupos acilo N-O o N-S no tiene lugar espontáneamente en condiciones normales de pH (Iwai and Ando 1967), pero puede producirse en condiciones acídicas.

La estructura de la proteína es importante para estabilizar el nuevo enlace (tio)éster producido debido a que el equilibrio de la reacción está desplazado hacia a la reorganización inversa de los grupos acilo O-N o S-N, formándose de nuevo el enlace peptídico. El plegamiento mediante interacciones electroestáticas y de van der Waals, mantiene una conformación que permite el ataque nucleofílico, la formación del enlace y su estabilización. Las proteínas desnaturalizadas son incapaces de formar el enlace (tio)éster.

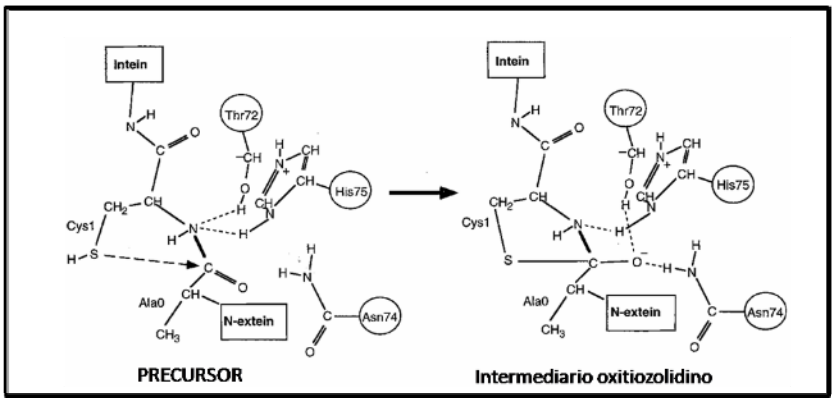


Fig. 11. Representación del ataque nucleofílico, reorganización de los grupos acilo N-S y la formación del enlace tioéster, a través de un intermediario oxitiozolidino, de la proteína GyrA de *M. Xenopi*. Representación modificada a partir de la esquematizada por Paulus (Klabunde, Sharma et al. 1998; Paulus 2000).

Cabe destacar que el proceso de formación del enlace (tio)éster presenta aspectos similares al descrito para las serinproteasas; incluso llegando a revisarse estos mecanismos autoproteolíticos en cierta literatura como serinproteasas no convencionales (Ekici, Paetzel et al. 2008).

El enlace (tio)éster puede sufrir diferentes reacciones y acabar rompiéndose o modificándose de la manera descrita en los diferentes mecanismos. En las enzimas en las que se forman derivados de piruvoilo, denominadas a partir de este momento como enzimas piruvoiladas, se produce la modificación del aminoácido carboxiterminal del enlace peptídico inicial; en las aminoterminal nucleófilo hidrolasas (Ntn hidrolasas), un ataque nucleofílico del solvente hidroliza el enlace; en las proteínas "erizo" (*Hedgehog* en inglés), se produce una transesterificación intermolecular con el colesterol y en el empalme de proteínas (*protein splicing* en inglés), la transesterificación tiene lugar dentro de la misma molécula (Fig.12).

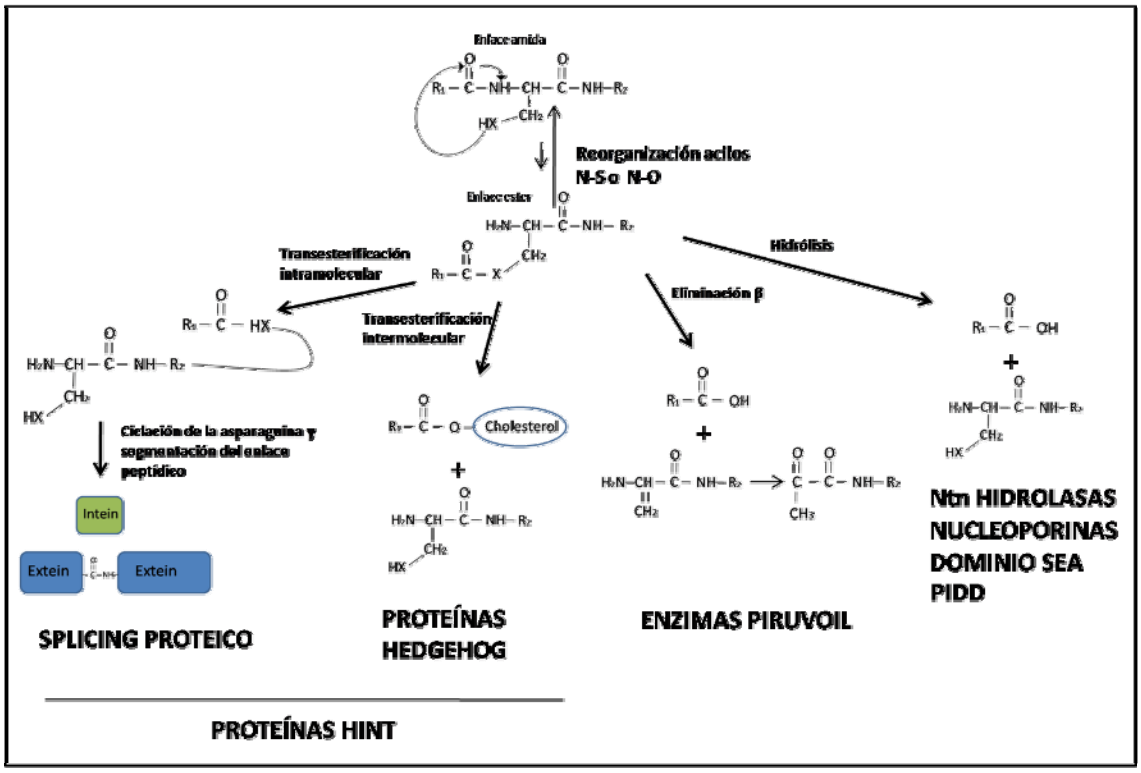


Fig. 12. Representación de los principales mecanismos de autoproteólisis intramolecular descritos a partir de la reorganización de los grupos acilo N-S o N-O y la formación del enlace (tio)éster. Modificación de la representación esquematizada por Paulus (Paulus 2000). La X en el esquema representa los átomos O o S. Ntn hidrolasas (aminoterminal nucleófilo hidrolasas), PIDD (proteína inducida por p53 con dominio de muerte)(Tinel, Janssens et al. 2007).

1.5.3.1.1 Enzimas piruvoiladas

Las enzimas piruvoiladas participan en la formación de aminas a partir del metabolismo catabólico de los aminoácidos en animales, plantas y microorganismos (revisado en (Recsei and Snell 1984; van Poelje and Snell 1990)). Estos pasos enzimáticos de formación de aminas incluyen descarboxilaciones y reducciones que utilizan el piridoxol como cofactor necesario para la reacción. En 1953 Rodwell postuló que la histidina decarboxilasa del organismo *Lactobacillus 30a* era independiente de piridoxol 5-fosfato (Rodwell 1953). A partir de este trabajo se fue identificando un grupo de enzimas del catabolismo de aminoácidos independiente del piridoxol 5-

P. Estas enzimas se caracterizan por tener un grupo piruvoilo unido a la proteína. El grupo prostético es capaz de formar una base de Schiff con el sustrato y permite la descarboxilación de éste, de una manera similar al cofactor piridoxol 5-fosfato. Estas enzimas se traducen como una cadena simple (π) y a partir de la formación del enlace éster, mediante la reorganización intramolecular entre dos serinas contiguas, y una eliminación β se produce la segmentación del enlace anterior (Fig.13). A partir de la eliminación β se obtienen dos cadenas polipeptídicas, una larga (α) y una más corta (β). Finalmente, se produce una transformación de la Ser de la cadena α a un grupo piruvoilo. La mayoría de estas enzimas presentan una oligomerización heterohexamérica en la que participan ambas subunidades (Recsei and Snell 1984), aunque existen algunos ejemplos en los que la oligomerización funcional se basa en la subunidad α únicamente. Se ha observado que mutaciones en los residuos implicados en el proceso de autocatálisis de la proteína inhiben la actividad de la enzima.

Ejemplos de las enzimas piruvoiladas son la histidina descarboxilasa, la S-adenosilmetionina descarboxilasa, la fosfatidilserina descarboxilasa, la α-aspartato descarboxilasa y la prolina reductasa. Los dos mecanismos de catálisis de aminoácidos (piruvoilo y piridoxol 5-fosfato) se distribuyen filogenéticamente, como por ejemplo en el caso de la histidina descarboxilasa que es de tipo enzima piruvoilo en bacterias gram positivas pero es piridoxol 5-fosfato dependiente en animales y bacterias gram negativas (Recsei and Snell 1984).

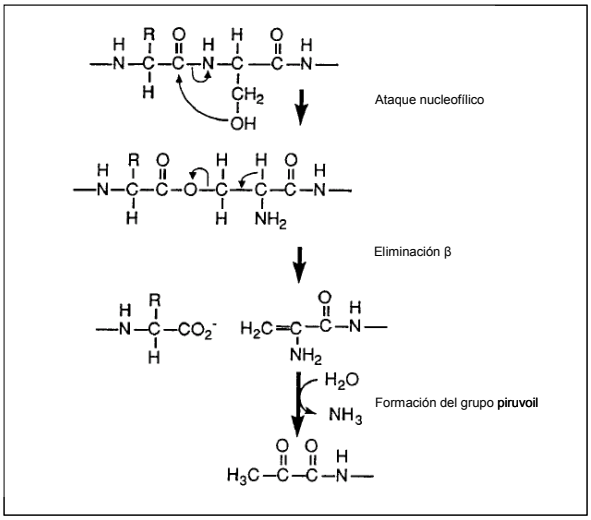


Fig. 13. Representación del ataque nucleofílico, formación del grupo éster, la eliminación β y la formación del grupo piruvoilo descrita para las enzimas piruvoiladas. Este esquema es una modificación del descrito por Xiong (Xiong, Stanley et al. 1997).

1.5.3.1.2 Ntn hidrolasas, nucleoporinas y proteínas con el dominio SEA

En 1995 Branningan y colaboradores (Brannigan, Dodson et al. 1995) definieron una nueva superfamilia de estructura de proteínas y la llamaron Ntn-hidrolasas (aminoterminal nucleofilo hidrolasas). Estas proteínas presentan un núcleo con una estructura αββα, en el cual dos láminas β antiparalelas empaquetadas entre ellas se encuentran rodeadas por hélices α antiparalelas (Artymiuk 1995; Brannigan, Dodson et al. 1995). Todas las Ntn hidrolasas conocidas se procesan mediante un mecanismo autocatalítico post-traduccional. A pesar de que su estructura es similar, las Ntn hidrolasas no tienen similitud de secuencia entre ellas, tan solo conservándose los residuos implicados en este proceso. Dentro de las Ntn hidrolasas se encuentran diferentes tipos de enzima como las penicilina G acilasas (Brannigan, Dodson et al. 1995; McDonough, Klei et al. 1999), las penicilina V acilasas

(Suresh, Pundle et al. 1999), los proteasomas (Bochtler, Ditzel et al. 1997; Groll, Ditzel et al. 1997), la glucosamina 6-fosfato sintasa (Isupov, Obmolova et al. 1996), la aspartilglucosaminidasa (Oinonen, Tikkanen et al. 1995), etc.

El procesamiento autocatalítico del enlace peptídico se produce a través de un residuo Thr, Ser o Cys carboxiterminal, siendo la reactividad de éste modulada por el tipo de residuo aminoterminal del enlace. El enlace éster entre los residuos es estabilizado por la formación de un intermediario covalente, a través de un estado de transición, gracias a residuos del denominado hoyo oxianión (*oxyanion hole*, del inglés) (Oinonen and Rouvinen 2000). Finalmente, el ataque nucleofílico de una molécula de agua permite la segmentación del enlace obteniéndose dos cadenas polipeptídicas (Fig.14). Cabe destacar el hecho que estos residuos de Thr, Ser o Cys son importantes tanto en el autoprocesamiento como en la actividad enzimática de la proteína.

Como en las enzimas piruvoiladas, estas proteínas pueden presentar una oligomerización con heterómeros de las dos cadenas recién segmentadas (Boanca, Sand et al. 2006) y el procesamiento de la preproteína se presenta indispensable para la actividad enzimática.

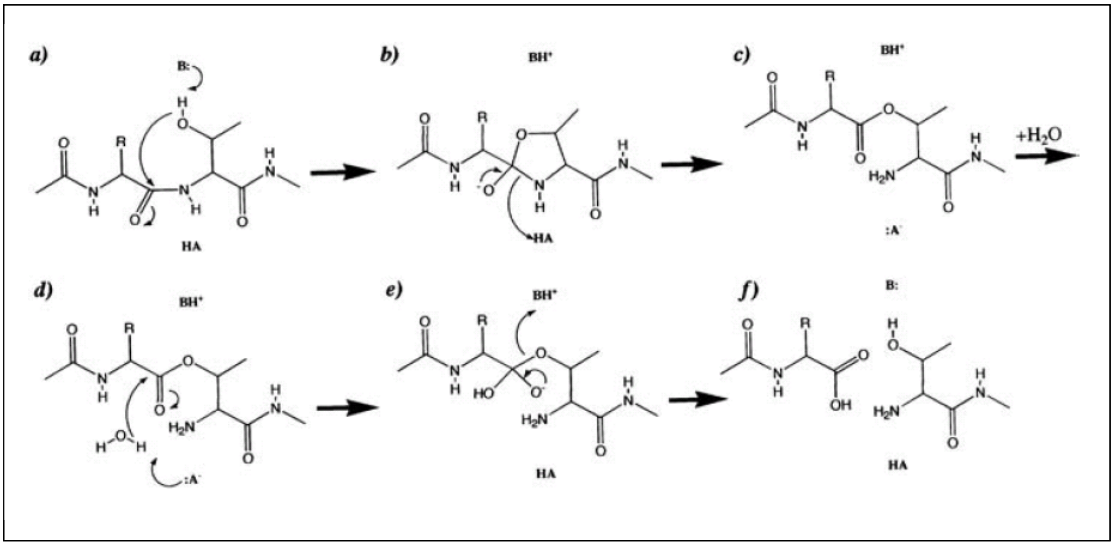


Fig. 14. Mecanismo de autoproteólisis propuesto por Saarela y colaboradores para la aspartilglucosaminidasa humana (Saarela, Oinonen et al. 2004).

Aunque hemos denominado a este grupo de procesos con el nombre de las familias de proteínas, hubiera sido más apropiado destacar la importancia del solvente en el proceso final de segmentación del enlace. Esto es debido a que en los últimos tiempos están apareciendo cada vez más proteínas en las que el agua acaba produciendo el ataque nucleofílico catalizador de la segmentación. Otros ejemplos de la importancia del solvente serían las nucleoporinas (Rosenblum and Blobel 1999) y las proteínas con el dominio SEA (Levitin, Stern et al. 2005; Johansson, Macao et al. 2008).

Las nucleoporinas tienen importancia en el transporte de ARN y proteínas entre el núcleo y el citosol (Rosenblum and Blobel 1999). Hasta la fecha se ha caracterizado el mecanismo de autoproteólisis intramolecular en la nucleoporina 98 de mamíferos (Rosenblum and Blobel 1999) y la nucleoporina 145 de *Saccharomyces cerevisiae* (Emtage, Bucci et al. 1997). El mecanismo es similar al descrito en las Ntn hidrolasas, en las cuales el

agua realiza el ataque nucleofílico que permite la segmentación del enlace éster creado entre la Phe^{863} y la Ser^{864} de la nucleoporina 98 (Rosenblum and Blobel 1999). Esta segmentación se produce a 6 kDa del extremo carboxiterminal de la proteína. Se ha demostrado la importancia de la estructura para la realización del proceso, corroborando la necesidad de la creación de un intermediario capaz de estabilizar el enlace éster. El proceso de autoproteólisis se ha caracterizado esencial para el ensamblaje del complejo del poro nuclear.

MUC1, una proteína transmembrana altamente expresada en varios tipos de carcinomas (Hilkens, Buijs et al. 1984), también sufre un mecanismo de autoproteólisis en el dominio SEA (se descubrió en una proteína del **e**sperma, en una **e**nteroquinasa y en la **a**grina) (Cohen and Parry 1986; Lillehoj, Han et al. 2003; Levitin, Stern et al. 2005). El mecanismo es similar al descrito para las Ntn hidrolasas y las nucleoporinas, donde el enlace éster es segmentado finalmente por el ataque nucleofílico del agua (Levitin, Stern et al. 2005). El procesamiento autocatalítico se produce entre los residuos Gly^{62} y Ser^{63}, siendo erradicado cuando el residuo serina es mutado a cualquier aminoácido con excepción de la cisteína o la treonina (Levitin, Stern et al. 2005).

El dominio SEA contiene alrededor de 120 aminoácidos conservados que se ha sugerido que contiene la estructura para hacer posible el mecanismo autoproteolítico (Levitin, Stern et al. 2005). Otras proteínas que contienen este dominio, como MUC3 (Khatri, Wang et al. 2003) y IgHepta (Abe, Fukuzawa et al. 2002), también sufren la autoproteólisis. La región citoplasmática de MUC1 puede ser fosforilada iniciando varias cascadas de señalización y, con ello, modificando el comportamiento celular. El mecanismo autoproteolítico podría permitir cambios alostéricos alterando el patrón de fosforilación e, incluso, la interacción entre las subunidades α y β podría constituir un mecanismo de interacción del tipo ligando-receptor (Levitin, Stern et al. 2005).

1.5.3.1.3 Proteínas *Hedgehog*

A partir de este grupo los procesos adquieren una mayor complejidad porque el enlace éster formado acaba dando con una nueva unión ya sea a si misma o a otra molécula.

La familia de proteínas *hedgehog* realiza una función de ruta de señalización importante en el desarrollo embrionario de diferentes especies de animales (Hao, Johnsen et al. 2006). Esta ruta de señalización es importante en la guía del desarrollo embrionario de los organismos multicelulares (Lee, von Kessler et al. 1992; Echelard, Epstein et al. 1993) y regula la especificación del destino, la proliferación y la supervivencia celulares. El precursor proteico de estas proteínas contiene un dominio aminoterminal de señalización (Hh-N) y un dominio carboxiterminal responsable del mecanismo de autoproteólisis (Hh-C). La reorganización intramolecular se produce mediante una cisteína completamente conservada, que lleva a cabo el ataque nucleofílico al residuo aminoacídico anterior, formándose el enlace tioéster. Posteriormente, se produce el ataque nucleofílico del enlace tioéster por el hidroxilo 3β de una molécula de colesterol dando lugar a la liberación del domino Hh-C y a la formación de un enlace éster entre el aminoácido carboxiterminal del dominio Hh-N y la molécula de colesterol (Hall, Porter et al. 1997). El dominio Hh-N sufre diferentes modificaciones para ser dirigida a la membrana celular y, finalmente, ser liberada al medio extracelular para cumplir su función de señalización mediante la unión a receptores de las células diana.

El dominio Hh-C muestra una similitud muy alta con la región N1 de las inteinas del empalme proteico (*protein splicing*, del ingles véase apartado 1.5.3.1.4), lo que sugiere mecanismos de reacción parecidos (Koonin 1995).

1.5.3.1.4 *Splicing* proteico

El mecanismo de *splicing* proteico es el que responde a una mayor complejidad de todos los mecanismos expuestos. De forma similar al *splicing* de ARN, una parte de la cadena preprocesada, la inteína, se escinde y se produce la unión de las exteínas adyacentes para dar lugar a la proteína madura. El descubrimiento del *splicing* proteico en 1990 (Hirata, Ohsumk et al. 1990; Kane, Yamashiro et al. 1990) ha abierto un amplio abanico de posibilidades biotecnológicas en el campo de la proteómica que van desde la purificación de proteínas nativas (eliminación de los añadidos para la purificación mediante dominios inteína) hasta la producción de péptidos cíclicos. Este proceso está presente en organismos unicelulares de eucariotas, bacterias y arqueas lo que sugiere la existencia de un antiguo ancestro común (Paulus 2000). Cabe destacar la presencia de *homing* endonucleasas dentro de ciertas inteínas.

Los trabajos de Pietrokovski (Pietrokovski 1994; Pietrokovski 1998) permitieron identificar zonas conservadas, tanto en las exteínas como en las inteínas, importantes estructuralmente para el proceso de splicing proteíco (Fig. 15).

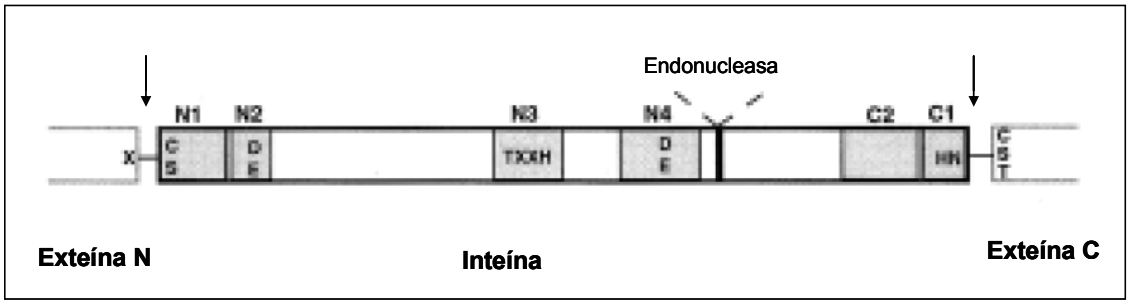


Fig. 15. Elementos conservados de una inteína. Modelo propuesto por Pietrokovski (Pietrokovski 1994; Pietrokovski 1998). Esquema modificado a partir del presentado por Paulus (Paulus 2000). En gris aparecen las zonas conservadas, las flechas indican las zonas de empalme. La zona de inserción de la endonucleasa también está representada.

El proceso de splicing proteico comprende cuatro pasos importantes. En el primer paso se forma el enlace (tio)éster en la zona de empalme de la zona aminoterminal de la inteína. Este enlace se origina por el ataque nucleofílico del residuo aminoterminal de la inteína al carboxilo del enlace peptídico que la precede. A continuación se produce un segundo enlace (tio)éster en la zona carboxiterminal de la inteína. En un segundo paso se produce una transesterificación intramolecular a causa del ataque nucleofílico del segundo enlace (tio)éster sobre el primero. A consecuencia de este paso se obtiene un polipéptido ramificado, en el que las exteínas ya están unidas por un enlace éster y la inteína está unida por su extremo carboxiterminal. En el siguiente paso, la amida β del residuo asparagina situado en la zona carboxiterminal de la inteína lleva a cabo un ataque nucleofílico sobre el grupo carbonilo del enlace peptídico produciéndose la liberación de la inteína y la ciclación de la asparagina para formar un intermediario aminosuccinimida. Finalmente, en el último paso la aminosuccinimida se hidroliza para dar lugar de nuevo a un residuo asparagina dentro de la inteína y el enlace éster entre las exteínas se transforma en un enlace amida, estabilizándose la estructura (Paulus 2000).

Es importante señalar que las proteínas que sufren *splicing* proteico no tienen una función común y entre ellas existen proteasas como ClpP (Wang and Liu 1997), polimerasas como la DNA polimerasa de *Pyrococcus sp*.

(Shao, Xu et al. 1995), subunidades de ATPasas como la subunidad VMA de *Saccharomyces cerevisiae* (Kawasaki, Makino et al. 1996; Kawasaki, Satow et al. 1997), etc.

En los últimos tiempos se han descubierto procesos de *splicing* proteico más complejos. Un ejemplo son los identificados en organismos que obtienen proteínas maduras a partir de varios genes, codificando cada gen para una exteína y parte de la inteína (Dassa, London et al. 2009). Al encontrarse las dos preproteínas se produce el *splicing*, dando lugar a una proteína madura. Otro de ellos sería el que presentan las proteínas bacterianas parecidas a inteína (*bacterial intein-like's* del inglés).

1.5.3.1.4.1 *Bacterial intein-like's* (BIL)

Este tipo de proteínas son muy parecidas a las *Hedgehog*/inteínas (Hall, Porter et al. 1997; Pietrokovski 1998) pero a diferencia de las inteínas éstas presentan motivos de secuencia únicos, no están integrados en zonas muy conservadas de proteínas esenciales y no tienen el dominio endonucleasa. A diferencia de las *hedgehog*, estas se encuentran en bacterias y no tienen los dominios específicos implicados en la actividad señalizadora. Los motivos específicos de las proteínas BIL pueden presentar residuos de ácido aspártico, ácido glutámico, asparagina, tirosina o alanina en las posiciones que intervienen en los ataques nucleofílicos de las inteínas o bien otros residuos en la asparagina que interviene en la formación del intermediario succinimida (Amitai, Belenkiy et al. 2003). Este hecho da a lugar *splicings* alternativos dependiendo de los aminoácidos y el paso implicados, que puede conducir a varios mecanismos de maduración diferentes (Fig. 16).

El hecho de encontrar tantas similitudes estructurales en los procesos de autoproteólisis entre las proteínas hedgehog y las inteínas ha provocado que se las acabe agrupando dentro de la misma familia, las proteínas HINT (**h**og y **int**eínas).

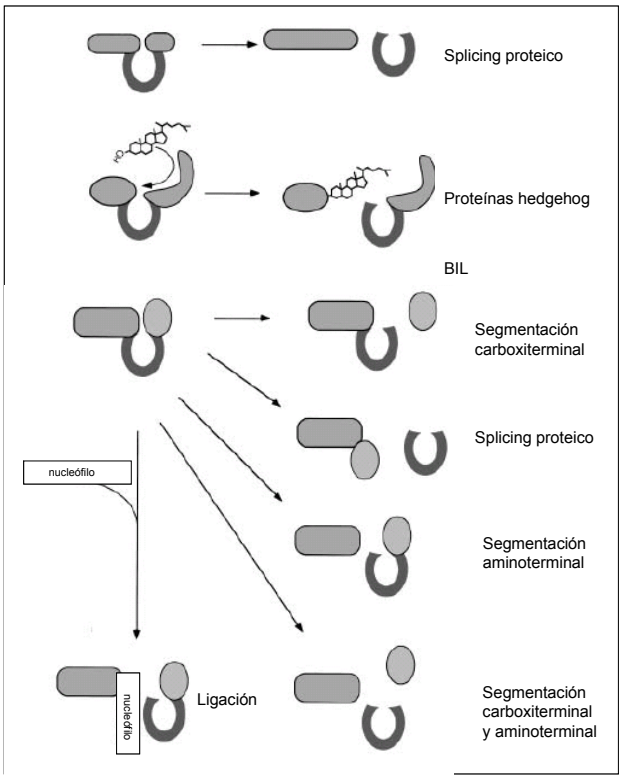


Fig. 16. Modificación del modelo de autoprocesamiento de las proteínas Hint presentado por Amitai (Amitai, Belenkiy et al. 2003).

Es interesante destacar el proceso evolutivo que han seguido las proteínas Hint. Ambos tipos de mecanismos, *splicing* proteico y proteínas *hedgehog*, tienen un origen común indudable pero en términos de función han seguido una curiosa dicotomía. El *splicing* proteico está presente en organismos unicelulares y tiene su centro de actuación como elemento genético parasítico debido a la presencia de las *homing* endonucleasas. Estas *homing* endonucleasas podrían tener la capacidad de esparcir las inteínas mediante transferencia genética horizontal incluso entre especies. En cambio, las proteínas *hedgehog* están presentes en organismos multicelulares, específicamente en animales, en los que desarrollan una función de señalización (Paulus 2000).

1.5.3.2 Mecanismos relacionados con el enlace peptídico Asx-Aa

De manera similar al capítulo anterior, dónde se describían diferentes procesos con un nexo común de unión que partía de la formación de un enlace (tio)éster, en este grupo se describen los procesos que parten de la formación de intermediarios cíclicos en el enlace entre un Asp o una Asn y otro aminoácido (Asx-aa). Estos intermediarios son los anhídridos cíclicos y las imidas cíclicas, y, aunque la relación entre los anhídridos cíclicos y la segmentación del enlace peptídico es directa esta relación es más difusa entre las imidas cíclicas y la segmentación del enlace peptídico. Aunque el mecanismo difiera entre ambas, se ha descrito que la energía libre de activación para estos dos intermediarios es similar (<1kcal mol-1) (Catak, Monard et al. 2008). Cabe destacar que se han descrito casos en que los residuos implicados en la producción de estos intermediarios son el Glu y la Gln.

1.5.3.2.1 Formación de anhídridos cíclicos

En la literatura se ha descrito el proceso de hidrólisis del enlace peptídico en residuos Asp o Asn expuestos al solvente (Shrier, Kenley et al. 1993; Shahrokh, Eberlein et al. 1994; Brange, Langkjaer et al. 1992; Li, Fort et al. 2009) en péptidos y proteínas. Aunque varios modelos se han presentado para el proceso de segmentación de este tipo de enlace, el trabajo de Oliyai et al. (Oliyai, Borchardt 1993) mostró que la explicación más plausible envuelve a un ataque nucleofílico intramolecular (catalizado por ácido) del carboxilo de la cadena lateral del Asp al carbonilo del enlace peptídico con el residuo n+1. Este ataque nucleofílico produce un intermediario anhídrido cíclico que puede hidrolizarse (mediante otro ataque nucleofílico) para dar los productos segmentados (Fig 17). Se ha descrito que el residuo adyacente al Asp afecta fuertemente el ratio de la hidrólisis (Li, Fort et al. 2009) siendo la Pro la que mayor incrementa la labilidad de este enlace.

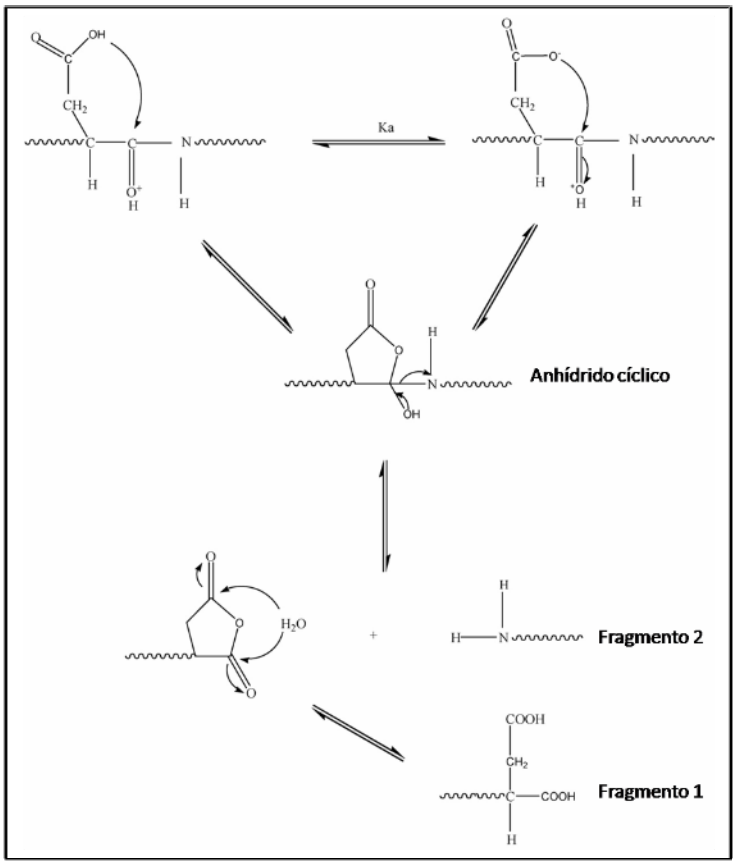


Fig.17 Modificación del modelo propuesto por Oliyai (Oliyai, Borchardt 1993) de la segmentación del enlace peptídico entre un Asp y otro residuo a través de la formación de un intermediario anhídrico cíclico.

1.5.3.2.2 Formación de succinimidas (imidas cíclicas)

La formación de succinimidas se ha incorporado en esta introducción en el capítulo de mecanismos autocatalíticos intramoleculares ya que, como hemos visto al describir el *splicing* proteico, puede estar envuelta en la segmentación del enlace peptídico. Como hemos comentado anteriormente, se ha propuesto que la amida β puede atacar al carbonilo del enlace peptídico, de manera similar a la formación del intermediario anhídrido cíclico, y dar lugar a la segmentación del enlace (Paulus 2000). Aún así, la formación de succinimidas se incluye en un proceso más complejo, la desamidación y la formación de ácido isoaspártico en las proteínas.

1.5.3.2.2.1 La desamidación y la formación de ácido isoaspártico en las proteínas

En condiciones favorables, se puede producir una reordenación de los residuos de ácido aspártico y asparagina a succinimidas (Kwong and Harris 1994) mediante la desamidación de la asparagina o la deshidratación del aspártico. La formación de succinimidas está favorecida cuando pequeños residuos hidrofílicos (glicina, serina o treonina) se encuentran adyacentes al aspártico/asparagina (Tyler-Cross and Schirch 1991; Wright 1991) y se presenta en zonas flexibles de la proteína (Kossiakoff 1988). Normalmente la formación de succinimidas se puede deducir a partir de la presencia de residuos de ácido isoaspártico (George-Nascimento, Lowenson et al. 1990; Violand, Schlittler et al. 1990; Di Donato, Ciardiello et al. 1993).

La succinimida se forma mediante el ataque nucleofílico de la amida α del enlace peptídico al carboxilo de la cadena lateral del aspártico/asparagina. Una vez formada la succinimida ésta se hidroliza produciendo un residuo de isoaspártico en una proporción de 70-85% y un residuo aspártico en una proporción de 15-30% (Reissner and Aswad 2003) (Fig 18.A). El isoaspártico se produce por la formación del nuevo enlace peptídico con el carboxilo β de la succinimida, dejando el carboxilo α al descubierto. El isoaspártico se produce en mayor cantidad por la asimetría de la estructura de la succinimida (Reissner and Aswad 2003). Estos residuos isoaspárticos son selectivamente metilados en el carboxilo α por la proteína carboxil metiltransferasa (PIMT E.C.2.1.1.77) (Diliberto and Axelrod 1976; O'Connor and Clarke 1985), lo que da lugar a una nueva ciclación produciéndose la succinimida (Fig 18.B). La especificidad de esta enzima sugirió la existencia de un proceso de reparación o extracción del isoaspártico (Johnson and Murray, et al. 1987).

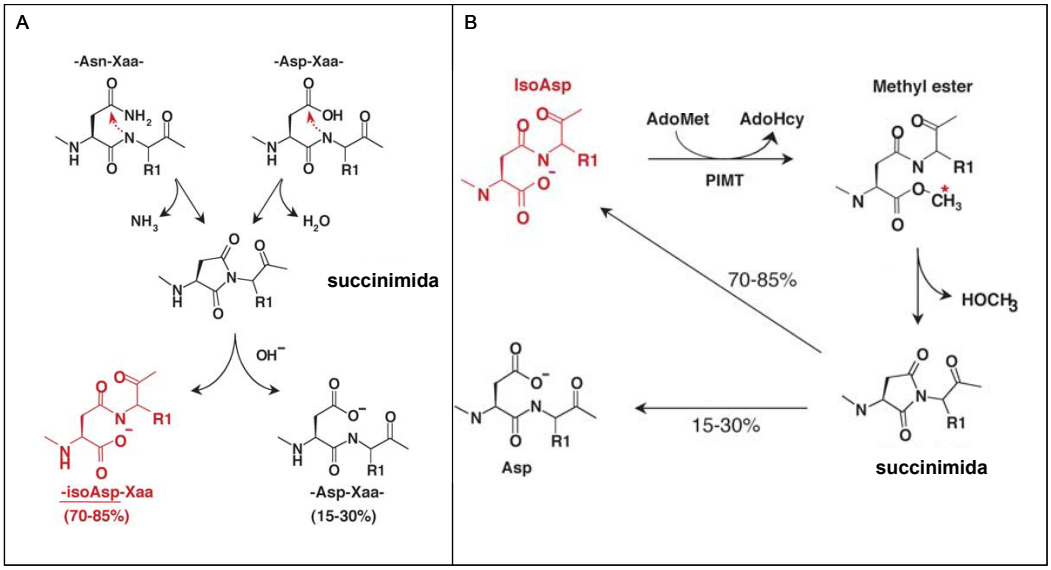


Fig. 18. A. Esquema del modelo de desamidación y formación del isoaspártico. B. Esquema del modelo de reparación del ácido isoaspártico por la enzima PIMT. Ambos esquemas son modificaciones del modelo presentado por Reissner y colaboradores (Reissner 2003).

El proceso de desamidación y de formación de residuos del ácido isoaspártico, así como su reciclaje por la proteína PIMT se ha asociado con un mecanismo relacionado con la edad de la proteína (reloj interno de la proteína (Robinson and Rudd 1974)) y, por lo tanto con la hipótesis de daño y reparación de la proteína. Esta hipótesis se ha corroborado con la calmodulina, ya que al reparar la formación del residuo isoaspártico se activa de 18 a 68% la actividad de la calcio/calmodulina proteina quinasa (Johnson, Langmack et al. 1987). Se ha caracterizado también que la acumulación de proteínas debido a la formación del isoaspártico como en la histona H2B o la sinapsina contribuye a enfermedades autoimmunes como el lupus eritematoso (Mamula, Gee et al. 1999). Paralelamente, también se ha descrito que existe una alta proporción de residuos isoaspártico en los péptidos Aβ acumulados en cerebros afectados de Alzheimer y que ratones deficientes en PIMT en el cerebro sufren cambios neurodegenerativos progresivos (Shimizu, Matsuoka et al. 2005).

A pesar de todos los ejemplos que sustentan esta hipótesis, en los últimos tiempos está apareciendo la duda de si esta modificación post-traduccional pueda desempeñar también otras importantes funciones. Una pista que indica otras posibles funciones es el hecho de que el proceso de la reparación/reciclaje es incompleto. En el caso del Asp la proteína PIMT restablece tanto el esqueleto peptídico normal como la cadena lateral del aminoácido,

pero en el caso de la desamidación de la Asn, aunque se restaura el esqueleto peptídico, la cadena lateral se transforma en un ácido aspártico (fig 17.B). Esto indica que podría ser un mecanismo de introducción de heterogeneidad en la especie proteica para modificar su comportamiento y/o su distribución celular (Pepperkok, Hotz-Wagenblatt et al. 2000). Un ejemplo de ésto último es la mayor concentración en el núcleo (1,5 a 2,3 veces) de la PKA-C nativa en comparación con la PKA-C desaminada en la Asn[2] (Pepperkok, Hotz-Wagenblatt et al. 2000).

Se ha sugerido también que PIMT (por la restitución del residuo Asp) tendría la función de asegurar que proteasas específicas pudieran degradar completamente estas proteínas. Esta hipótesis se sustenta en el hecho de que los residuos de ácido isoaspártico interfieren en el reconocimiento de las proteasas (Dorer, Haley et al. 1968; Johnson, Murray et al. 1987) y también en los altos niveles de isoaspárticos encontrados en enfermedades en las que hay una acumulación de proteínas antes citadas. Incluso se contempla la posibilidad de que la formación de isoaspártico/aspártico constituya un mecanismo reversible de regulación de la función de estas enzimas.

OBJETIVOS

2. OBJETIVOS

1. Estudio mediante mecánica molecular de la 4-difosfocitidil-2C-metil-D-eritritol quinasa (CMK) de *Escherichia coli* a partir de la estructura cristalina obtenida por difracción de rayos X ya descrita (PDB 1oj4). Obtención de trayectorias de dinámica molecular para el estudio de las interacciones proteína-proteína de la formación del complejo homodimérico.

2. Realización de hipótesis farmacofóricas para la búsqueda de compuestos que puedan inhibir la formación del dímero. Búsqueda virtual en bases de compuestos mediante las hipótesis farmacofóricas realizadas.

3. Caracterización mediante métodos empíricos de la unión de los compuestos obtenidos con la CMK. Estudio del efecto de los compuestos obtenidos sobre la actividad enzimática de la CMK.

4. Caracterización del marco abierto de lectura (ORF) TM_1770 como posible homólogo de la enzima 1-desoxi-D-xilulosa 5-fosfato sintasa (DXS) en *Thermotoga maritima*. Caracterización bioquímica de la enzima.

5. Estudio del mecanismo proteolítico observado en la DXS de *Escherichia coli, Thermotoga maritima* y *Arabidopsis thaliana.*

INFORME DEL DIRECTOR DEL FACTOR DE IMPACTO DE LOS ARTICULOS PUBLICADOS

Santiago Imperial Ródenas, Profesor Titular del Departament de Bioquímica i Biología Molecular de la Universitat de Barcelona como director de la tesis titulada:

Molecular studies of two methylerythritol 4-phosphate pathway enzymes of isoprenoid biosynthesis: the 4-diphosphocytidyl-2C-methyl-D-erythritol kinase and the 1-deoxy-D-xylulose 5-phosphate synthase (Estudios moleculares de dos enzimas de la ruta del metileritritol 4-fosfato de biosíntesis de isoprenoides: la 4-difosfocitidil-2C-metil-D-eritritol quinasa y la 1-desoxi-D-xilulosa 5-fosfato sintasa)

HACE CONSTAR QUE:

Esta Tesis Doctoral se presenta como "Compendio de Publicaciones" vertebrándose en cinco Capítulos de los cuales el doctorando es primer autor de todos ellos. El Capítulo I ha sido publicado en la revista Journal of Molecular Modeling (2,018 de índice de impacto, posición 19 de 94 de su categoría: computer science, interdisciplinary applications) y corresponde a un estudio realizado mediante mecánica molecular de la 4-difosfocitidil-2C-metil-D-eritritol quinasa (CMK) *de Escherichia coli* en el que se han obtenido trayectorias de dinámica molecular para el estudio de las interacciones proteína-proteína de la formación del complejo homodimérico de la enzima. El capitulo II corresponde a un artículo aceptado con revisiones en Journal of Molecular Recognition (3,160 de índice de impacto, posición 19/70 categoría: biophysics). En él se describe la realización de hipótesis farmacofóricas para la búsqueda de compuestos que puedan inhibir la formación del dímero de la enzima CMK. En el capítulo III se estudia el efecto de los compuestos encontrados con capacidad de unión a la CMK sobre la actividad enzimática de ésta. También se realizan estudios de oligomerización de la CMK y se compara con otros miembros de la familia de quinasas GHMP. Está previsto que este trabajo sea enviado para su publicación a la revista Biochimica et Biophysica Acta, General Subjects con un índice de impacto de 2,713 (posición 131/276 categoría: biochemistry and molecular biology). Los resultados presentados en el capítulo IV se centran en la caracterización del marco abierto de lectura TM_1770 de *Thermotoga maritima* como homólogo de la enzima desoxixilulosa 5-fosfato sintasa (DXS) de *E. coli*. Se describe el carácter termofílico y algunas constantes bioquímicas de la proteína codificada por TM_1770. Este trabajo se enviará para su publicación a la revista Extremophiles, con un índice de impacto de 1,782 (posición 59/91, categoría: microbiology). Por último los resultados resumidos en el capítulo V, describen el mecanismo de autoproteolisis que experimenta la enzima DXS durante su purificación que podrían estar asociados a mecanismos post-traduccionales de la proteína. Está previsto enviar el manuscrito a la revista FEBS Journal con un índice de impacto de 3,139 (posición 105/276, categoría: biochemistry and molecular biology)

Por lo que respecta a la contribución del doctorando a cada uno de los capítulos, Víctor Giménez-Oya fue el encargado de llevar a cabo el grueso de la investigación y el desarrollo de los experimentos, exceptuando los correspondientes al capítulo I en los que comparte la posición de primer autor. Cabe destacar que el trabajo de coautoría no se ha presentado en ninguna otra tesis doctoral.

Dr. Santiago IMPERIAL RÓDENAS

Departament de Bioquímica i Biología Molecular
Universitat de Barcelona

RESUMEN GLOBAL: RESULTADOS, DISCUSIÓN Y CONCLUSIONES

3.1 RESULTADOS

La presente Tesis se ha planteado de una forma multidisciplinar. El planteamiento y los resultados ahondan tanto en campos con posibles aplicaciones finales terapéuticas como en conocimientos de ciencia básica. Abarcando para ello disciplinas que van desde la química computacional hasta la bioquímica de proteínas asistida por aplicaciones biotecnológicas. Todo ello se ha aplicado a una ruta muy interesante, la ruta del MEP de síntesis de precursores de compuestos isoprenoides. Destacable es, al fin, la utilización de modelos *in silico*, modelos *in vitro* y modelos *in vivo*.

En el primer capítulo se describe el estudio realizado mediante mecánica molecular de la CMK de *Escherichia coli*. La primera parte del estudio describe las interacciones proteína-proteína del complejo homodimérico y la importancia del solvente en su formación. En la segunda parte del estudio se describe la hipótesis farmacofórica diseñada a partir del análisis de las interacciones con el objetivo de encontrar moléculas capaces de interrumpir estas interacciones. El capítulo acaba con la caracterización de la unión de un compuesto a la CMK según la técnica de "disminución de la intensidad" con láser de desorción/ionización asistido por matriz-tiempo de vuelo (Intensity fading MALDITOF-MS).

En el segundo capítulo se describe la incorporación de varios puntos farmacofóricos adicionales más al farmacóforo previamente diseñado. La importancia del solvente en la formación del dímero, descrita en el primer capítulo, condujo a intentar mejorar la hipótesis farmacofórica. Adicionalmente, se diseñó un péptido cíclico que también imita la zona de interacción. En el capítulo se describe además la caracterización empírica mediante resonancia del plasmón en superfície de la unión a CMK de un compuesto obtenido a partir del nuevo farmacóforo y también del péptido cíclico. Finalmente, se presentan modelos hipotéticos para describir las uniones observadas mediante dinámica molecular.

En el tercer capítulo se estudia el efecto de los compuestos encontrados con capacidad de unión a la CMK sobre la actividad enzimática de ésta. También se analiza el grado de oligomerización de la CMK y se compara con el de otros miembros de la familia de quinasas GHMP.

El cuarto capítulo se centra en la caracterización del marco abierto de lectura TM_1770 de *Thermotoga maritima* como homólogo de la DXS. En este capítulo se describe el carácter termofílico y algunas de las propiedades de la proteína codificada por TM_1770.

Finalmente, en el último capítulo se describe el proceso proteolítico observado durante la purificación de las enzimas DXS de *Escherichia coli, Arabidopsis thaliana y Thermotoga maritima*. Los resultados sugieren que la proteína sufre mecanismos de intramoleculares post-traduccionales y se intentan asociar al proceso de proteólisis.

A continuación, se exponen los resultados obtenidos en los diferentes capítulos para, posteriormente, proceder a una discusión de los mismos y a enumerar las conclusiones obtenidas.

3.1.1 Imitando interacciones directas proteína-proteína e interacciones mediadas por solvente en el homodímero CDP-ME quinasa: Una aproximación de búsqueda virtual dirigida a través de farmacóforo.

Para realizar el estudio por mecánica molecular de la CMK se partió de la estructura tridimensional 1oj4 (Miallau, Alphey et al. 2003) obtenida a partir de la base de datos Protein Data Bank (http://www.rcsb.org/pdb/home/home.do). La estructura cristalina presenta una disposición homodimérica con aparente posición invertida de los monómeros. La estabilización del dímero se produce mediante la interacción de dos pequeñas zonas de la proteína. En la zona de interacción A se encuentran las interacciones intermoleculares del enlace iónico entre la Arg^{21} y el Asp^{80} y el puente de hidrógeno entre el carbonilo de la Ala^{22} y la amida de la Gly^{87}. En la zona B se encuentran las mismas interacciones, aunque la segunda de ellas se presenta a través de la mediación por solvente. Por otro lado, también se observan un gran número de interacciones mediadas por solvente entre los dos monómeros de la estructura cristalina.

A la estructura, una vez incluidos los hidrógenos, se le minimizó la energía simulando la constante dieléctrica del agua. A continuación se le añadieron cajas de solvatación para simular el solvente en el estado natural de la proteína y se procedió a realizar de nuevo la minimización de la energía. A partir de aquí, se fue aumentando la temperatura del sistema hasta alcanzar los 300ºK con una serie de restricciones al volumen y a la presión para luego ir paulatinamente reduciendo dichas restricciones. Una vez el sistema hubo convergido en un mínimo de energía se llevó a cabo el estudio estructural, para ello se realizaron 1000 ps de trayectoria.

Los estudios de la trayectoria mostraron que, a excepción del enlace iónico entre la Arg^{21} y el Asp^{80} de la zona de interacción A, las demás interacciones electrostáticas que estabilizan el dímero se muestran como puentes de hidrógeno mediados por solvente. Debido a la importancia del solvente, se llevó a cabo la caracterización de las aguas que intervenían en las interacciones durante 200 ps. Destacables son las interacciones que se observan de los residuos Ala^{22} y Asp^{23} de la zona de interacción A. En el primer caso una molécula de agua actúa como centro de una red de puentes de hidrógeno entre el carbonilo de la Ala^{22} de un monómero y el esqueleto de los residuos Gly^{49}, Pro^{85}, Gly^{87} y Ser^{88} del otro monómero. Aunque esta red de interacciones se mantiene estable a lo largo de la trayectoria, la molécula de agua que actúa como centro va variando. En el caso del Asp^{23} éste establece dos interacciones mediadas por solvente. La cadena lateral establece la interacción con la cadena lateral de la Lys^{76} del otro monómero, mientras que el carbonilo del esqueleto establece una red con el carbonilo de la Gly^{49} y la cadena lateral de la Arg^{72}.

El estudio de las moléculas de agua mostró su gran importancia en la superficie de reconocimiento entre los dos monómeros. La superficie de interacción entre ellos aumenta un 19% (de 355 $Å^2$ a 422 $Å^2$) cuando se tiene en cuenta la superficie de las moléculas de agua implicadas en las interacciones mediadas por solvente.

El cálculo de las energías de van der Waals permitió inferir la participación de dos zonas de residuos en la interacción de carácter hidrofóbica. Estas zonas interaccionan con las respectivas zonas del otro monómero. Al cálcular la energía electroestática entre los monómeros se obtuvo un valor de -27 kcal mol^{-1}. Este valor tan pequeño podría venir dado por la carga que tiene la CMK ya que ésta presenta en cada monómero 36 residuos con carga negativa y 26 residuos con carga positiva. Cuando se incluyó el solvente en el cálculo, el valor

obtenido aumentó hasta -5.537 kcal mol^{-1}. Según este resultado el solvente podría estar compensando las cargas en exceso que tiene la proteína, tal como se observa al calcular las energías electroestáticas de cada uno de los residuos por separado y al observar que las repulsiones de los aminoácidos cargados negativamente disminuyen al incluir el solvente en el cálculo.

Dada la poca superficie que formaba las zonas de interacción del dímero y el hecho de que las principales interacciones se centraban únicamente en tres residuos, se decidió realizar la hipótesis farmacofórica para imitarlos. Esta hipótesis farmacofórica se realizó con el objetivo de encontrar pequeñas moléculas capaces de impedir la formación del dímero. El farmacóforo contiene 5 puntos farmacofóricos de los cuales tres tienen carácter hidrofóbico, uno corresponde a un dador de protones (imitando la cadena lateral de la Arg^{21}) y el último un aceptor de protones (imitando la molécula de agua de la interacción mediada por solvente entre el Asp^{23} y la Lys^{76}). Las coordenadas del farmacóforo se obtuvieron a partir de los máximos y los mínimos de los grupos a imitar obtenidos durante la trayectoria de la dinámica molecular.

Con el programa Catalyst™ se realizó una búsqueda de moléculas que cumplieran las restricciones de la hipótesis farmacofórica en diferentes bases de datos de compuestos. Como resultado se obtuvieron 214 conformaciones de moléculas de ACD, 512 de SPECS y 835 de ChemDiv que cumplían las restricciones de carga y estéricas definidas en el farmacóforo. Estas conformaciones se anclaron a la proteína y se descartaron aquellas que pudieran presentar impedimentos estéricos con la CMK. Las conformaciones positivas se ordenaron mediante el programa X-Score y finalmente, tras una selección de los compuestos por el compromiso entre los valores del X-score y parámetros de visualización, se compraron 11 compuestos (Anexo 1).

Para comprobar la interacción de las moléculas seleccionadas con la CMK, y así corroborar el modelo propuesto, se realizaron estudios de "disminución de la intensidad" por MALDI-TOF MS. En este capítulo se describe la interacción de forma cualitativa de un compuesto (compuesto 1) con la CMK según este método. Al realizar los tiempos de vuelo del compuesto 1 se obtuvieron dos picos de m/z, 508 y 510. El pico de 508 correspondía al peso molecular de la molécula. Al añadir la CMK en concentraciones crecientes a la mezcla, se observó que los picos correspondientes al compuesto 1 iban disminuyendo en intensidad, lo que se puede asociar con la formación del complejo compuesto1-CMK.

3.1.2 Diseño de una molécula pequeña y un péptido cíclico natural como ligandos de la CDP-metileritritol quinasa imitando interacciones proteína-proteína e interacciones mediadas por solvente.

El agua tiene un importante papel en la estabilidad de los complejos biológicos y este hecho está aumentando la importancia de su papel en el diseño racional de fármacos (Lloyd, Garcia-Sosa et al. 2004; Leroux, Gresh et al. 2006). En consecuencia, se decidió a incrementar los puntos farmacofóricos de la hipótesis farmacofórica imitando más interacciones mediadas por solvente. Gracias al estudio de las moléculas de agua importantes para las interacciones entre los monómeros de la CMK realizado en el capítulo anterior, se identificaron dos nuevas posiciones. Estas dos posiciones cumplían los requisitos de mediar una interacción entre los monómeros y de estar localizadas próximas a la zona de interacción imitada en el farmacóforo del capítulo 1. La primera

molécula de agua era la que formaba la red de puentes de hidrógeno entre la Ala^{22} de un monómero y los residuos Gly^{49}, Pro^{85}, Gly^{87} y Ser^{88} del otro monómero de la CMK. La segunda molécula de agua era la que formaba interacciones entre el carbonilo del Asp^{23} y los residuos Gly^{49} y Arg^{72} del otro monómero.

Se implementaron estos puntos al farmacóforo descrito en el capítulo anterior y se volvió a realizar la búsqueda en las bases de compuestos con el programa Catalyst™ (Accelrys). Se obtuvieron 639 conformaciones que cumplían los requisitos impuestos por el farmacóforo de entre las bases de datos ACD, Mini Maybridge, ChemDiv databases, Maybridge 2004, SPECS, Nat Diverse, IBS 2004, NCI, Aurora y Derwent.

Las diferentes conformaciones se anclaron a la CMK, como en el capítulo 1, descartando aquellas que pudieran presentar impedimentos estéricos. Las que superaron este proceso se ordenaron con el programa X-Score y se compraron 4 compuestos (anexo 1).

Los residuos implicados en la interacción entre las subunidades de la CMK utilizados en el farmacóforo se localizaban en un pequeño rizo situado entre dos láminas β que se inserta en un pequeño hueco entre una helice α y dos rizos del otro monómero de la CMK. Este rizo está conservado entre las CMK de diferentes organismos (Wada, Kuzuyama et al. 2003) con una secuencia consenso de aminoácidos Arg-X-Asp-Gly-Tyr-His-X-Leu-X-Thr-X-Phe. Con la idea de poder determinar la importancia de este rizo, el cual ya estábamos intentando imitar mediante la búsqueda de moléculas pequeñas, se sintetizó un péptido cíclico con la secuencia de aminoácidos correspondiente al rizo de la CMK de *Escherichia coli*. La ciclación de péptidos puede producir un incremento de los valores de afinidad debido al decremento de la perdida de entropía en la unión respecto a los péptidos lineales. Esta ganancia de afinidad ya se ha descrito anteriormente (Bach II, Eyermann et al. 1994; Morgan, Holland et al. 1994; Zang, Yu et al. 1998; Dumez, Snaith et al. 2002; Andrews, McInnes et al. 2004; Nam, Ye et al. 2004; Chang and Chu 2005).

En este capítulo se estudia el modo de unión de una molécula pequeña (derivado de 7H-furo[3,2-g]chromo-7-ona) y el péptido cíclico a la CMK utilizando el protocolo MMGBSA (Kollman, Massova et al. 2000). En la ordenación de las moléculas pequeñas nos encontramos con que esta molécula presentaba valores buenos (relativos a otras moléculas) con dos conformaciones diferentes para la desviación de la raíz cuadrada de las medias (RMSD) y para el X-Score. Por ello, se realizaron los estudios estructurales para las dos conformaciones diferentes. Se realizó una trayectoria de 2ns de dinámica molecular para poder caracterizar las uniones de la molécula pequeña (2 trayectorias) y del péptido cíclico a la CMK.

Se obtuvo un valor de energía libre de unión del complejo de -20,61 kcal mol^{-1} y -14,94 kcal mol^{-1} para las conformaciones de la molécula pequeña con bajo RMSD y alto valor de X-Score, respectivamente. La energía libre de unión obtenida para el complejo péptido cíclico/CMK fue de -13,21 kcal mol^{-1}. El protocolo MMGBSA también permite llevar a cabo el estudio por residuo, por lo que en este capítulo se incluye un modelo de interacción para cada caso, describiendo los residuos implicados en las principales interacciones hidrofóbicas y electroestáticas. Para poder predecir la capacidad de los compuestos para competir por la zona de interacción entre los monómeros se calculó la energía libre de unión del dímero, y se obtuvo un valor de -11,1 kcal mol^{-1}. Ambos compuestos presentan una estimación de la energía libre de unión más favorable que el propio complejo homodimérico aunque cabe destacar que estos valores són una predicción en la que no se estima la entropía.

Para poder corroborar los modelos obtenidos se caracterizó la unión de estos compuestos a la CMK mediante la técnica de resonancia del plasmón en superficie. En una primera serie de ensayos se fijó la CMK al chip y se probaron las moléculas como ligandos. A partir de esta técnica se obtuvo un valor de K_D $2{,}431 \times 10^{-4}$ M y una

energía libre de unión de -4,93 kcal mol^{-1} para el péptido cíclico, y un valor de K_D 2,121x10^{-4} M y una energía libre de unión de -5,01 kcal mol^{-1} para la molécula pequeña. Una forma de corroborar el modelo de unión predicho es mediante la mutación de los residuos implicados, con este fin se prepararon mutantes de la CMK (Asp80Ala, Lys76Ala), de los cuales sólo se pudo purificar el mutante Asp80Ala. Esta forma mutada se fijó al chip y se analizó la interacción con el péptido cíclico (el residuo Asp^{80} no aparecía como importante en el modelo molécula pequeña/CMK). El ensayo de resonancia del plasmón en superficie mostró que el péptido cíclico no interaccionaba con el mutante Asp80Ala. Adicionalmente, se fijó el péptido cíclico en el chip y se utilizó la CMK como ligando, obtuviendose unos valores de K_D 1,14x10^{-5} M y una energía de libre de unión de -6,74 kcal mol^{-1}.

3.1.3 Análisis del estado oligomérico de la CDP-metileritritol quinasa: la formación del homodímero no es necesaria para la actividad enzimática.

La CMK de *Escherichia coli* había sido caracterizada como una enzima homodimérica mediante experimentos de espectrometría de masas de láser de desorción/ionización asistida por matriz-tiempo de vuelo (MALDI-TOF MS), de cromatografía de exclusión molecular (Miallau, Alphey et al. 2003) y de ultracentrifugación diferencial (Gabrielsen, Bond et al. 2004). A pesar de ello, debido a controversias observadas en la oligomerización se procedió, pues, a un análisis detallado de la oligomerización de la CMK por diferentes técnicas. Nuestros resultados obtenidos por experimentos de cromatografía de exclusión molecular mostraron que las preparaciones de CMK de *E.coli* contenían una mezcla de especies monoméricas y diméricas, siendo la especie monomérica la predominante. Cuando las fracciones cromatográficas separadas se guardaban a 4ºC y se volvían a analizar por columna de exclusión molecular se volvían a obtener picos bien separados que correspondían a los pesos moleculares de las fracciones monoméricas y diméricas. Estos resultados sugerían que con el transcurso del tiempo existía un proceso dinámico de interconversión de las distintas formas de oligomerización.

Al analizar el estado de oligomerización de las fracciones de CMK de *E.coli* mediante electroforesis nativa en geles de diferente porcentaje de poliacrilamida y representar la migración relativa respecto al porcentaje de acrilamida se obtuvo una recta con un valor de la pendiente de 5,42. Este valor, al interpolarlo en la recta patrón obtenida de los valores correspondientes a otras proteínas de pesos moleculares (y grado de oligomerización) conocidos correspondía con un tamaño de 31kDa, lo que indicaba que, principalmente, la CMK se encontraba en estado monomérico.

Para comprobar el equilibrio dinámico se realizaron experimentos de unión química cruzada ("*cross-linking*") con formaldehído en la CMK de *E.coli*. Se observó el aumento en la intensidad de una banda de 70 kDa de peso molecular aparente, cuyo tamaño molecular es compatible con el tamaño del dímero. La intensidad de dicha banda no aumentaba a una velocidad muy elevada a lo largo del tiempo del experimento por lo que, o bien la eficiencia de unión química cruzada no era muy buena o el equilibrio dinámico no tiene una velocidad muy elevada. Aunque la aparición de la banda correspondiente al dímero a lo largo del tiempo experimental sí que corroboraba la hipótesis del equilibrio dinámico.

Los experimentos de dispersión dinámica de la luz (DLS) mostraron la CMK de *E. coli* con una oligomerización monomérica aunque, sin embargo, las fracciones purificadas utilizadas presentaban una gran polidispersidad.

Esta polidispersidad se transformaba en monodispersidad si se realizaban los experimentos de DLS con las fracciones correspondientes a la especie monomérica obtenidas mediante cromatografía de exclusión molecular. La polidispersidad, por tanto, podría venir dada por la presencia de otras especies oligoméricas.
Paralelamente, se comprobó que las moléculas obtenidas por diseño racional, descritas en el primer y segundo capítulo, no ejercían ningún efecto sobre la actividad enzimática de la CMK. Además, los valores de actividad enzimática de las fracciones correspondientes a la especie monomérica obtenidas mediante cromatografía de exclusión molecular mostraban valores de actividad similares a los descritos previamente (Bernal, Mendez et al. 2005) para las preparaciones purificadas de la enzima que contenían una mezcla de diferentes especies oligoméricas.

3.1.4 Clonación molecular y caracterización de la 1-desoxi-D-xilulosa 5-fosfato de la bacteria termofílica *Thermotoga maritima*

La 1-desoxi-D-xilulosa 5-fosfato sintasa (DXS) es una enzima crucial en varias rutas metabólicas. Se han caracterizado homólogos de la DXS en gran cantidad de organismos pero, hasta la fecha, no se había caracterizado ninguno de un organismo hipertermófilo. Con este objetivo se escogió el marco abierto de lectura (ORF) TM_1770 de *Thermotoga maritima*, el cual se había descrito como una DXS potencial (Nelson, Clayton et al. 1999) durante el proceso de secuenciación del genoma de este organismo.
El ORF TM_1770 se clonó en varios vectores de expresión compatibles para la expresión y la purificación en *Escherichia coli*. Para comprobar que TM_1770 presentaba actividad enzimática DXS se transformó la cepa de *Escherichia coli dxs*-deficiente *EcAB4-2* con las construcciones obtenidas. La cepa *EcAB4-2* es capaz de crecer en un medio con mevalonato (MVA) debido a que a su genoma se le ha insertado el operón mevalonato y es capaz de sintetizar IPP y DMAPP a partir de MVA. Sin la presencia de MVA esta cepa es incapaz de crecer debido que el gen *dxs* está disruptado por la inserción del gen *cloranfenicol acetil transferasa*. Se observó que la cepa *EcAB4-2* transformada con las construcciones de TM_1770 era capaz de crecer en ausencia de MVA, lo que indicaba que este marco abierto de lectura codificaba para una proteína con actividad enzimática DXS.
Para poder purificar la proteína en cantidades suficientes para ser caracterizada se utilizó la cepa de expresión BL21 Rosetta2 que contiene un plásmido que codifica para 7 RNAs de transferencia que se encuentran en baja frecuencia en *Escherichia coli*. La secuencia de DXS de *Thermotoga maritima* contenía un 24% de estos codones poco frequentes debido, muy probablemente, a que el genoma de *Thermotoga* contiene un número elevado de genes de arqueobacterias y, para poder producirse la transferencia genética horizontal, genomas y genes insertores se tienen que parecer. Esto corroboraría la diferenciación de *Thermotoga* de las bacterias Gram- así como su situación filogenética como puente entre los dos reinos. La enzima purificada se caracterizó funcionalmente presentando un óptimo de actividad a una temperatura de 70 ºC, con una energía de activación de 34,3 KJ mol^{-1} K^{-1}. La enzima presentó una cinética típica de Michaelis-Mentem para el piruvato presentando una K_M de 75,7 μM, una V_{max} de 4,55 μmol min^{-1}, una K_{cat} de 5,16 s^{-1} y una K_{cat}/K_M of 6,8 x 10e^4 $M^{-1}s^{-1}$, a 70 ºC y un pH 7,8.
Los residuos His48 de la DXS de *Escherichia coli* (Querol, Rodriguez-Concepcion et al. 2001) y His49 de *Agrobacterium tumesfaciens* (Lee, Oh et al. 2007) se han descrito como cruciales para la actividad enzimática de la DXS. Al comparar estas secuencias con la secuencia de DXS de *Thermotoga maritima* se observó que el

residuo His estaba conservado y se presentaba en la posición 35. Se preparó el mutante TmDXS His35Ala y se observó que la cepa *EcAB4-2* transformada con el mutante era incapaz de superar el bloqueo de la *dxs*. Este hecho corrobora el papel crucial de esta His conservada en DXS y sugiere que la His35 de *Thermotoga maritima* llevaría a cabo una función similar al modelo presentado para la His48 de *Escherichia coli* y la His49 de *Agrobacterium tumesfaciens*.

3.1.5 Procesamiento autoproteolítico de la 1-desoxi-D-xilulosa 5-fosfato sintasa

La DXS, como enzima crucial en varias rutas enzimáticas, está sujeta a una estricta regulación siendo posible que mecanismos post-traduccionales puedan estar asociados a esta regulación. Hasta la fecha, se ha sugerido la proteólisis mediada por proteasa en los homólogos de *Escherichia coli* (Xiang, Usunow et al. 2007) y *Arabidopsis thaliana* (Flores-Perez, Sauret-Gueto et al. 2008). Se describió que la resolución de la estructura de la DXS de *Escherichia coli* (Xiang, Usunow et al. 2007) se resolvió gracias a la contaminación por un hongo durante las cribas de las condiciones de cristalización. A consecuencia de esto, se producían cadenas polipeptídicas de menor tamaño de DXS que la correspondiente a la secuencia completa por lo que se asumió la acción de una proteasa proveniente de este hongo. En el caso de *Arabidopsis thaliana,* la caracterización de un mutante que en teoría impedía la síntesis de proteínas en los plastos, dio un fenotipo de supresión de bloqueo por inhibidor de la ruta del MEP, por un incremento y mayor estabilidad de las enzimas de la ruta y un aumento de la proteasa Clp1 (Flores-Perez, Sauret-Gueto et al. 2008). Se asumió que el aumento observado en los niveles de la proteasa se debía a que no era funcional, ya que era sintetizada en los plastos, y la mayor estabilidad de las enzimas de la ruta del MEP se consideró debida a que resultaban diana de ésta.

En este capítulo se describe la proteólisis que sufren los homólogos de DXS de los organismos *Escherichia coli*, *Arabidopsis thaliana* y *Thermotoga maritima*. El trabajo parte de la purificación de las tres proteínas recombinantes a partir de *Escherichia coli*. Al resolver los tamaños mediante electroforesis en geles de poliacrilamida-SDS (SDS-PAGE) se observó en cada caso que, además del peso molecular correspondiente a la secuencia entera de aminoácidos, otras cadenas polipeptídicas de menor tamaño. Mediante la técnica de "western blot" se determinó que eran cadenas polipeptídicas relacionadas con la DXS. La immunodetección se realizó con un anticuerpo específico contra las histidinas insertadas en las proteínas recombinantes y, aunque principalmente las clonaciones se realizaron añadiendo la cola de His en el extremo carboxiterminal, en el caso de la DXS de *Escherichia coli* también se le añadió en el extremo aminoterminal. Los resultados por "western blot" contra la cola de His carboxiterminal de fracciones purificadas de DXS (DXS-His) de *Escherichia coli* y *Thermotoga maritima*, los cuáles tienen un peso molecular parecido, presentaban un patrón similar en donde aparecían bandas del peso molecular esperado para la secuencia completa y bandas adicionales de peso molecular alrededor de 50 kDa y 40 kDa. Al realizar la immunodetección contra fracciones purificadas que contenían las histidinas en el extremo aminoterminal (His-DXS) se observó, a parte del peso molecular de la proteína completa, unas bandas de pesos moleculares de aproximadamente de 50 kDa, 35 kDa y 25 kDa. El patrón de proteólisis observado no variaba al realizar pruebas con distintos compuestos inhibidores de proteasas durante la purificación de las proteínas.

Para poder detectar el punto de segmentación de la proteólisis producida y, basándonos en los resultados de la

immunodetección de las DXS-His, se realizaron experimentos de secuenciación aminoterminal de Edman. La secuenciación de la banda de peso molecular de alrededor de 50 kDa de *Escherichia coli* daba como resultado la secuencia Ser-Leu-Arg-Glu-Gly. Este péptido sugería la segmentación del enlace peptídico entre la Ser^{206} y la Ser^{207}, que justamente están situadas en una de las zonas no resueltas de la estructura de la DXS de *Escherichia coli* (Xiang, Usunow et al. 2007). La predicción del peso molecular del péptido carboxiterminal de este punto de segmentación sería de 46,2 kDa. La secuenciación de la banda de peso molecular de alrededor de 40 kDa de *Escherichia coli* dió como resultado la secuencia Pro-Ile-Thr-Phe-His, sugiriendo la segmentación del enlace peptídico entre el Asp^{294} y la Pro^{295} que se ubica dentro de la segunda zona no resuelta de la estructura cristalina.
Los experimentos de secuenciación de Edman en *Thermotoga maritima* se realizaron sobre los péptidos de mayor abundancia. Los resultados a partir de la banda de alrededor de 40 kDa dieron una secuencia Ser-Val-Val-His-Val y a partir de la banda de 30kDa dieron una secuencia Met-Leu-Leu-Asp-Glu. Éstas correspondían al extremo carboxiterminal de la posible segmentación del enlace peptídico entre la Ser^{264} y la Ser^{265} y el extremo aminoterminal de la proteína respectivamente. La secuenciación de la banda de alrededor de 46 kDa de *Arabidopsis thaliana* dio un péptido Ala-Ala-Lys-Val-Asp que correspondería al enlace entre Leu^{301}-Ala^{302}. La presencia de Ser en los residuos implicados en la proteólisis de dos de las tres proteínas indicaba su importancia en el proceso. Como se ha descrito en la Introducción (véase apartado 1.5.3.1.), la formación del enlace éster mediante el ataque nucleofílico de la cadena lateral de residuos tipo Ser, Cys o Thr, es el primer paso y crucial para la formación de un proceso autoproteolítico de dicho tipo. Así mismo, residuos Asp también aparecían en algunos de los enlaces peptídicos segmentados (véase apartado 1.5.3.2.).
En los ensayos *in vitro* de transcripción y traducción de las tres proteínas homologas, realizado en reticulocito de conejo marcando la proteína con L-[4,5-^{3}H] Leucina, se observaron varias bandas que sugieren que mecanismos de proteólisis también tienen lugar en este sistema. En estos ensayos se observó un patrón de proteólisis similar en todas, donde aparecía la banda correspondiente al peso molecular de la secuencia completa y una banda correspondiente a unos 50 kDa, en *Escherichia coli* y *Thermotoga maritima* y de peso molecular algo mayor en *Arabidopsis thaliana* (debido al mayor peso molecular de este homólogo). La proteólisis observada en el ensayo de transcripción y traducción *in vitro* aportaba más indicios sobre la posibilidad de un procesamiento autocatalítico ya que, un punto importante en la caracterización de estos procesos es su presencia en modelos o ensayos libres de células.
Para confirmar esta hipótesis se desnaturalizó la DXS-His de *Escherichia coli* y se resolvió por SDS-PAGE. La banda correspondiente a la secuencia entera de la proteína (70 kDa) se recortó del gel y se electroeluyó para tenerla en solución. Una vez electroeluida se renaturalizó en un tampón de renaturalización, realizándose también un control en un agente desnaturalizante (tampón con una concentración 8M de urea). Al realizar la immunodetección de la banda de 70 kDa renaturalizada y desnaturalizada, se observó que la muestra renaturalizada había vuelto a sufrir el proceso de proteólisis mientras que la muestra desnaturalizada no mostraba este proceso. En el patrón de proteólisis de la banda renaturalizada, se observó que aunque la banda de tamaño 40kDa se mantenía similar, la banda de alrededor 50 kDa de tamaño observada en las fracciones purificadas de la enzima migraba con un tamaño aparente algo mayor de 50kDa. Aunque no se pueda descartar, se presenta difícil la posibilidad que existiera una proteasa específica de DXS que copurificara con ella, que fuera de tamaño similar y que estuviera presente en reticulocitos de un organismo superior que no contiene secuencia codificante de DXS en su genoma. Además, el hecho de que la proteólisis se estuviera produciendo con la proteína

renaturalizada indicaba la importancia de la estructura en el proceso de segmentación de los enlaces peptídicos.

El plegamiento de la proteína se ha descrito de vital importancia para la viabilidad de los procesos autoproteolíticos. El correcto plegamiento puede sustentar toda una serie de interacciones que al fin permiten la formación de las reorganizaciones en el enlace peptídico. Por tanto, la zona implicada en el autoprocesamiento conlleva una región más amplia que el simple enlace peptídico segmentado. En consecuencia, insertar los aminoácidos importantes en el proceso autoproteolítico puede producir la autoproteólisis en otra proteína quimérica. Con esta idea se preparó una construcción que codificaba para la glutatión-S-transferasa (GST) fusionada a la proteína verde fluorescente (GFP). En esta construcción se insertó en pauta de traducción la secuencia que codificaba para la zona no resuelta en la cristalización de la DXS de *Escherichia coli*. La zona no resuelta en *Escherichia coli* (del residuo 183 al 238) se amplió a la zona que codificaba del residuo 158 al 240 para evitar la no inclusión de residuos importantes. Se purificaron y realizaron los ensayos de immunodetección de esta proteína quimérica utilizando la proteína GST-GFP como control. La immunodetección con anticuerpos específicos contra la cola de His y contra la GST mostró, principalmente, la presencia de una banda de peso molecular correspondiente a toda la proteína quimérica y otra de un peso molecular algo mayor que las proteínas de fusión que las constituían que correspondería a una de las proteínas de fusión más una parte del rizo insertado. La presencia de esta banda de menor peso molecular se observó en las proteínas quiméricas con el rizo insertado en comparación a la proteína GST-GFP lo que sugiere que la inserción de esta región de la proteína induce el proceso de proteólisis.

La hidroxilamina es un agente nucleófilo muy reactivo con especies (tio)éster (Jencks, Cordes et al. 1960) y succinimidas (Kwong and Harris 1994) pero demasiado débil como para romper en enlace amida (Rosenblum and Blobel 1999). Con esta premisa se procedió a caracterizar la posible hidroxilaminosis sobre la DXS-His de *Escherichia coli* con el correcto plegamiento proteico y desnaturalizada. La incubación con hidroxilamina aumentó el proceso de proteólisis en comparación con la muestra control. El ensayo por immunodetección del extremo carboxiterminal mostró un patrón de migración de las bandas similar al observado en el ensayo después de renaturalizar la banda de 70kDa, donde aparecían bandas de unos 55 kDa y la de 40 kDa. También se observó que la hidroxilaminosis se producía tanto en la proteína correctamente plegada como en la proteína desnaturalizada con urea.

La literatura describe extensamente la importancia del plegamiento de la proteína para mantener las interacciones que permiten la formación del enlace éster, ya que en caso contrario, el equilibrio de esta reacción esta favorecido hacía la formación del enlace amida (Rosenblum and Blobel 1999; Paulus 2000; Johansson, Macao et al. 2008). La hidroxilaminosis observada tanto en la proteína plegada como en la proteína desnaturalizada sugería que su causa era debida, no a la formación del enlace éster, sino a la presencia de succinimidas/intermediarios anhídridos cíclicos estables.

Para poder corroborar esto se secuenció el extremo aminoterminal de la banda de 55 kDa aparecida por hidroxilaminosis dando una secuencia Gly-Ala-Ile-Thr-Ala y la de 40 kDa dando una secuencia Pro-Ile-Thr-Phe-His. Estas secuencias sugerían la segmentación de los enlaces peptídicos con Asp implicados entre los residuos Asp^{152}-Gly^{153} y Asp^{294}-Pro^{295} lo que corroboraría la formación de succinimidas/intermediarios anhídridos cíclicos. Cabe destacar que, aunque la secuenciación de la banda de 40 kDa obtenida por hidroxilaminosis de la DXS sugería el mismo punto de segmentación que el acaecido en la banda de similar tamaño de las fracciones purificadas, también sugería la posible formación de otros intermediarios anhídridos cíclicos en el Asp^{305} debido

a la secuenciación de otro extremo aminoterminal (PSSG).

Finalmente, para comprobar la importancia del proceso proteolítico en la actividad enzimática de la DXS, se realizaron diferentes construcciones de DXS de *Escherichia coli* deleccionando el rizo (no resuelto en la estructura) donde se ubican los residuos Ser observados en la proteólisis. Los ensayos de complementación con la cepa *EcAB4-2* mostraron que las construcciones con el rizo deleccionado tenían actividad DXS.

3.2 DISCUSIÓN

En los tres primeros capítulos de esta Tesis se han estudiado temas relacionados con la oligomerización de la enzima CDP-ME quinasa (CMK). A partir de la estructura cristalina de la enzima de *E.coli* (Miallau, Alphey et al. 2003), en la cual se la define como a un homodímero, se han buscado moléculas por métodos *in silico* capaces de competir por las interacciones proteína-proteína del dímero. A varias de estas moléculas se le ha determinado su capacidad de interaccionar con la CMK mediante la utilización de métodos empíricos y, en algunos casos, se le han predicho valores de afinidad de unión a la zona de interacción homodimérica mejores que la propia formación del dímero.

Al llevar a cabo la caracterización teórica se ha visto que el solvente es un elemento crucial tanto para estabilizar las interacciones del dímero de la CMK aumentando su superficie, como para compensar su carga negativa favoreciendo los valores de energía electroestática. Esta importancia se ha visto corroborada por la posibilidad de incluir hasta tres puntos farmacofóricos, dentro de las hipótesis farmacofóricas, que imitan interacciones mediadas por agua y obtener moléculas capaces de interaccionar con la CMK.

La controversia respecto al estado de oligomerización de la CMK se observa al iniciar el estudio de los efectos de las moléculas encontradas sobre dicha oligomerización. A pesar de haber sido caracterizada como un dímero (Miallau, Alphey et al. 2003; Gabrielsen, Bond et al. 2004), los ensayos de cromatografía de exclusión molecular y dispersión dinámica de la luz obtenidos muestran la existencia de una mezcla de especies oligoméricas en las preparaciones purificadas de CMK. En esta mezcla la forma monomérica es la especie preponderante, aunque también se observa la presencia del tamaño correspondiente al dímero. Estos resultados se confirmaron mediante electroforesis en geles de poliacrilamida en condiciones nativas. La no detección de las formas diméricas puede ser debida a que su presencia estaba por debajo del umbral de detección de la tinción general de proteínas por Comassie blue R-250. La observación empírica de la mezcla de especies en la oligomeración de la CMK puede ser consecuente con el bajo valor obtenido, en comparación con otros complejos diméricos, de la predicción de la energía libre de unión del dímero.

En los ensayos de caracterización de la CMK de *Escherichia coli* realizados por Gabrielsen y sus colaboradores, en los que se la describe como una enzima dimérica, la concentración de sal en la solución es tres veces superior a la utilizada en nuestros estudios. Por el contrario, en la caracterización de la CMK de *Thermus thermophilus* (Wada, Kuzuyama et al. 2003), en la cual se la describe como enzima monomérica, la concentración de sal se asemeja más a la utilizada en nuestros experimentos que a los realizados por Gabrielsen. Aunque, es posible que el estado de oligomerización de la CMK pueda variar en función de la fuerza iónica del medio tal y como se ha descrito para otras proteínas, no se puede descartar otras condiciones experimentales que puedan influenciar el estado de oligomerización. En este sentido, el hecho que, al aplicar la proteína en diferentes concentraciones en la columna de exclusión molecular se obtuvieran diferentes ratios dímero:monómero, aunque siempre con preponderancia del monómero, podría apoyar esta idea.

Las estructuras cristalinas de la CMK de *Thermus thermophilus* (Wada, Kuzuyama et al. 2003) y de *Aquifex aeolicus* (Sgraja, Alphey et al. 2008) muestran las enzimas en forma monomérica. No se han observado cambios alostéricos significativos entre estas tres estructuras por lo que la oligomerización, en lo que respecta a la estructura monomérica, no se presenta como un factor importante. Un factor que podría influir en la oligomerización de la estructura tridimensional podría ser las fuerzas de empaquetamiento cristalinas (Carugo

and Argos 1997). Éstas podrían haber forzado, durante la cristalización de la CMK de *Escherichia coli,* a adoptar la forma dimérica. Las fuerzas de empaquetamiento podrían haber desplazado el aparente equilibrio dinámico observado mediante los ensayos de unión química cruzada con formaldehído y cromatografía de exclusión molecular, hacía la forma dimérica durante la producción de los cristales.

Se ha observado que las fracciones monoméricas de la CMK de *E. coli* presentan valores de actividad enzimática similares a los obtenidos con preparaciones purificadas de la enzima que contienen una mezcla de especies con diferente oligomerización. Por otro lado, ninguna de las moléculas que mostraban interacción con la CMK tuvo algún efecto significativo sobre la actividad. Estos resultados sugieren que no existe relación directa entre oligomerización y actividad enzimática, o, en otras palabras, que el estado de oligomerización no tiene efecto sobre la actividad enzimática. El centro catalítico de la CMK está localizado en una cavidad cerca de los residuos implicados en la unión de los sustratos. Estos residuos están situados dentro un mismo monómero y el centro catalítico, por tanto, se encuentra alejado de la zona de interacción proteína-proteína de las dos subunidades del dímero. El hecho de que cada monómero de la CMK disponga de un centro catalítico y los residuos implicados en la actividad es consistente con que el monómero presente actividad enzimática.

Como ya se ha comentado en la Introducción, la CMK pertenece a la superfamilia GHMP de quinasas. Esta superfamilia de pequeñas proteínas, principalmente quinasas, mantiene un plegamiento proteico α+β. En todas las proteínas de esta superfamilia el centro catalítico se presenta alejado de la superficie de interacción proteína-proteína con todos los residuos catalíticos dentro de una misma cadena polipeptídica. Pese a esto, se ha observado que la controversia relacionada con la oligomerización de estas proteínas no solo atañe a la CMK, sino que esta presente en toda la superfamília.

Las mevalonato quinasas (MVK) de S*treptococcus pneumoniae*, de *Rattus norvergicus* y de *Homo sapiens* se han descrito como complejos homodiméricos tanto en solución como en su estructura cristalina (Fu, Wang et al. 2002; Andreassi and Leyh 2004; Fu, Voynova et al. 2008). Sin embargo, otros homólogos de la enzima como la MVK de *Leishmania major* se caracterizó como monómero en solución (Sgraja, Smith et al. 2007). De forma inversa, la MVK de *Methanococcus jannaschii* se ha descrito como dimérica en solución y monomérica en la estructura cristalina (Yang, Shipman et al. 2002).

Un caso muy parecido a la controversia presentada en esta tesis aparece en las mevalonato 5-difosfato decarboxilasas (MDD). Las estructuras cristalinas de la MDD de *Trypanosoma brucei*, de *Sthapylococcus aureus* y de *Saccharomyces cerevisiae* presentaron una oligomerización dimérica, sin embargo, los ensayos de cromatografía de exclusion molecular y de sedimentación mostraron variabilidad en el estado de oligomerización para las enzimas (Byres, Alphey et al. 2007). Mientras que en solución el homólogo de *Trypanosoma brucei* se presentaba como monomérico y el de *Sthapylococcus aureus* homodimérico, la MDD de *Saccharomyces cerevisiae* se presentaba como una mezcla de especies en experimentos de exclusión molecular y en forma dimérica en experimentos de sedimentación. En el caso de la MDD la interacción proteína-proteína entre las subunidades es muy parecida a la observada para la CMK, donde se muestran grupos de residuos hidrofóbicos, un par de interacciones salinas y una red de interacciones mediadas por solvente. Los autores concluyeron que la estabilidad del dímero en solución dependía de la superficie de interacción proteína-proteína y del número de interacciones salinas.

Las galactoquinasas de *Homo sapiens* y *Saccharomyces cerevisiae* se caracterizaron como monoméricas en solución y presentaron una estructura dimérica en la estructura cristalina (Timson, Ross et al. 2002; Timson and

Reece 2003). En este caso, se postuló la variabilidad observada en función de las fuerzas de empaquetamiento cristalino o, incluso, a su desconocimiento del estado *in vivo* de la proteína.

Como hemos ido comentando a través de ejemplos en la literatura, el grado de oligomerización no parece tener una importancia directa en los ensayos in *vitro* de la actividad enzimática en esta familia de proteínas. Sin embargo, no se puede descartar que otras posibles funciones estén relacionadas con la oligomerización, como por ejemplo, la estabilidad de la enzima *in vivo*. Un ejemplo de ésto, podría ser el observado a partir de la secuenciación de la mutación Ile268Thr en la enzima MVK en pacientes con síndrome de hiperimmunoglobulemia D/fiebres periódicas. Estos pacientes presentaban un 50% menos de proteína MVK por lo que los autores concluyeron que las mutaciones en residuos importantes para la formación del dímero podrían causar inestabilidad de la proteína (Hinson, Ross et al. 1999; Fu, Wang et al. 2002). Otra posibilidad, aunque existe controversia sobre la existencia de los canales metabólicos en la ruta del MEP (Gabrielsen, Bond et al. 2004; Lherbet, Pojer et al. 2006), fue la aportada por Gabrielsen al llevar a cabo experimentos de sedimentación. En ellos describían a la CMK como una enzima dimérica. Este grado de oligomerización dimérica se deducía al interpretar los tamaños moleculares que se producían al interaccionar la CMK con otras enzimas de la ruta del MEP. Además, se proponía que las interacciones observadas se debían a la posibilidad de formación de canales metabólicos en esta ruta anabólica.

Se ha ido describiendo la variabilidad de la oligomerización de las proteínas de la superfamília GHMP, lo que denota, hasta que se determinen correctamente las variables que puedan regular la oligomerización, la gran labilidad que presenta la interacción del complejo dimérico en esta familia de proteínas. En la caracterización empírica de la unión de los compuestos a la CMK descrita en el segundo capítulo se obtuvieron unas constantes de asociación relativamente no muy buenas (del orden 10^{-4}M). El carácter lábil de la interacción proteína-proteína entre las subunidades del homodímero presente en esta superfamilia mostraría que, quizá, estos valores son más optimistas de lo que cabría pensar, ya que se habrían encontrado compuestos capaces de interaccionar con la superficie cruda de la proteína. Este hecho permite corroborar la gran fiabilidad de la técnica de búsqueda de compuestos asistida por ordenador, mostrada en esta Tesis, a la hora de encontrar moléculas capaces de interaccionar con las dianas seleccionadas.

La DXS es una enzima que desempeña un papel crucial en varias rutas metabólicas y es por esa razón que se han ido caracterizando bastantes homólogos de diferentes organismos. En esta Tesis se presenta la primera caracterización de la DXS a partir de un microorganismo hipertermófilo. La elucidación de las estructuras terciarias a partir de homólogos termofílicos es una aproximación bastante utilizada en proteínas difíciles de cristalizar. En el capítulo cuarto se ha caracterizado el marco abierto de lectura TM_1770, predicho como DXS en el proyecto de secuenciación genómica de *Thermotoga maritima* (Nelson, Clayton et al. 1999), como codificante de una proteína con actividad enzimática DXS. Su secuencia aminoacídica presenta el motivo de unión a TPP $GDGX_{25\text{-}30}N$, el motivo consenso de transcetolasas situado en el extremo aminoterminal de la proteína, y los residuos caracterizados en la estructura cristalina de la DXS de *Escherichia coli* importantes para la unión de los sustratos. La capacidad enzimática de producir desoxi-D-xilulosa 5-fosfato y la homología de secuencia confirmó que TM_1770 era un homólogo de la DXS en *Thermotoga maritima*. Mediante mutagénesis dirigida se comprobó que el residuo His35 de la DXS de *Thermotoga maritima* era crucial para la transferencia

de protones en la catálisis de modo similar a la His48 de la DXS de *Escherichia coli* (Querol, Rodriguez-Concepcion et al. 2001) y la His49 de la DXS de *Agrobacterium tumesfaciens* (Lee, Oh et al. 2007).

Filogenéticamente, la DXS de *Thermotoga maritima* es más afín a la clase *Clostridia* del filo *Firmicutes*, que son bacterias Gram+, o a plantas que a otras bacterias Gram-. En la mayoría de las bacterias los genes que codifican para las enzimas de la ruta del MEP se presentan en forma de operones sin embargo, en el caso de *Thermotoga maritima* se presentan dispersos en el genoma. Una explicación podría venir dada por el proceso de la transferencia genética horizontal a partir de diferentes organismos. La similitud con el filo *Firmicutes* se hizo patente en todas las enzimas de la ruta del MEP lo que descartaría la transferencia genética horizontal de estos genes a partir de varios grupos diferentes de organismos. La similitud encontrada del marcador genético de la 16S ARNr 7-metilguanosina metiltransferasa de *Thermotoga maritima* con homólogas del filo *Firmicutes* descarta la transferencia genética horizontal de los genes de la ruta del MEP a partir de este filo y muestra la similitud filogenética con estas bacterias Gram+. La similitud entre *Thermotoga maritima* y este filo de bacterias Gram+ puede venir dada por las características especiales de este organismo que se ha descrito como uno de los más profundos y más lentamente evolucionados dentro de las bacterias Gram- (Achenbach-Richter, Gupta et al. 1987). *Thermotoga maritima* representa, probablemente, un puente evolutivo entre el reino de las eubacterias y el reino de las arqueas (Nelson, Clayton et al. 1999) debido a que posee un número elevado de genes similares a los de este último grupo.

La actividad de la proteína recombinante presentó un óptimo de temperatura a 70ºC, consistente con caracterizaciones anteriores de otras enzimas de *Thermotoga maritima*. Una de los mecanismos que confieren estabilidad a temperaturas altas es el aumento de residuos cargados en detrimento de pequeños residuos polares no cargados. Se ha descrito que este aumento produce una mayor rigidez en la estructura y evita posibles desaminaciones (Fukuchi and Nishikawa 2001). Cuando se comparan las proporciones de estos tipos de aminoácidos en homólogos de la DXS de diferentes organismos se observa que en la DXS de *Thermotoga maritima* éstos se presentan con unos niveles similares a los de otros organismos termófilos. El aumento de los residuos cargados en detrimento de los polares no cargados podría ser uno de los factores que explicarían el carácter termofílico de la enzima, aunque no se descarta que las interacciones con carácter hidrofóbico puedan tener también un papel importante. Cabe destacar que aunque existen otros organismos mesófilos que presentan un alto tanto por ciento de residuos cargados, éstos no son en detrimento de los residuos polares no cargados. Uno de estos casos es el de las DXS dentro del filo *Firmicutes.* El alto porcentaje de residuos cargados que contienen podría ser una de las razones de la similitud tan alta con la DXS de *Thermotoga maritima* observada en los estudios filogenéticos.

La caracterización de mecanismos autoproteolíticos presenta una ardua tarea ya que, además del proceso de caracterización en si, las proteasas se presentan de gran importancia en infinidad de procesos celulares. Para iniciar esta discusión partiremos del proceso lógico seguido durante el proceso experimental.

El proceso de proteólisis se observó en los tres homólogos tanto en la expresión recombinante en células de *Escherichia coli in vivo* y después de purificar por cromatografía de afinidad, como en un sistema de expresión *in vitro* libre de células.

Uno de los principales pasos a seguir, descritos en la bibliografía en este tipo de procesos, es la obtención de

mutantes en los residuos implicados para obtener la proteína con una velocidad disminuida del proceso (incluso a una velocidad 0). Para obtener una especie molecular de la proteína preprocesada y observar el proceso se decidió realizar el proceso de separar la proteína desnaturalizada, recortar del SDS-PAGE la proteína preprocesada y renaturalizarla. La observación de que la proteólisis se volvía a producir en el proceso de renaturalización y que ésta no se observaba cuando la proteína estaba desnaturalizada en urea, daba fuerza a la idea de que un autoprocesamiento estaba ocurriendo. Aunque no se había obtenido un mutante parcialmente procesado o con una velocidad menor, el proceso de desnaturalización-escisión de la banda-renaturalización en principio partía de la proteína preprocesada que es lo que se busca con los procesos de mutagénesis. En este experimento, a la par, también se demostraba la importancia crucial de la estructura de la proteína en el proceso.

Johanson y colabores definieron 6 puntos para poder aceptar la relación de plegamiento y proteólisis (Johansson, Macao et al. 2008). Aunque la mayoría de los puntos se refieren a los residuos implicados en el ataque nucleofílico y las energías establecidas por el enlace tensionado que se produce, la posibilidad de disminuir la velocidad del proceso autoproteólitico mediante mutaciones distantes del enlace o solventes desnaturalizantes como la urea es un punto importante.

Experimentos de proteólisis a lo largo del tiempo mostraron una no dependencia en segundo grado de la concentración inicial por lo que se descartó un proceso de autoproteólisis intermolecular.

El hecho de poder immunodetectar la DXS de *Escherichia coli* como DXS-his y his-DXS nos permitía observar dentro del patrón de bandas las posibles localizaciones del proceso. La immunodetección de DXS-his marcaba, a parte de la proteína preprocesada, dos cadenas polipeptídicas principalmente de alrededor de 50 kDa y 40 kDa y la immunodetección de His-DXS, a parte de la proteína preprocesada, marcaba tres cadenas polipeptídicas de alrededor de 50 kDa, 35 kDa y 25 kDa, respectivamente. Aunque la obtención de estas proteínas fueron experimentos por separado, se intentó observar si los péptidos presentaban complementariedad. Los experimentos de hidroxilaminosis, además de la proteólisis observada *in vivo* mostraban que las cadenas polipeptídicas 40 kDa-C y 35 kDa-N podrían ser complementarias. Sea como fuere, el patrón de péptidos observado indicaba que existen como mínimo dos puntos de procesamiento en la DXS de *Escherichia coli*.

Uno de los experimentos descritos en las caracterizaciones de proteínas que sufren mecanismos autoproteolíticos intramoleculares es la capacidad de producir la segmentación de dos proteínas de fusión cuando la zona implicada en el proceso es insertada entre ellas. La inserción de las zonas implicadas en nuestros experimentos mostraba la capacidad de estas zonas de proteolizarse, pero el patrón de los pesos moleculares descartaba la capacidad de unir las dos proteínas de fusión.

La secuenciación del extremo aminoterminal de la cadena polipeptídica de 50 kDa de las fracciones purificadas de la DXS-His de *Escherichia coli* acabó descartando un proceso de *splicing,* ya que, relacionaba el punto de segmentación entre los residuos Ser^{206} y Ser^{207}, que como hemos comentado ya, se presenta en una de las zonas no resueltas en la estructura. Este hecho, y la secuenciación de la cadena polipeptídica de 40 kDa de las fracciones purificadas de DXS de *Thermotoga maritima* que relacionaba la segmentación del enlace entre los residuos Ser^{264} y Ser^{265}, aún concordaban con la idea de autoproteólisis mediada por una reorganización de los grupos acilo N-O y la formación de un enlace éster. Sin embargo, la secuenciación de la cadena polipeptídica de 40 kDa de las fracciones purificadas de DXS de *Escherichia coli* sugería la presencia de mecanismos relacionados con el Asp.

La hidroxilamina es un fuerte nucleófilo capaz de reaccionar con enlaces éster y residuos succinimida,

segmentando el enlace éster y el enlace con el residuo carboxiterminal de la succinimida. En ambos casos es una prueba concluyente que dichos procesos están ocurriendo. La principal diferencia es que mientras en la formación del enlace éster, como ya se ha comentado, el correcto plegamiento de la proteína es trascendental para la formación del enlace éster y la hidroxilaminosis; en la formación de succinimidas/intermediarios anhídridos cíclicos, si éstos son estables, el plegamiento de la proteína no es necesario para la hidroxilaminosis. Los resultados mostraron un incremento de la proteólisis (hidroxilaminosis), con el mismo patrón tanto en la DXS de *Escherichia coli* correctamente plegada como en la proteína desnaturalizada. Esto sugería la formación de succinimidas/intermediarios anhídridos cíclicos estables. Interesantemente, mientras la cadena polipeptídica de 40 kDa se mantenía con una misma migración tratando con hidroxilamina y sin tratar, la cadena polipeptídica de alrededor de 50 kDa en la proteína no tratada, migraba con un tamaño de unos 4 o 5 kDa mayor tratada con hidroxilamina. La migración era similar a la cadena polipeptídica obtenida después de renaturalizar la proteína. La secuenciación de los residuos de la cadena polipeptídica de algo más de 50 kDa por hidroxilaminosis, mostró que el enlace entre Asp^{152} y Gly^{153} se había segmentado. Existe extensa literatura describiendo la formación de intermediarios succinimidas en la combinación de residuos Asp-Gly, lo que sumado a los experimentos de hidroxilaminosis corroboraba la hipótesis de que se podría estar formando succinimidas como mecanismo intramolecular en dicho Asp. La secuenciación del extremo aminoterminal de la cadena polipeptídica de 40 kDa mostraba la segmentación del enlace Asp^{294}-Pro^{295}. Este caso es particular ya que esta combinación no puede formar succinimidas debido a la imposibilidad del nitrógeno del enlace peptídico para poder realizar el ataque nucleofílico sobre el carboxilo de la cadena lateral (Li 2009). Sin embargo, se crean intermediarios anhídridos cíclicos mediante el ataque del carboxilo de la cadena lateral del Asp al carbonilo del enlace peptídico con el residuo Pro. Está descrita en la literatura la asociación de esta reordenación intramolecular a procesos de segmentación del enlace peptídico (Li 2009, Oliyai 1993) lo que podría explicar la observación de esta cadena polipeptídica como la predominante en los ensayos *in vivo* (dentro de las cadenas procesadas).

El patrón en la DXS de *Escherichia coli* de relación entre residuos Asp-Gly y Asp-Pro con el tamaño de los posibles péptidos carboxiterminales también se repite en *Thermotoga maritima* en los enlaces Asp^{138}-Gly^{139} y Asn^{283}-Pro^{284}. Este último, aunque difiera en la Asn por el Asp, también se ha asociado a procesos de segmentación del enlace peptídico mediante el ataque nucleofílico de la amida β sobre el carbonilo del enlace peptídico, formándose una aminosuccinimida. También el patrón es similar en *Arabidopsis thaliana* en los enlaces Asp^{218}-Gly^{219} y Asp^{385}-Pro^{386}. Este último enlace, presenta homología con el segundo extremo N-terminal observado mediante hidroxilaminosis en la cadena polipeptídica de 40 kDa de la DXS de *Escherichia coli*.

El tratamiento con hidroxilamina, aunque nos permite concluir la formación de succinimidas, nos plantea la duda de si la cadena polipeptídica de alrededor de 50 kDa formada, cuyo punto de segmentación está entre los residuos Ser-Ser, es capaz de producirse por un mecanismo de formación del enlace éster o por otro mecanismo. La hidroxilamina, como ya hemos comentado, no puede distinguir entre el enlace éster y la succinimida/intermediario anhídrido cíclico por lo que el tratamiento con ella podría estar creando un efecto "epistático" de estos intermediarios sobre el enlace éster con la consecuente no observación de la segmentación entre los residuos Ser. Esto podría explicarse por que la segmentación del enlace en la succinimida conllevaría a una desestructuración del plegamiento, implicando el reordenamiento inverso de los grupos acilo O-N para volver a formar el enlace amida entre los residuos serina. Tampoco es descartable un proceso secuencial dónde a

partir de la segmentación en la succinimida otros procesos estén implicados para obtener la cadena polipeptídica con la segmentación entre las dos Ser. El experimento de desnaturalización-escisión de la proteína preprocesada-renaturalización, en donde la proteína renaturalizada presenta la banda de migración similar a la obtenida por hidroxilaminosis sería consecuente con esta dicotomía. La aparición de esta cadena polipeptídica, a la que le suponemos la hidrólisis por la succinimida descrita en el experimento de hidroxilaminosis, podría indicar que principalmente la segmentación se produce en este punto y después se tendría que producir el procesamiento hasta los residuos serina, pero también podría indicar que aunque formadas las succinimidas, el correcto plegamiento implicado en el enlace éster no se habría constituido.

Llegados a este punto vale la pena remarcar que, teniendo indicios claros que la proteína sufre un proceso de proteólisis dependiente de su estructura, habiéndose demostrado la formación de succinimidas y encontrado residuos proclives a la formación de enlaces éster en los extremos aminoterminales de los péptidos procesados no se puede concluir si alguna de las dos modificaciones post-traduccionales como origen del procesamiento de la cadena polipeptídica de alrededor de 50 kDa. Esto es debido, por un lado, a que aunque la presencia de succinimidas se ha descrito en los procesos de splicing proteico como inductor de la segmentación del enlace peptídico existe extensa literatura que le atribuye otras funciones. Y, por otro lado, como hemos comentado los experimentos de hidroxilaminosis no pudieron detectar la presencia del enlace éster. En cambio, en referencia a la cadena polipeptídica de 40 kDa, la formación de intermediarios anhídridos cíclicos podría explicar la segmentación del enlace que se observa a lo largo de todos los ensayos realizados con la enzima.

Para discernir esto sería necesaria la mutagénesis dirigida de los residuos implicados en la formación de succinimidas por separado de la mutación de los residuos propensos a la formación del enlace éster. Además, la secuenciación de los residuos aminoterminales de la cadena polipeptídica algo mayor de 50kDa obtenida por renaturalización podría ayudar a corroborar la hipótesis de la succinimida.

Por otra parte, todos los mecanismos descritos en la literatura tienen una importante función para la proteína. Para poder encontrar lazos entre el posible mecanismo que está sucediendo en la DXS y su función, se realizaron una serie de construcciones deleccionando la parte no resuelta estructuralmente de la DXS de *Escherichia coli*. Estas construcciones fueron la DXS de *Escherichia coli* con el rizo no resuelto en la estructura que conecta la 4ª y 5ª laminas β (del residuo 190 al 240) deleccionado y ésta con la inserción de un rizo flexible de glicinas (Pro-7xGly-Ser). La cepa *EcAB4-2* transformada con estas construcciones era capaz de crecer en ausencia de MVA por lo que la delección de este rizo no elimina la actividad enzimática aunque la cepa transformada con el rizo flexible mostraba una mayor velocidad de crecimiento. La importancia de este rizo se mantiene elusiva ya que, aunque parezca no tener una relación directa con la actividad en los experimentos de complementación, puede que tenga importancia en otras funciones. La delección de este rizo se ha observado en la DXS2 respecto a la DXS1 de *Arabidopsis thaliana* lo que podría implicar una diferenciación funcional de esta familia multigénica, y podría corroborar esta idea.

Por otra parte, la construcción de la enzima con esta delección nos permitió tener una construcción con el residuo implicado en la succinimida (Asp^{152}) sin tener los residuos del posible enlace ester (Ser^{206} y Ser^{207}), por lo que en ausencia de mutantes de estos últimos se decidió a purificar la proteína. La proteína mostraba casi una total supresión del proceso de proteólisis mostrando la importancia del rizo en este proceso (y quizá con ello la formación del enlace éster), aunque no es descartable que la supresión de un segmento tan grande de la proteína no pudiera modificar la formación de la succinimida/intermediarios anhídridos cíclicos.

Abstrayéndonos de la relación entre el proceso de autoproteólisis y la presencia de succinimidas en la DXS, tan sólo la caracterización de estas succinimidas ya merece una mención especial. La formación de succinimidas normalmente se infiere a partir de la presencia de isoaspartatos (George-Nascimento, Lowenson et al. 1990; Violand, Schlittler et al. 1990; Di Donato, Ciardiello et al. 1993). Éstos se transfieren a succinimidas mediate la enzima PIMT y a partir de un tratamiento con hidroxilamina se puede observar el número de isoaspartatos/succinimidas formados. Nuestros experimentos de hidroxilaminosis nos indican que parte de las moléculas de DXS ya contienen los Asp^{152} en forma de succinimidas estables. Además, la formación de la succinimida en estos residuos puede tener una gran importancia en la actividad enzimática.

El residuo Asp^{152} es crucial para la actividad ya que forma parte del motivo de unión a TPP $GDGX_{25\text{-}30}N$, formando la estructura piramidal que mantiene la coordinación octaedrica del ión magnesio enlazado al grupo difosfato del TPP. La posibilidad de que este residuo entrara en el ciclo de Asp/Isoasp/succinimida y su posible regulación por la enzima PIMT, debido a la importancia de estos residuos Asp en la actividad enzimática, podría mostrar un mecanismo de regulación de su actividad más que un proceso de envejecimiento de la proteína (véase apartado 1.5.3.2.2.1) que podría ser extrapolable a todas las enzimas TPP dependientes. La alta conservación de ambos residuos Asp en todos los homólogos de DXS abriría la posibilidad de que estuviéramos ante un mecanismo general de regulación post-traduccional de la DXS.

3.3 CONCLUSIONES

1. Los estudios por mecánica molecular han permitido apreciar la importancia del agua para la estabilización *in silico* del complejo homodimérico de la CMK de *Escherichia coli*. Se han estudiado las interacciones proteína-proteína e interacciones mediadas por solvente del complejo a través de una trayectoria de dinámica molecular. Se han realizado dos hipótesis farmacofóricas con la intención de disruptar estas interacciones y se ha obtenido un gran número de moléculas que cumplen estas hipótesis. Paralelamente, se ha sintetizado un péptido cíclico imitando el rizo implicado en la interacción del complejo dimérico.
2. La unión de dos moléculas pequeñas, una por cada hipótesis farmacofórica, y el péptido cíclico con la CMK se ha caracterizado empíricamente, corroborando las hipótesis farmacofóricas planteadas, y se han realizado modelos *in silico* de los complejos de unión formados.
3. Los estudios de oligomerización de la CMK de *Escherichia coli* han mostrado una mezcla de especies entre la forma monomérica y la dimérica, siendo la primera la predominante, presentando un equilibrio dinámico entre ellas. La forma monomérica de la CMK presenta valores de actividad similares a la mezcla de especies, lo que sumado a que las moléculas diseñadas para disruptar la interacción del complejo dimérico no presentan ningún efecto inhibidor, permite concluir que las homodimerización de la CMK no es necesaria para su actividad. La controversia observada en relación a la oligomerización no es exclusiva de la CMK y se hace extensiva a toda la familia GHMP de quinasas con paralelismos muy cercanos dentro de las mevalonato quinasas y mevalonato 5-difosfato decarboxilasas.
4. El marco abierto de lectura TM_1770 del organismo *Thermotoga maritima* se ha caracterizado como DXS. Se han caracterizado sus principales constantes bioquímicas, presentando un óptimo de temperatura a 70º C. El residuo conservado histidina 35 se ha mostrado crucial en la actividad enzimática, tal y como se había descrito para esta histidina en otros homólogos. Se han presentado análisis filogenéticos para esta enzima y se han formulado hipótesis estructurales que sustenten la capacidad termofílica de la enzima.
5. Se ha observado un patrón de proteólisis similar durante la purificación de homólogos de DXS recombinantes de los organismos *Escherichia coli, Arabidopsis thaliana y Thermotoga maritima*. La proteólisis también se ha observado durante la transcripción/traducción *in Vitro* de estas proteínas. El plegamiento de la DXS de *Escherichia coli* se ha mostrado importante para el proceso de proteólisis, aportando indicios de un posible mecanismo autocatalítico. Aunque la caracterización de los residuos implicados en los posibles puntos de segmentación aportan indicios de un posible mecanismo relacionado con la formación de un enlace éster, los experimentos de hidroxilaminosis no han podido corroborar la presencia del enlace éster por la presencia de succinimidas/intermediarios anhídridos cíclicos en residuos Asp muy conservados en todas las DXS. La formación de intermediarios anhídridos cíclicos a partir del enlace Asx(Asp,Asn)-Pro, correspondiente a la segmentación observada que produce la cadena polipeptídica de 40 kDa, sugiere a ésta como el origen de la segmentación debido a la alta reactividad descrita de dicho enlace. Además, la presencia de succinimidas en residuos importantes para la actividad enzimática muestra la posibilidad de un posible mecanismo autoprocesivo de regulación enzimática post-traduccional.

BIBLIOGRAFÍA

4. Bibliografía

Abe, J., T. Fukuzawa, et al. (2002). "Cleavage of Ig-Hepta at a "SEA" module and at a conserved G protein-coupled receptor proteolytic site." J Biol Chem **277**(26): 23391-8.

Accelrys Inc., U. "CATALYST™."

Achenbach-Richter, L., R. Gupta, et al. (1987). "Were the original eubacteria thermophiles?" Syst Appl Microbiol **9**: 34-9.

Adam, P., S. Hecht, et al. (2002). "Biosynthesis of terpenes: studies on 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate reductase." Proc Natl Acad Sci U S A **99**(19): 12108-13.

Amitai, G., O. Belenkiy, et al. (2003). "Distribution and function of new bacterial intein-like protein domains." Mol Microbiol **47**(1): 61-73.

Andreassi, J. L., 2nd and T. S. Leyh (2004). "Molecular functions of conserved aspects of the GHMP kinase family." Biochemistry **43**(46): 14594-601.

Andrews, M. J., C. McInnes, et al. (2004). "Design, synthesis, biological activity and structural analysis of cyclic peptide inhibitors targeting the substrate recruitment site of cyclin-dependent kinase complexes." Org Biomol Chem **2**(19): 2735-41.

Artymiuk, P. J. (1995). "A sting in the (N-terminal) tail." Nat Struct Biol **2**(12): 1035-7.

Bach II, A., C. Eyermann, et al. (1994). "Structural studies of a Family of high affinity ligands for GPIIb/IIIa " J Am Chem Soc **116**(8): 3207-3219

Bean, L. E., W. H. Dvorachek, Jr., et al. (2001). "Analysis of the pdx-1 (snz-1/sno-1) region of the Neurospora crassa genome: correlation of pyridoxine-requiring phenotypes with mutations in two structural genes." Genetics **157**(3): 1067-75.

Berendsen, H., J. Postma, et al. (1984). "Molecular dynamics with coupling to an external bath." J Chem Phys **81**(8): 3684.

Bernal, C., E. Mendez, et al. (2005). "A spectrophotometric assay for the determination of 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase activity." Anal Biochem **340**(2): 245-51.

Boanca, G., A. Sand, et al. (2006). "Uncoupling the enzymatic and autoprocessing activities of Helicobacter pylori gamma-glutamyltranspeptidase." J Biol Chem **281**(28): 19029-37.

Bochtler, M., L. Ditzel, et al. (1997). "Crystal structure of heat shock locus V (HslV) from Escherichia coli." Proc Natl Acad Sci U S A **94**(12): 6070-4.

Bonanno, J. B., C. Edo, et al. (2001). "Structural genomics of enzymes involved in sterol/isoprenoid biosynthesis." Proc Natl Acad Sci U S A **98**(23): 12896-901.

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

Borrmann, S., S. Issifou, et al. (2004). "Fosmidomycin-clindamycin for the treatment of Plasmodium falciparum malaria." J Infect Dis **190**(9): 1534-40.

Boucher, Y. and W. F. Doolittle (2000). "The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways." Mol Microbiol **37**(4): 703-16.

Brange, J., L. Langkjaer et al. (1992). "Chemical stability of insulin. 1. Hydrolytic degradation during storage of pharmaceutical preparations" Pharm Res. 1992 Jun;9(6):715-26.

Brannigan, J. A., G. Dodson, et al. (1995). "A protein catalytic framework with an N-terminal nucleophile is capable of self-activation." Nature **378**(6555): 416-9.

Byres, E., M. S. Alphey, et al. (2007). "Crystal structures of Trypanosoma brucei and Staphylococcus aureus mevalonate diphosphate decarboxylase inform on the determinants of specificity and reactivity." J Mol Biol **371**(2): 540-53.

Cane D.E, H. Y., Cornish J.A, Robinson J.K., Spencer I.D. (1998). "Biosynthesis of Vitamin B6: The Oxidation of 4-(Phosphohydroxy)-L-threonine by PdxA." J. Am. Chem. Soc. **1998**(120): 1936-1937.

Carugo, O. and P. Argos (1997). "Protein-protein crystal-packing contacts." Protein Sci **6**(10): 2261-3.

Case, D. A., T. E. Cheatham, 3rd, et al. (2005). "The Amber biomolecular simulation programs." J Comput Chem **26**(16): 1668-88.

Casey, W. M., G. A. Keesler, et al. (1992). "Regulation of partitioned sterol biosynthesis in *Saccharomyces cerevisiae*." J Bacteriol **174(22)**(Nov): 7283-7288.

Catak, S., G. Monard, et al. (2008). "Computational study on non enzymatic peptide cleavage at Asparagine and Aspartic residues" J Phys Chem 112(26): 8752–8761.

Cohen, C. and D. A. D. Parry (1986). "alfa-helical coiled-coils- a widespread motif in proteins." Trends Biochem Sci **11**: 245-8.

Chang, Y. P. and Y. H. Chu (2005). "Using surface plasmon resonance to directly determine binding affinities of combinatorially selected cyclopeptides and their linear analogs to a streptavidin chip." Anal Biochem **340**(1): 74-9.

Chapman, A. and B. Hill (2000). Dictionary of Natural Products on CD-ROM (version 9.1.). London, New York, CRC Press.

Chappell, J. (1995). "Biochemistry and Molecular Biology of the Isoprenoid Biosynthetic Pathway in Plants." Annu Rev Plant Physiol Plant Mol Biol **46**: 521-547.

Cheek, S., H. Zhang, et al. (2002). "Sequence and structure classification of kinases." J Mol Biol **320**(4): 855-81.

Dassa, B., N. London, et al. (2009). "Fractured genes: a novel genomic arrangement involving new split inteins and a new homing endonuclease family." Nucleic Acids Res **37**(8): 2560-73.

Di Donato, A., M. A. Ciardiello, et al. (1993). "Selective deamidation of ribonuclease A. Isolation and characterization of the resulting isoaspartyl and aspartyl derivatives." J Biol Chem **268**(7): 4745-51.

Diliberto, E. J., Jr. and J. Axelrod (1976). "Regional and subcellular distribution of protein carboxymethylase in brain and other tissues." J Neurochem **26**(6): 1159-65.

Dorer, F. E., E. E. Haley, et al. (1968). "The hydrolysis of beta-aspartyl peptides by rat tissue." Arch Biochem Biophys **127**(1): 490-5.

Dumez, E., J. S. Snaith, et al. (2002). "Synthesis of macrocyclic, potential protease inhibitors using a generic scaffold." J Org Chem **67**(14): 4882-92.

Echelard, Y., D. J. Epstein, et al. (1993). "Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity." Cell **75**(7): 1417-30.

Ehrenshaft, M., P. Bilski, et al. (1999). "A highly conserved sequence is a novel gene involved in de novo vitamin B6 biosynthesis." Proc Natl Acad Sci U S A **96**(16): 9374-8.

Eisenreich, W., F. Rohdich, et al. (2001). "Deoxyxylulose phosphate pathway to terpenoids." Trends Plant Sci **6**(2): 78-84.

Eisenreich, W., M. Schwarz, et al. (1998). "The deoxyxylulose phosphate pathway of terpenoid biosynthesis in plants and microorganisms." Chem Biol **5**(9): R221-33.

Ekici, O. D., M. Paetzel, et al. (2008). "Unconventional serine proteases: variations on the catalytic Ser/His/Asp triad configuration." Protein Sci **17**(12): 2023-37.

Emtage, J. L., M. Bucci, et al. (1997). "Defining the essential functional regions of the nucleoporin Nup145p." J Cell Sci **110 (Pt 7)**(Pt 7): 911-25.

Estevez, J. M., A. Cantero, et al. (2001). "1-Deoxy-D-xylulose-5-phosphate synthase, a limiting enzyme for plastidic isoprenoid biosynthesis in plants." J Biol Chem **276**(25): 22901-9.

Flores-Perez, U., S. Sauret-Gueto, et al. (2008). "A mutant impaired in the production of plastome-encoded proteins uncovers a mechanism for the homeostasis of isoprenoid biosynthetic enzymes in Arabidopsis plastids." Plant Cell **20**(5): 1303-15.

Frank, R. A., F. J. Leeper, et al. (2007). "Structure, mechanism and catalytic duality of thiamine-dependent enzymes." Cell Mol Life Sci **64**(7-8): 892-905

Fu, Z., N. E. Voynova, et al. (2008). "Biochemical and structural basis for feedback inhibition of mevalonate kinase and isoprenoid metabolism." Biochemistry **47**(12): 3715-24.

Fu, Z., M. Wang, et al. (2002). "The structure of a binary complex between a mammalian mevalonate kinase and ATP: insights into the reaction mechanism and human inherited disease." J Biol Chem **277**(20): 18134-42.

Fukuchi, S. and K. Nishikawa (2001). "Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria." J Mol Biol **309**(4): 835-43.

Gabrielsen, M., C. S. Bond, et al. (2004). "Hexameric assembly of the bifunctional methylerythritol 2,4-cyclodiphosphate synthase and protein-protein associations in the deoxy-xylulose-dependent pathway of isoprenoid precursor biosynthesis." J Biol Chem **279**(50): 52753-61.

George-Nascimento, C., J. Lowenson, et al. (1990). "Replacement of a labile aspartyl residue increases the stability of human epidermal growth factor." Biochemistry **29**(41): 9584-91.

Giebel, L. B., R. T. Cass, et al. (1995). "Screening of cyclic peptide phage libraries identifies ligands that bind streptavidin with high affinities." Biochemistry **34**(47): 15430-5.

Groll, M., L. Ditzel, et al. (1997). "Structure of 20S proteasome from yeast at 2.4 A resolution." Nature **386**(6624): 463-71.

Hahn, F. M., L. M. Eubanks, et al. (2001). "1-Deoxy-D-xylulose 5-phosphate synthase, the gene product of open reading frame (ORF) 2816 and ORF 2895 in Rhodobacter capsulatus." J Bacteriol **183**(1): 1-11.

Hall, T.M., J.A. Porter, et al. (1997) "Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins." Cell **91**(1):85-97.

Hao, L., R. Johnsen, et al. (2006). "Comprehensive analysis of gene expression patterns of hedgehog-related genes." BMC Genomics **7**(280): 280.

Harker, M. and P. M. Bramley (1999). "Expression of prokaryotic 1-deoxy-D-xylulose-5-phosphatases in Escherichia coli increases carotenoid and ubiquinone biosynthesis." FEBS Lett **448**(1): 115-9.

Hilkens, J., F. Buijs, et al. (1984). "Monoclonal antibodies against human milk-fat globule membranes detecting differentiation antigens of the mammary gland and its tumors." Int J Cancer **34**(2): 197-206.

Hinson, D. D., R. M. Ross, et al. (1999). "Identification of a mutation cluster in mevalonate kinase deficiency, including a new mutation in a patient of Mennonite ancestry." Am J Hum Genet **65**(2): 327-35.

Hirata, R., Y. Ohsumk, et al. (1990). "Molecular structure of a gene, VMA1, encoding the catalytic subunit of H(+)-translocating adenosine triphosphatase from vacuolar membranes of Saccharomyces cerevisiae." J Biol Chem **265**(12): 6726-33.

Humphrey, W., A. Dalke, et al. (1996). "VMD: visual molecular dynamics." J Mol Graph **14**(1): 33-8, 27-8.

Ingram, R. N., P. Orth, et al. (2006). "Stabilization of the autoproteolysis of TNF-alpha converting enzyme (TACE) results in a novel crystal form suitable for structure-based drug design studies." Protein Eng Des Sel **19**(4): 155-61.

Isupov, M. N., G. Obmolova, et al. (1996). "Substrate binding is required for assembly of the active conformation of the catalytic site in Ntn amidotransferases: evidence from the 1.8 A crystal structure of the glutaminase domain of glucosamine 6-phosphate synthase." Structure **4**(7): 801-10.

Iwai, K. and T. Ando (1967). Methods in enzymology **XI**: 262-282.

Jencks, W. P., S. Cordes, et al. (1960). "The free energy of thiol ester hydrolysis." J Biol Chem **235**: 3608-14.

Johansson, D. G., B. Macao, et al. (2008). "SEA domain autoproteolysis accelerated by conformational strain:

mechanistic aspects." J Mol Biol **377**(4): 1130-43.
Johnson B.A, E.L Langmack, et al. (1987). "Partial repair of deamidation-damaged calmodulin by protein carboxyl methyltransferase" J Biol Chem **262**(25):12283-7
Johnson B.A, E.D Murray, et al. (1987). "Protein carboxyl methyltransferase facilitates conversion of atypical L-isoaspartyl peptides to normal L-aspartyl peptides" J Biol Chem **262**(12):5622-9
Jomaa, H., J. Wiesner, et al. (1999). "Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs." Science **285**(5433): 1573-6.
Jorgensen, W., J. Chandrasekhar, et al. (1983). "Comparison of simple potential functions for simulating liquid water." J Chem Phys **79**: 926.
Kane, P. M., C. T. Yamashiro, et al. (1990). "Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase." Science **250**(4981): 651-7.
Kawasaki, M., S. Makino, et al. (1996). "Folding-dependent in vitro protein splicing of the Saccharomyces cerevisiae VMA1 protozyme." Biochem Biophys Res Commun **222**(3): 827-32.
Kawasaki, M., Y. Satow, et al. (1997). "Protein splicing in the yeast Vma1 protozyme: evidence for an intramolecular reaction." FEBS Lett **412**(3): 518-20.
Khatri, I. A., R. Wang, et al. (2003). "SEA (sea-urchin sperm protein, enterokinase and agrin)-module cleavage, association of fragments and membrane targeting of rat intestinal mucin Muc3." Biochem J **372**(Pt 1): 263-70.
Klabunde, T., S. Sharma, et al. (1998). "Crystal structure of GyrA intein from Mycobacterium xenopi reveals structural basis of protein splicing." Nat Struct Biol **5**(1): 31-6.
Kollman, P. A., I. Massova, et al. (2000). "Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models." Acc Chem Res **33**(12): 889-97.
Koonin, E. V. (1995). "A protein splice-junction motif in hedgehog family proteins." Trends Biochem Sci **20**(4): 141-2.
Kossiakoff, A. A. (1988). "Tertiary structure is a principal determinant to protein deamidation." Science **240**(4849): 191-4.
Kuzuyama, T., T. Shimizu, et al. (1998). "Fosmidomycin, a Specific Inhibitor of 1-Deoxy-d-Xylulose 5-Phosphate Reductoisomerase in the Nonmevalonate Pathway for Terpenoid Biosynthesis." Tetrahedron Letters **39**(43): 7913-7916.
Kuzuyama, T., S. Takahashi, et al. (2000). "Characterization of 1-deoxy-D-xylulose 5-phosphate reductoisomerase, an enzyme involved in isopentenyl diphosphate biosynthesis, and identification of its catalytic amino acid residues." J Biol Chem **275**(26): 19928-32.
Kwong, M. Y. and R. J. Harris (1994). "Identification of succinimide sites in proteins by N-terminal sequence analysis after alkaline hydroxylamine cleavage." Protein Sci **3**(1): 147-9.
Lange, B. M. and R. Croteau (1999). "Isopentenyl diphosphate biosynthesis via a mevalonate-independent pathway: isopentenyl monophosphate kinase catalyzes the terminal enzymatic step." Proc Natl Acad Sci U S A **96**(24): 13714-9.
Lange, B. M., M. R. Wildung, et al. (1998). "A family of transketolases that directs isoprenoid biosynthesis via a mevalonate-independent pathway." Proc Natl Acad Sci U S A **95**(5): 2100-4.
Lee, J. J., D. P. von Kessler, et al. (1992). "Secretion and localized transcription suggest a role in positional signaling for products of the segmentation gene hedgehog." Cell **71**(1): 33-50.
Lee, J. K., D. K. Oh, et al. (2007). "Cloning and characterization of the dxs gene, encoding 1-deoxy-d-xylulose 5-phosphate synthase from Agrobacterium tumefaciens, and its overexpression in Agrobacterium tumefaciens." J Biotechnol **128**(3): 555-66.
Leroux, V., N. Gresh, et al. (2006). "Role of water molecules for binding inhibitors in the SH2 domain of Grb2: A molecular dynamics study." teochem **806**(1-3): 51-66.
Levitin, F., O. Stern, et al. (2005). "The MUC1 SEA module is a self-cleaving domain." J Biol Chem **280**(39): 33374-86.
Lherbet, C., F. Pojer, et al. (2006). "Absence of substrate channeling between active sites in the Agrobacterium tumefaciens IspDF and IspE enzymes of the methyl erythritol phosphate pathway." Biochemistry **45**(11): 3548-53.
Li, N., F. Fort et al. (2009) "Factors affecting cleavage at aspartic residues in model decapeptides" J Pharm Biomed Anal. 2009 Aug 15;50(1):73-8.
Lichtenthaler, H. K. (1999). "The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants." Annu Rev Plant Physiol Plant Mol Biol **50**: 47-65.
Lichtenthaler, H. K. (2000). "Non-mevalonate isoprenoid biosynthesis: enzymes, genes and inhibitors." Biochem Soc Trans **28**(6 Dec): 785-9.
Lichtenthaler, H. K., M. Rohmer, et al. (1997). "Two independent biochemical pathways for isopentenyl diphosphate and isoprenoid biosynthesis in higher plants." Physiologia plantarum **101**: 643-652.
Lichtenthaler, H. K., Schwender, J., Disch, A., Rohmer, M. (1997). "Biosynthesis of isoprenoids in higher plant chloroplasts proceeds via a mevalonate-independent pathway." FEBS Lett **400**: 271-274.
Lillehoj, E. P., F. Han, et al. (2003). "Mutagenesis of a Gly-Ser cleavage site in MUC1 inhibits ectodomain shedding." Biochem Biophys Res Commun **307**(3): 743-9.
Lipinski, C. A., F. Lombardo, et al. (2001). "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings." Adv Drug Deliv Rev **46**(1-3): 3-26.
Lois, L. M., N. Campos, et al. (1998). "Cloning and characterization of a gene from Escherichia coli encoding a transketolase-like enzyme that catalyzes the synthesis of D-1-deoxyxylulose 5-phosphate, a common

precursor for isoprenoid, thiamin, and pyridoxol biosynthesis." Proc Natl Acad Sci U S A **95**(5): 2105-10.
Luttgen, H., F. Rohdich, et al. (2000). "Biosynthesis of terpenoids: YchB protein of Escherichia coli phosphorylates the 2-hydroxy group of 4-diphosphocytidyl-2C-methyl-D-erythritol." Proc Natl Acad Sci U S A **97**(3): 1062-7.
Luz, J. G., C. A. Hassig, et al. (2003). "XOL-1, primary determinant of sexual fate in C. elegans, is a GHMP kinase family member and a structural prototype for a class of developmental regulators." Genes Dev **17**(8): 977-90.
Lloyd, D. G., A. T. Garcia-Sosa, et al. (2004). "The effect of tightly bound water molecules on the structural interpretation of ligand-derived pharmacophore models." J Comput Aided Mol Des **18**(2): 89-100.
Mamula, M. J., R. J. Gee, et al. (1999). "Isoaspartyl post-translational modification triggers autoimmune responses to self-proteins." J Biol Chem **274**(32): 22321-7.
Margulis, L. (1975). "Symbiotic theory of the origin of eukaryotic organelles; criteria for proof." Symp Soc Exp Biol(29): 21-38.
Mathews (2003). "Bioquímica." (3ª edición).
McDonough, M. A., H. E. Klei, et al. (1999). "Crystal structure of penicillin G acylase from the Bro1 mutant strain of Providencia rettgeri." Protein Sci **8**(10): 1971-81.
McGarvey, D. J. and R. Croteau (1995). "Terpenoid metabolism." Plant Cell **7**(7): 1015-26.
Miallau, L., M. S. Alphey, et al. (2003). "Biosynthesis of isoprenoids: crystal structure of 4-diphosphocytidyl-2C-methyl-D-erythritol kinase." Proc Natl Acad Sci U S A **100**(16): 9173-8.
Missinou, M. A., S. Borrmann, et al. (2002). "Fosmidomycin for malaria." Lancet **360**(9349): 1941-2.
Morgan, B., D. Holland, et al. (1994). "Structure-based design of an inhibitor of the zinc peptidase Thermolysin." J Am Chem Soc **116**(8): 3251-3260.
Mueller, C., J. Schwender, et al. (2000). "Properties and inhibition of the first two enzymes of the non-mevalonate pathway of isoprenoid biosynthesis." Biochem Soc Trans **28**(6): 792-3.
Nam, N. H., G. Ye, et al. (2004). "Conformationally constrained peptide analogues of pTyr-Glu-Glu-Ile as inhibitors of the Src SH2 domain binding." J Med Chem **47**(12): 3131-41.
Nelson, K. E., R. A. Clayton, et al. (1999). "Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima." Nature **399**(6734): 323-9.
O'Connor, C. M. and S. Clarke (1985). "Specific recognition of altered polypeptides by widely distributed methyltransferases." Biochem Biophys Res Commun **132**(3): 1144-50.
Oinonen, C. and J. Rouvinen (2000). "Structural comparison of Ntn-hydrolases." Protein Sci **9**(12): 2329-37.
Oinonen, C., R. Tikkanen, et al. (1995). "Three-dimensional structure of human lysosomal aspartylglucosaminidase." Nat Struct Biol **2**(12): 1102-8.
Oliyai, C., R.T. Borchardt (1993) "Chemical pathways of peptide degradation. IV. Pathways, kinetics, and mechanism of degradation of an aspartyl residue in a model hexapeptide." **Pharm Res. 1993 Jan;10(1):95-102.**
Osmani, A. H., G. S. May, et al. (1999). "The extremely conserved pyroA gene of Aspergillus nidulans is required for pyridoxine synthesis and is required indirectly for resistance to photosensitizers." J Biol Chem **274**(33): 23565-9.
Pandian, S., Saengchjan, S., Raman, T.S. (1981). "An alternative pathway for the biosynthesis of isoprenoid compounds in bacteria." Biochem J **196(3)**(Jun 15): 675-681.
Paulus, H. (2000). "Protein splicing and related forms of protein autoprocessing." Annu Rev Biochem **69**: 447-96.
Pepperkok, R., A. Hotz-Wagenblatt, et al. (2000). "Intracellular distribution of mammalian protein kinase A catalytic subunit altered by conserved Asn2 deamidation." J Cell Biol **148**(4): 715-26.
Pietrokovski, S. (1994). "Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins." Protein Sci **3**(12): 2340-50.
Pietrokovski, S. (1998). "Modular organization of inteins and C-terminal autocatalytic domains." Protein Sci **7**(1): 64-71.
Puente, X. S., L. M. Sanchez, et al. (2005). "A genomic view of the complexity of mammalian proteolytic systems." Biochem Soc Trans **33**(Pt 2): 331-4.
Querol, J., M. Rodriguez-Concepcion, et al. (2001). "Essential role of residue H49 for activity of Escherichia coli 1-deoxy-D-xylulose 5-phosphate synthase, the enzyme catalyzing the first step of the 2-C-methyl-D-erythritol 4-phosphate pathway for isoprenoid Synthesis." Biochem Biophys Res Commun **289**(1): 155-60.
Rawn, J. D. (1989). Bioquímica. Madrid, McGraw Hill-Interamericana de España.
Recsei, P. A. and E. E. Snell (1984). "Pyruvoyl enzymes." Annu Rev Biochem **53**: 357-87.
Reissner, K. J. and D. W. Aswad (2003). "Deamidation and isoaspartate formation in proteins: unwanted alterations or surreptitious signals?" Cell Mol Life Sci **60**(7): 1281-95.
Robinson, A. B. and C. J. Rudd (1974). "Deamidation of glutaminyl and asparaginyl residues in peptides and proteins." Curr Top Cell Regul **8**(0): 247-95.
Rodríguez-Concepción, M. and A. Boronat (2002). "Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieved through genomics." Plant Physiol **130**(3): 1079-89.
Rodwell, A. W. (1953). "The occurrence and distribution of amino-acid decarboxylases within the genus Lactobacillus." J Gen Microbiol **8**(2): 224-32.
Rohdich, F., S. Hecht, et al. (2002). "Studies on the nonmevalonate terpene biosynthetic pathway: Metabolic role of IspH (LytB) protein." Proc Natl Acad Sci USA **99 Feb 5**(3): 1158-1163.
Rohdich, F., J. Wungsintaweekul, et al. (1999). "Cytidine 5'-triphosphate-dependent biosynthesis of isoprenoids:

YgbP protein of Escherichia coli catalyzes the formation of 4-diphosphocytidyl-2-C-methylerythritol." Proc Natl Acad Sci USA **96**(21): 11758-63.

Rohdich, F., J. Wungsintaweekul, et al. (2000). "Biosynthesis of terpenoids: 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase from tomato." Proc Natl Acad Sci U S A **97**(15): 8251-6.

Rohmer, M. (1999). "The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants." Nat Prod Rep **16**(5): 565-74.

Rosenblum, J. S. and G. Blobel (1999). "Autoproteolysis in nucleoporin biogenesis." Proc Natl Acad Sci U S A **96**(20): 11370-5.

Rubio-Martinez, J., M. Pinto, et al. (2005). "Dock_dyn: a program for fast molecular docking using molecular dynamics information ".

Saarela, J., C. Oinonen, et al. (2004). "Autoproteolytic activation of human aspartylglucosaminidase." Biochem J **378**(Pt 2): 363-71.

Sacchettini, J. C. and C. D. Poulter (1997). "Creating isoprenoid diversity." Science **277**(5333): 1788-9.

Seemann, M., N. Campos, et al. (2002). "Isoprenoid biosynthesis via the methylerythritol phosphate pathway: accumulation of 2-C-methyl--erythritol 2,4-cyclodiphosphate in a gcpE deficient mutant of Escherichia coli." Tetrahedron Letters **43**(5): 775-778.

Sgraja, T., M. S. Alphey, et al. (2008). "Characterization of Aquifex aeolicus 4-diphosphocytidyl-2C-methyl-d-erythritol kinase - ligand recognition in a template for antimicrobial drug discovery." Febs J **275**(11): 2779-94.

Sgraja, T., T. K. Smith, et al. (2007). "Structure, substrate recognition and reactivity of Leishmania major mevalonate kinase." BMC Struct Biol **7**: 20.

Shao, Y., M. Q. Xu, et al. (1995). "Protein splicing: characterization of the aminosuccinimide residue at the carboxyl terminus of the excised intervening sequence." Biochemistry **34**(34): 10844-50.

Shahrokh, Z., G. Eberlein et al. (1994). "Major degradation products of basic fibroblast growth factor: detection of succinimide and iso-aspartate in place of aspartate" Pharm Res. 1994 Jul;11(7):936-44.

Shi, W., J. Feng, et al. (2007). "Biosynthesis of isoprenoids: characterization of a functionally active recombinant 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (IspD) from Mycobacterium tuberculosis H37Rv." J Biochem Mol Biol **40**(6): 911-20.

Shimizu, T., Y. Matsuoka, et al. (2005). "Biological significance of isoaspartate and its repair system." Biol Pharm Bull **28**(9): 1590-6.

Shrier, JA., R.A Kenley, et al. (1993). "Degradation pathways for recombinant human macrophage colony-stimulating factor in aqueous solution**"** Pharm Res. 1993 Jul;10(7):933-44.

Sprenger, G. A., U. Schorken, et al. (1997). "Identification of a thiamin-dependent synthase in *Escherichia coli* required for the formation of the 1-deoxy-D-xylulose 5-phosphate precursor to isoprenoids, thiamin, and pyridoxol." Proc Natl Acad Sci USA **94(24)**(Nov 25): 12857-12862.

Sprenger, G. A., Schorken, U., Wiegert, T., Grolle, S., de Graaf, A.A., Taylor, S.V., Begley, T.P., Bringer-Meyer, S., Sahm, H. (1997). "Identification of a thiamin-dependent synthase in Escherichia coli required for the formation of the 1-deoxy-D-xylulose 5-phosphate precursor to isoprenoids, thiamin, and pyridoxol." Proc Natl Acad Sci USA **94(24)**(Nov 25): 12857-12862.

Strassburger, E. (1988). Tratado de Botánica. Barcelona, Ediciones Omega.

Suresh, C. G., A. V. Pundle, et al. (1999). "Penicillin V acylase crystal structure reveals new Ntn-hydrolase family members." Nat Struct Biol **6**(5): 414-6.

Timson, D. J. and R. J. Reece (2003). "Functional analysis of disease-causing mutations in human galactokinase." Eur J Biochem **270**(8): 1767-74.

Timson, D. J., H. C. Ross, et al. (2002). "Gal3p and Gal1p interact with the transcriptional repressor Gal80p to form a complex of 1:1 stoichiometry." Biochem J **363**(Pt 3): 515-20.

Tinel, A., S. Janssens, et al. (2007). "Autoproteolysis of PIDD marks the bifurcation between pro-death caspase-2 and pro-survival NF-kappaB pathway." Embo J **26**(1): 197-208.

Titiz, O., M. Tambasco-Studart, et al. (2006). "PDX1 is essential for vitamin B6 biosynthesis, development and stress tolerance in Arabidopsis." Plant J **48**(6): 933-46.

Tyler-Cross, R. and V. Schirch (1991). "Effects of amino acid sequence, buffers, and ionic strength on the rate and mechanism of deamidation of asparagine residues in small peptides." J Biol Chem **266**(33): 22549-56.

Ursu, O., Diudea, M.V., NAKAYAMA, S. (2006). "3D Molecular similarity: Methods and algorithms." J Comput Chem **5**(1).

Utsumi, T., A. Levitan, et al. (1993). "Effects of truncation of human pro-tumor necrosis factor transmembrane domain on cellular targeting." J Biol Chem **268**(13): 9511-6.

van Poelje, P. D. and E. E. Snell (1990). "Pyruvoyl-dependent enzymes." Annu Rev Biochem **59**: 29-59.

Villanueva, J., O. Yanes, et al. (2003). "Identification of protein ligands in complex biological samples using intensity-fading MALDI-TOF mass spectrometry." Anal Chem **75**(14): 3385-95.

Violand, B. N., M. R. Schlittler, et al. (1990). "Formation of isoaspartate 99 in bovine and porcine somatotropins." J Protein Chem **9**(1): 109-17.

Wada, T., T. Kuzuyama, et al. (2003). "Crystal structure of 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase, an enzyme in the non-mevalonate pathway of isoprenoid synthesis." J Biol Chem **278**(32): 30022-7.

Wagner, S., A. Bernhardt, et al. (2006). "Analysis of the Arabidopsis rsr4-1/pdx1-3 mutant reveals the critical function of the PDX1 protein family in metabolism, development, and vitamin B6 biosynthesis." Plant Cell **18**(7): 1722-35.

Wang, S. and X. Q. Liu (1997). "Identification of an unusual intein in chloroplast ClpP protease of Chlamydomonas

eugametos." J Biol Chem **272**(18): 11869-73.
Wang, R., L. Lai, et al. (2002). "Further development and validation of empirical scoring functions for structure-based binding affinity prediction." J Comput Aided Mol Des 16(1): 11-26.
Wright, H. T. (1991). "Nonenzymatic deamidation of asparaginyl and glutaminyl residues in proteins." Crit Rev Biochem Mol Biol **26**(1): 1-52.
Xiang, S., G. Usunow, et al. (2007). "Crystal structure of 1-deoxy-D-xylulose 5-phosphate synthase, a crucial enzyme for isoprenoids biosynthesis." J Biol Chem **282**(4): 2676-82.
Xiong, H., B. A. Stanley, et al. (1997). "Processing of mammalian and plant S-adenosylmethionine decarboxylase proenzymes." J Biol Chem **272**(45): 28342-8.
Yanes, O., A. Nazabal, et al. (2006). "Detection of noncovalent complexes in biological samples by intensity fading and high-mass detection MALDI-TOF mass spectrometry." J Proteome Res **5**(10): 2711-9.
Yang, D., L. W. Shipman, et al. (2002). "Structure of the Methanococcus jannaschii mevalonate kinase, a member of the GHMP kinase superfamily." J Biol Chem **277**(11): 9462-7.
Zang, X., Z. Yu, et al. (1998). "Tight-binding streptavidin ligands from a cyclic peptide library." Bioorg Med Chem Lett **8**(17): 2327-32.
Zobel, K., L. Wang, et al. (2006). "Design, synthesis, and biological activity of a potent Smac mimetic that sensitizes cancer cells to apoptosis by antagonizing IAPs." ACS Chem Biol **1**(8): 525-33.

PUBLICACIONES

Imitando interacciones directas proteína-proteína e interacciones mediadas por solvente en el homodímero CDP-metileritritol quinasa: Una aproximación de búsqueda virtual dirigida a través de farmacóforo.

Victor Giménez-Oya[1*], Óscar Villacañas[2*$], Xavier Fernàndez-Busquets[3,4], Jaime Rubio-Martinez[2] , Santiago Imperial[1,4#]

[1] *Dept. de Bioquímica i Biología Molecular, Universitat de Barcelona (UB), Avda Diagonal 645, E-08028 Barcelona, Spain.*
[2] *Dept. de Química Física, Universitat de Barcelona (UB) and the Institut de Recerca en Química Teòrica i Computacional (IQTCUB), Martí i Franquès 1, E-08028 Barcelona, Spain.*
[3] *Biomolecular Interactions Team, Nanobioengineering Group, Institute for Bioengineering of Catalonia, Barcelona Science Park, Baldiri Reixac 10, E-08028 Barcelona.*
[4] *Nanoscience and Nanotechnology Institute, Universitat de Barcelona.*
[*] *Both Authors have contributed equally to the work.*
[$] *Current Address: Intelligent Pharma, S.L. Barcelona Science Park, C/ Baldiri Reixac 10, E-08028, Barcelona, Spain.*

Running title: Interacciones proteína-proteína en el homodímero CMK.

Palabras clave**:** CMK, MEP, MVA, interacciones mediadas por solvente, interacciones proteína-proteína, Dinámica molecular, Diseño de fármacos, Intensity-fading MALDI-TOF mass spectrometry.

Este estudio ha sido publicado en la revista "Journal of Molecular Modeling" 2009 Aug;15(8):997-1007, (ISSN: 1610-2940 (print version), ISSN: 0948-5023 (electronic version), Journal no. 894, Springer) cuya publicación se muestra en el anexo 2.

RESUMEN

La ruta del 2C-metilerithritol 4-fosfato (MEP) para la biosíntesis de isopentenil pirofosfato y su isómero dimetilalil pirofosfato, precursores de los isoprenoides, está presente en plantas, en el parásito causante de la malaria *Plasmodium falciparum* así como en muchas otras eubacterias, incluyendo agentes patógenos. Hongos y animales no tienen la ruta del MEP y sintetizan los precursores de los isoprenoides a través de la ruta del mevalonato (MVA). Dadas las características de la ruta del MEP, sus enzimas se convierten en dianas potenciales para la generación de nuevos agentes antibacterianos, antimaláricos y herbicidas altamente selectivos. En este trabajo nos hemos centrado en la enzima 4-(citidina 5'-difosfo)-2-C-metil-D-eritritol quinasa (CMK), que cataliza la cuarta reacción de la ruta del MEP. Se realizó un estudio de dinámica molecular para el complejo dimérico de la CMK y se estudiaron las interacciones proteína-proteína incluyendo las interacciones mediadas por solvente entre monómeros. Con el objetivo de encontrar pequeñas moléculas que interaccionen con la enzima CMK y sean capaces de impedir la formación del dímero, las interacciones observadas durante la dinámica molecular se utilizaron para

modelar el farmacóforo utilizado en las búsquedas en bases de datos de moléculas. La capacidad de interaccionar con la CMK de un compuesto se describió mediante la técnica de disminución de la intensidad MALDI-TOF de espectrometría de masas. La información presentada en este estudio muestra que la aproximación de búsqueda virtual puede ser usada para identificar moléculas candidatas para impedir el complejo CMK-CMK. La estrategia presentada aquí puede contribuir a aumentar la velocidad de descubrimiento de nuevos antimaláricos, antibacterianos y herbicidas.

Mimicking direct protein-protein and solvent-mediated interactions in the CDP-methylerythritol kinase homodimer : A pharmacophore-directed virtual screening approach.

Victor Giménez-Oya[1*], Óscar Villacañas[2*$], Xavier Fernàndez-Busquets[3,4], Jaime Rubio-Martinez[2] , Santiago Imperial[1,4#]

[1] *Dept. de Bioquímica i Biología Molecular, Universitat de Barcelona (UB), Avda Diagonal 645, E-08028 Barcelona, Spain.*
[2] *Dept. de Química Física, Universitat de Barcelona (UB) and the Institut de Recerca en Química Teòrica i Computacional (IQTCUB), Martí i Franquès 1, E-08028 Barcelona, Spain.*
[3] *Biomolecular Interactions Team, Nanobioengineering Group, Institute for Bioengineering of Catalonia, Barcelona Science Park, Baldiri Reixac 10, E-08028 Barcelona.*
[4] *Nanoscience and Nanotechnology Institute, Universitat de Barcelona.*
[*] *Both Authors have contributed equally to the work.*
[$] *Current Address: Intelligent Pharma, S.L. Barcelona Science Park, C/ Baldiri Reixac 10, E-08028, Barcelona, Spain.*

Running title: Protein-protein interactions in the CMK homodimer

Key words**:** CMK, MEP, MVA, Solvent-mediated interactions, Protein-protein interactions, Molecular dynamics, Drug design, Intensity-fading MALDI-TOF mass spectrometry.

ABSTRACT

The 2C-methylerythritol 4-phosphate (MEP) pathway for the biosynthesis of isopentenyl pyrophosphate and its isomer dimethylallyl pyrophosphate, which are the precursors of isoprenoids, is present in plants, in the malaria parasite *Plasmodium falciparum* and in most eubacteria, including pathogenic agents. However, the MEP pathway is absent from fungi and animals, which have exclusively the mevalonic acid pathway. Given the characteristics of the MEP pathway, its enzymes represent potential targets for the generation of selective antibacterial, antimalarial and herbicidal molecules. We have focused on the enzyme 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase (CMK), which catalyses the fourth reaction step of the MEP pathway. A molecular dynamics simulation was carried out on the CMK dimer complex, and protein-protein interactions analysed, considering also water-mediated interactions between monomers. In order to find small molecules that bind to CMK and disrupt dimer formation, interactions observed in the dynamics trajectory were used to model a pharmacophore used in database searches. Using an intensity-fading matrix assisted laser desorption/ionization time-of-flight mass spectrometry approach, one compound was found to interact with CMK. The data presented here indicate that a virtual

screening approach can be used to identify candidate molecules that disrupt the CMK-CMK complex. This strategy can contribute to speeding up the discovery of new antimalarial, antibacterial, and herbicidal compounds.

INTRODUCTION

Isopentenyl pyrophosphate and its isomer dimethylallyl pyrophosphate (DMAPP) [1,2] are the universal five-carbon precursors of one of the largest family of natural products, isoprenoids. Isoprenoids include hopane triterpenes, ubiquinones and menaquinones in bacteria; carotenoids, plastoquinones, mono-, sesqui-, di and tri- terpenes and the prenyl side chains of chlorophylls in plants; and quinones, dolichols, steroids and retinoids in mammals [3] and account for more than 30000 naturally occurring molecules of both primary and secondary metabolism.

After the discovery of the mevalonic acid (MVA) pathway in yeast and animals, it was assumed that IPP was synthesised from acetyl-CoA via MVA and then isomerized to DMAPP in all organisms [2-5]. However, an alternative MVA-independent route for the biosynthesis of IPP was recently identified by labeling experiments in bacteria [6,7] and plants [8]. It is named MEP pathway, after what is currently considered its first committed precursor, 2C-methylerythritol 4-phosphate (for recent reviews see [9] and [10]).

Experimental evidence has shown that most organisms only use one of the two pathways for the biosynthesis of their isoprenoid precursors. Thus, the MEP pathway is essential in the malaria parasite *Plasmodium falciparum* and in most eubacteria, including the causal agents for diverse and serious human diseases like leprosy, bacterial meningitis, various gastrointestinal and sexually transmitted infections, tuberculosis, and certain types of pneumonia. The MEP pathway is absent in archaeobacteria, fungi and animals, which synthesise their isoprenoids exclusively through the operation of the MVA pathway, whereas plants use both pathways for isoprenoid biosynthesis, although they are localised in different compartments [9-16].

Given the absence of the MEP pathway in mammals, its enzymes represent potential targets for the development of selective antibacterial, antimalarial and herbicidal compounds[15-18]. An example of this new class of chemotherapeutic agents is fosmidomycin, a natural antibiotic originally isolated from *Streptomyces lavendulae,* which inhibits the second enzyme of the MEP pathway, deoxyxylulose 5-phosphate reductoisomerase [16]. Fosmidomycin and its derivatives have been shown to inhibit bacterial growth [19] and to be effective against *Plasmodium* infections [20-22]. A combined therapy of fosmidomycin and clyndamicin, a chemical that inhibits the prokaryote-like translation machinery of the *P. falciparum* apicoplast, proved to be highly efficient against malaria, with a 3-day regime resulting in an overall cure rate of 95% [23]. These results are very promising and validate the usefulness of MEP pathway enzymes as drug targets.

In this work we have studied the enzyme 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase (CMK), which catalyses the fourth step in the MEP pathway, i.e. the ATP-dependent phosphorylation of the intermediate CDP-ME at the C-2 hydroxyl group to yield CDP-MEP. The enzyme from *Escherichia coli* has been cloned, overexpressed, purified, and crystallised. X-ray structure studies showed that CMK is organised as a homodimer. The asymmetric unit consists of two subunits, A and B, of approximate total

mass 62 kDa, which assemble with C2 symmetry to form an extended homodimer, but only 4% of the total surface area of the two monomers is involved in dimer formation. Despite such a small area of interaction, a dimer is observed in gel filtration experiments that can be detected in matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) [24] and in analytical ultracentrifugation assays [25].

Disruption of protein-protein interactions through small molecules that target *hotspots* in the contact regions between protein surfaces [26] is an attractive therapeutic approach that can be very useful for drug design [27-30]. Thus, destabilising the CMK complex could represent a new strategy for the development of antibacterial and antimalarial agents. Although few direct protein-protein interactions are observed in the crystal structure of CMK, several water-mediated interactions seem be important for dimer stabilisation. Solvent-mediated protein-protein interactions play an important role in molecular recognition [31-35] and are being increasingly considered when modeling protein complexes [36]. Here, we have carried out molecular dynamics (MD) simulations on the CMK dimer in order to give insight into such protein-protein interactions. With the objective of studying the intermolecular interactions of the complex in more detail, both direct and solvent-mediated protein-protein interactions were monitored throughout the dynamics trajectory. A pharmacophore was designed based on the position of atoms involved in the interactions, and used in the database search for small molecules capable of disrupting the CMK dimer. The association between candidate compounds and CMK was empirically checked by intensity-fading MALDI-TOF MS [37,38]. This approach provides a quick method for the detection of complex formation between the protein and a ligand based on the reduction of ion intensities of the mixture, compared to the intensity of both protein and ligand alone. One hit provided by the pharmacophore-guided database search turned out to decrease dimer formation.

MATERIALS AND METHODS

System design

All minimizations and MD simulations were done with the AMBER 7 suite of programs [39] and all-atom *parm99* and *gaff* force fields[40,41]. For organic small molecules, atom types were assigned with the antechamber module of AMBER7 and atomic charges were calculated following the BCC-AM1 method [42]. Starting coordinates of the CMK complex were downloaded from the Protein Data Bank (PDB) [24], identification number PDBid = 1oj4. The complex consists of two CMK proteins (each monomer comprising the full sequence, from residue 1 to residue 283), two molecules of adenosine 5′-(β, γ-imido)triphosphate (AMP-PNP), two molecules of 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol (CDP-ME), 192 water molecules and 1 chloride ion. The structure was resolved by X-Ray diffraction at a resolution of 2.01 Å. Some side-chains presented two possible conformations. Because they were not involved in protein-protein close contacts, only the conformations marked in the PDB file as 'A' were considered for further purposes. Selenomethionine residues were replaced by methionines. Hydrogen atoms were placed where necessary with the LEaP module of AMBER7. The system was neutralized with Na^+ ions (31 ion atoms were required) placed at the most negative electrostatic potential points following a grid-based box-shaped procedure for mapping electrostatic potential surface (as implemented in LEaP).

Molecular dynamics simulations

A first minimization step was carried out *in vacuo* with a distance-dependent dielectric constant of *4r*. Then, a rectangular box of TIP3P water [43] was added, with a minimum distance between the protein and the edges of the box along the Cartesian axis of 10 Å. Water molecules closer than 1.8 Å to protein/CDP-ME/AMP-PNP were removed. The system was then relaxed by a two-step procedure. First, water molecule positions were optimised and, second, the whole system was optimised. MD simulations were performed using the particle mesh Ewald (PME) summation method [44] to compute long-range electrostatic energy. Bonds involving hydrogen atoms were constrained with the SHAKE algorithm [45]. The time step was set to 1 fs and the non-bonded pair list was updated every 25 steps, applying a *cutoff* value of 9 Å. Temperature and pressure regulation was simulated using Berendsen's algorithm[46], with time coupling constants of 1.0 and 0.2 ps respectively, unless specified otherwise. The system was heated gradually to 300 ºK increasing 30 ºK each 10 ps, at constant volume. Then, a two-step pressure equilibration process followed. A temperature coupling constant of 1.0 ps was applied during the first 20 ps and increased to 2.0 ps for a further 40ps. Up to this point, ions and protein atoms had been restrained under harmonic potentials with a force constant values of 1 and 5 kcal mol^{-1} $Å^{-1}$ respectively. Next, restrains were reduced gradually for 60 ps. Finally, simulations were carried out in the NVT ensemble with no restrains. The last 1000 ps with stable total energy were considered as the production run for further analysis.

Energy analysis

Lennard-Jones (LJ) and Coulombic energies between every side chain of the first monomer protein and

the entire second monomer protein, and viceversa, were computed for snapshots every picosecond throughout the 1000 ps of production time with the ANAL module of AMBER 7. These calculations were also performed after adding water molecules to one of the interacting partners.

Solvent-mediated interaction analysis

Water molecules that established hydrogen bonds simultaneously with both proteins throughout the last 200 ps were analysed with a program developed by our group. Hydrogen bond distances were monitored and occupancy was defined as the percentage of time for which there existed at least one hydrogen bond to each monomer with a distance value lower than 2.5 Å.

Design of interaction models from MD trajectory

During the last 200 ps of production time, the positions of those atoms in one monomer that effectively interacted with its partner monomer were monitored. If several atoms contributed to the same effective interaction, the centre of masses of the mean positions was calculated and it was treated as a single interaction point. Maximum and minimum distances between every pair of points were used to construct the model for database searches and mean positions of points as Cartesian coordinates were used for the docking procedure. Because all points must be referred to the same reference system, each snapshot was superimposed onto the backbone heavy atoms. An interaction feature was assigned to every point: hydrogen donor/acceptor, hydrophobic.

Database search and docking procedure

A selected interaction model was introduced as input for the compound search. Catalyst (Accelrys, Inc., USA) software together with Available Chemical Database (ACD), National Cancer Institute (NCI) database, ChemDiv database and Specs database were used. Matches were filtered and molecules with unusual bonds or with a molecular mass higher than 750 Da were not selected. Hits were saved as *mol2* format files containing all the conformations given by Catalyst. Each conformation was docked to the CMK monomer as follows. Pharmacophoric features were automatically assigned to the ligands, and distances between pharmacophoric points were calculated for all conformations. Next, for each conformation, we tested whether a combination that satisfied the interaction model existed and, in such cases, the ligand pharmacophoric points were superimposed to those of the interaction model, optimising the root mean square deviation (RMSD). At this point, the CMK monomer structure was introduced in the model to detect van der Waals (vdW) clashes between the ligand and the protein. In order to avoid bad vdW contacts, a slight translational movement of the ligand was allowed. Atoms were not allowed to be closer than half of the sum of their vdW radii. If there were no forbidden vdW contacts, the dissociation constant was evaluated with X-Score (Wang, Lai et al. 2002).

A visual structure analysis was carried out on ligands with best docking, defined as those with best X-Score and RMSD consensus values. Selected compounds were finally classified according to their structure.

Preparation of recombinant CMK

Recombinant CMK from *E. coli* was prepared as previously described[47]. Briefly, M15[pREP4] *E. coli* cells were transformed with the pQE30-*ispE* construct containing the coding region of the the *ispE* gene of *E.coli*. Transformed cells were grown at 37 °C in LB medium supplemented with ampicillin (100 μg/mL) and kanamycin (50 μg/mL) until an OD_{600} of 0.3–0.4 was reached. Induction was then performed with 0.4mM IPTG for 3 h at 30 °C. Cells were harvested by centrifugation (3000 x *g*, 10 min) and resuspended in 40mM Tris–HCl buffer, pH 8, containing 0.5mM EDTA, lysozyme (1mg/mL) and complete-mini EDTA-free (Roche) (one tablet/10mL buffer). Cells were incubated at 4 °C for 20min and sonicated (5 x 30s bursts with 1min cooling at 0ºC). The extract was centrifuged at 11,000 x *g* for 30 min at 4 °C; the supernatant was recovered and supplemented with protamine sulphate (1.25mg/mL), incubated at room temperature for 15 min, and then centrifuged at 13,000 x *g* for 20 min at 4 °C.
Hi-trap chelating columns (GE Healthcare, Little Chalfont, UK) were used for the purification of recombinant CMK. The resin was washed with 40mM Tris–HCl buffer, pH 8.0, containing 0.1M NaCl, and 10mM imidazole. His6-CMK was eluted with an imidazole gradient (10–500mM) in 40mM Tris–HCl buffer, pH 8.0, containing 0.1M NaCl. The fractions containing His6-CMK were 95% pure as judged from SDS–polyacrylamide gel electrophoresis and Coomassie brilliant blue R-250 staining and appeared as a major band corresponding to an apparent molecular mass of 31kDa. CMK preparations were desalted by gel filtration on disposable Sephadex G-25 columns (PD10) (GE Healthcare) equilibrated in 40mM Tris–HCl buffer pH 8.0 containing 0.1M NaCl, distributed into aliquots, snap-frozen with liquid N_2 and stored at -80ºC until use. Protein concentration was measured by the method of Bradford [48] using bovine serum albumin as a standard.

Intensity-fading MALDI-TOF MS

Intensity-fading MALDI-TOF MS was used to study the formation of a complex between one of the selected compounds in the database search, 2,4,6-trichloro-N-[5,6-dimethyl-1-(2-methylbenzyl)-1H-1,3-benzimidazol-4-yl]benzenesulfonamide (Bionet Research, Key Organics, Camelford, UK), Compound 1 (C1), and purified CMK from *E. coli*. In these experiments a direct protocol could be used because both C1 and CMK were soluble in the concentration range used. This protocol requires also that the complex remains intact in solution at least in part during the process of MALDI-TOF MS determination [37]. Samples were prepared as follows: a 1mM C1 solution prepared in water was mixed (1:1, v:v) with different dilutions in the purification buffer of recombinant CMK ranging from 0.002 mM to 0.2mM. Then, the samples were mixed (1:1, v:v) with a 10mg/mL solution of α-ciano-4-hydroxycinnamic acid (CHCA)(Aldrich; Sigma-Aldrich, St. Louis, MO) matrix solution in acetonitrile/deionised water (1:1, v:v). One microlitre of the mixture was overlaid onto the MALDI-TOF plate and dried using the dried droplet method [49]. MALDI-TOF mass spectra were recorded in a 4700 Proteomics Analyzer instrument (Applied Biosystems, Foster City, CA). Acquisition of mass spectra was performed in the MS reflector positive-ion mode. Typical parameters were set to source and grid voltages 20 and 14kV, respectively, power laser from 5200 to 5800, signal/noise threshold 5, and a noise windows width of 50.

Drawings

Figures depicting complex structure were constructed with Chimera [50] or VMD [51].

RESULTS AND DISCUSSION

Interactions deduced from X-ray crystal structure

The X-ray crystal structure of CMK [24] shows a homodimeric complex, where the two monomers are apparently positioned in an inverted position. The protein monomers interact with one another through two small interfaces: **A** and **B** (see figure 1), in which similar interactions are established. Residues of monomer 1 (chain A in PDB original file) found in interface **A** are the same as those of monomer 2 (chain B in PDB original file) in interface **B**, and vice versa. In the experimental structure of the complex, three direct protein-protein electrostatic interactions are detected. Two of them are observed in interface **A**, an intermolecular salt bridge between Arg^{21} and Asp^{80} (hereinafter, monomer 1 and monomer 2, respectively) and a hydrogen bond between Ala^{22}(carbonyl) and Gly^{87}(amine). While the former is also observed in interface **B** (the third protein-protein electrostatic interaction), the latter interaction is detected in interface **B** not as a direct protein-protein interaction but as a solvent-mediated interaction (Table 1), although with very distorted hydrogen bond angles between water molecule and protein atoms. A more relaxed geometry is found in the solvent-mediated interaction between Ser^{88}(carbonyl) and Ala^{22}(carbonyl) (Table 1). Table 1 lists all solvent-mediated hydrogen bonding networks between monomers.

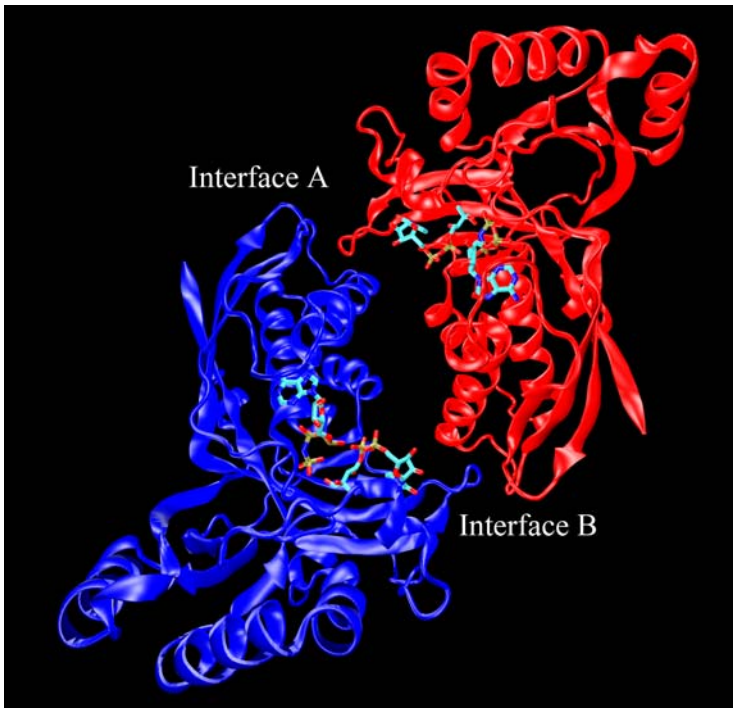


Fig. 1 X-ray structure of *E. coli* 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase **(**CMK) dimer. Monomers are represented in the cartoon mode and adenosine 5'-(β, γ-imido) triphosphate (AMP-PNP) and 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol (CDP-ME) molecules in the licorice mode as implemented in VMD visualization software (Humphrey, Dalke et al. 1996). According to the reference used in the text, first monomer in PDB is colored in blue and the second in red.

Residues of monomer 1	Residues of monomer 2	Distance 1 (Å)	Angle 1 (degrees)	Distance 2 (Å)	Angle 2 (degrees)	Interface
Ala^{22} (amine)	**Thr^{86} (side chain)**	3.3	83/87	3.4	117	A
Ala^{22} (carbonyl)	**Ser^{88} (carbonyl)**	3.2	124	2.7	138	A
Gly^{49} (carbonyl)	**Ala^{22} (carbonyl)**	2.8	121	3.1	131	B
Lys^{76} (side chain)	**Leu^{136} (carbonyl)**	2.9	103	3.0	157	B
Asp^{80} (side chain)	**Glu^{158} (side chain)**	3.5	117	3.4	116	B
Ser^{88} (amine)	**Ala^{22} (carbonyl)**	3.2	93/102	2.1	141	B
Ser^{88} (carbonyl)	**Ala^{22} (carbonyl)**	2.8	124	2.1	141	B
Gly^{87} (amine)	**Ala^{22} (carbonyl)**	3.6	131/88	2.1	141	B
Gly^{87} (amine)	**Ala^{22} (carbonyl)**	2.8	118/119	3.7	88	B

Table 1 Distances and angles of hydrogen bonds detected in the crystal structure of 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase (CMK). Distance 1 and angle 1 are calculated between water and residue of the monomer 1, Distance 2 and angle 2 are calculated between water and residue of monomer 2. When a protein amino group is involved, both carbon-bonded atoms were used to calculate angles (carbon of the N-terminal side / carbon of C-terminal side).

Interactions from dynamics trajectory

Hydrogen bonds and salt bridges

The scenario obtained throughout the MD production time was slightly different from that suggested by the crystal structure. Our analysis revealed only one direct monomer-monomer electrostatic interaction, a salt bridge between Arg^{21} and Asp^{80} in interaction interface **A** (mean minimum hydrogen bond distance was 1.81 Å, with an root mean square (RMS) of 0.12 Å). Although the interaction in the other contact region (interaction interface **B**) was not of protein-protein type, a more detailed study revealed that a water bridge joined both side chains through hydrogen bonds (Fig.2). The hydrogen bond distances were analysed at every snapshot extracted from the last 200 ps of the production time, and the respective percentages of time with values below 2.5 Å (occupancy) are noted in Table 2. The analysis showed that a water bridge (always the same water molecule) between Asp^{80} and Arg^{21} that established hydrogen bonds in 96% and 99% of the time, respectively. As far as the interaction between Ala^{22} and Gly^{87} backbones is concerned (detected in crystal structure), it was observed as water-mediated in both interfaces. In interface **A**, a water molecule acted as the center of a hydrogen bond network and established favourable interactions with Ala^{22}(carbonyl) of monomer 1 and the backbones of Gly^{49}, Pro^{85}, Gly^{87}, and Ser^{88} of monomer 2. The occupancies of hydrogen bond distances were 83% and 97%, respectively (Table 2). In this latter case, the water molecule involved was labile and, therefore, different molecules were observed to play the role of mediating the interaction. However, the equivalent

interaction in the other interface (**B**) was mediated by a water bridge, with respective hydrogen bond occupancy values of 65% and 99% (Table 2).

Summarising, the intermolecular protein-protein hydrogen bonds elicited from the crystal structure were also observed in the dynamics trajectory but in a different structural pattern. Most direct protein-protein interactions in the crystal were observed in the dynamics trajectory as mediated by solvent through hydrogen bond networks. The role of water molecules in mediating interactions between both monomers is also observed in the crystal structure. The CMK homodimer complex appears to overcome its low protein-protein interaction surface by water-mediated interactions. Further interactions of this kind were resolved from the trajectory and will be detailed in the following section.

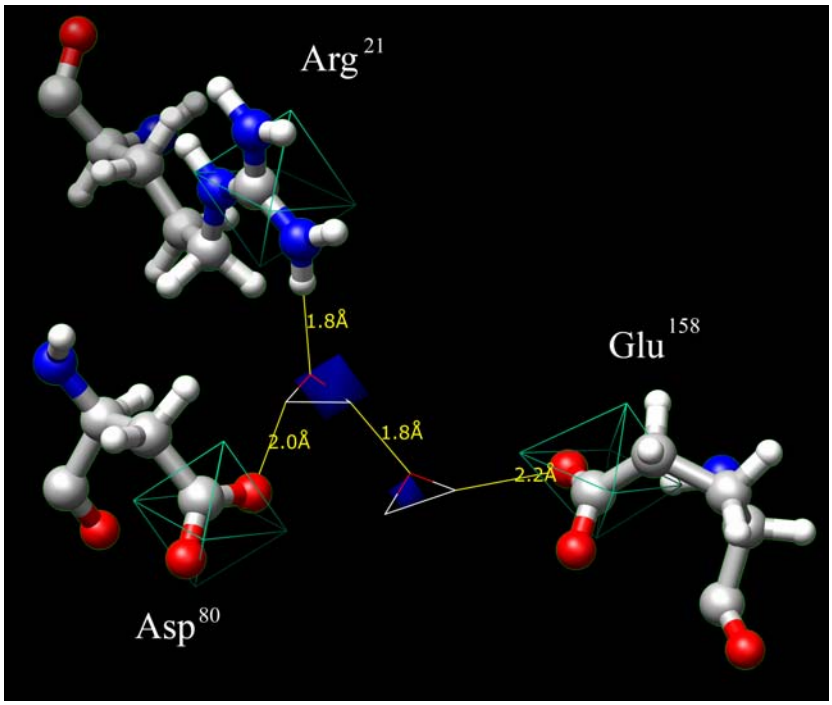


Fig. 2 Water-mediated interactions between Asp^{80} of monomer 1 and Arg^{21} and Glu^{158} of monomer 2 in interface B. Occupational analysis was carried out with Chimera software [50]. Grid volumes in which an oxygen atom of a water molecule is found at least in 70% (approximately) of the time throughout the last 200 ps of the production time is colored in solid blue. Carbon atoms of carboxylic (Asp^{80} and Glu^{158}) and guanidinium (Arg^{21}) groups and the oxygen atom of water bridge between Asp^{80} and Arg^{21} were used to superimpose snapshots before analysis. Cutoff and grid spacing values were set to 7 and 1.8 Å, respectively. For validation, volume that comprises carbon atoms mentioned above in 100% of the same time is outlined in light blue. As an exemple, the first snapshot is represented in the figure.

Residues of monomer 1	Residues of monomer 2	% HB 1	% HB 2	Interface
Ala^{22}(amine)	**Ala^{79}(carbonyl)**	19	94	A
Ala^{22}(carbonyl)	**Pro^{85}(carbonyl), Gly^{87}(amine), Ser^{88}(carbonyl/amine) and Gly^{49}(carbonyl)**	83	97	A
Asp^{23}(side chain)	**Lys^{76}(side chain)**	98	97	A
Asp^{23}(carbonyl)	**Gly^{49}(carbonyl), Arg ^{72}(side chain)**	94	99	A
Glu^{158}(side chain)	**Lys^{76}(side chain), Asp^{80}(carbonyl/side chain)**	99	99	A
Pro^{85}(carbonyl), Thr^{86}(carbonyl), Ser^{88}(amine)	**Ala^{22}(carbonyl)**	65	99	B
Asp^{80}(side chain)	**Arg^{21}(side chain)**	96	99	B
Ala^{79}(carbonyl)	**Ala^{22}(amine)**	62	97	B

Table 2 Percentage of hydrogen bond (HB) distances under 2.5 Å in protein-protein solvent-mediated interactions throughout the last 200 ps of the production time. % HB 1 refers to hydrogen bond between water and monomer 1; % HB 2 refers to hydrogen bond between water and monomer 2.

Protein-protein solvent-mediated interactions

In addition to those described above, other stable water-mediated protein-protein interactions were found to be stable in the last 200 ps of production time (Table 2). In the interaction interface **A**, three further water-mediated interactions were elucidated, two of which involving both the side chain and the carbonyl group of Asp^{23} in monomer 1. The side chain established a solvent-mediated interaction with the Lys^{76}(side chain) of monomer 2, whereas the carbonyl group did the same with Gly^{49}(carbonyl) and Arg^{72}(side chain) of monomer 2. Finally, the side chain of Glu^{158} in monomer 1 interacted with Lys^{76}(side chain) and Asp^{80}(side chain/carbonyl) of monomer 2 through hydrogen bonds established with water molecules. In the interface **B**, one additional solvent-mediated interaction was observed between Ala^{79}(carbonyl) and Ala^{22}(amine). Similarities and differences between the crystal structure and the dynamics trajectory can be deduced by comparing values from Table 1 and Table 2. The addition of temperature to the system changed the water-mediated monomer-monomer interaction pattern.

The interaction involving Glu^{158} and Asp^{80}(side chain)/Lys^{76}(side chain) in interface **A** was not detected in interface **B**. However, some evidence led us to carry out a spatial occupancy analysis on the oxygen atoms of the water molecules (Chimera software [50]). The results revealed that this interaction was described as mediated by two water molecules (Fig. 2), which most of the time filled optimal sites for establishing a hydrogen bond network between the residues. There was also evidence of a further example of a double water-mediated interaction pattern between Ala^{22}(amine) and Ala^{79}(carbonyl) in interface **A**. However, occupational analysis did not show a clear picture of two oxygen atoms located in optimal positions (data not shown). In this last case, a single water-mediated interaction pattern was

observed in 19% of the time. It is likely that both patterns exchanged with one another several times throughout the trajectory.

On the whole, the results extracted from dynamics analysis show the importance of water-mediated intermolecular interactions in the CMK dimer complex. It has been argued that bound water should be considered as an integral part of the tertiary structure of proteins (Levy and Onuchic 2006), a proof of which is the conservation of buried water molecules in homologous proteins (Sreenivasan and Axelsen 1992). Here, in the CMK complex, most water molecules that were involved in intermolecular solvent-mediated interactions were highly labile. Because of the small monomer interaction surfaces, these water molecules are highly exposed, that is, they are not buried in protein cavities, and have high exchange rates. The existence of the hydrogen bond network was detected in most of the 200 ps time, suggesting that the role of water molecules is crucial for binding and recognition, although they are not tightly bound to the proteins. The computation of the protein-protein interaction surface throughout 20 snapshots extracted from the production time (carried out with MolSurf as implemented in AMBER7), gave a value of 355 $Å^2$. When considering water molecules involved in solvent-mediated interactions as an integral part of monomers the surface increased to 422 $Å^2$, a 19% increase. Thus, taking into account water molecules significantly raised the interaction surface.

Van der Waals interactions

Van der Waals (vdW) interactions were analyzed in terms of energy, as described in *Materials and Methods*. Interaction vdW energy between every side chain of one monomer and the entire counterpart protein is depicted in Fig. 3. The results revealed two main groups of amino acid residues with favorable interaction energies; the first group is composed of Arg^{21} (with a mean Lennard-Jones (LJ) energy value, considering both cases, of -4.7 kcal/mol), Ala^{22} (-2.6 kcal/mol), Asp^{23} (-4.7 kcal/mol), and Tyr^{25} (-1.4 kcal/mol). The second group was formed by Met^{75} (-1.9 kcal/mol), Lys^{76} (-3.0 kcal/mol), Ala^{79} (-2.3 kcal/mol), Leu^{84} (-2.4 kcal/mol) and Thr^{86} (-1.2 kcal/mol). In fact, these two groups interacted with one another when considered in different monomers.

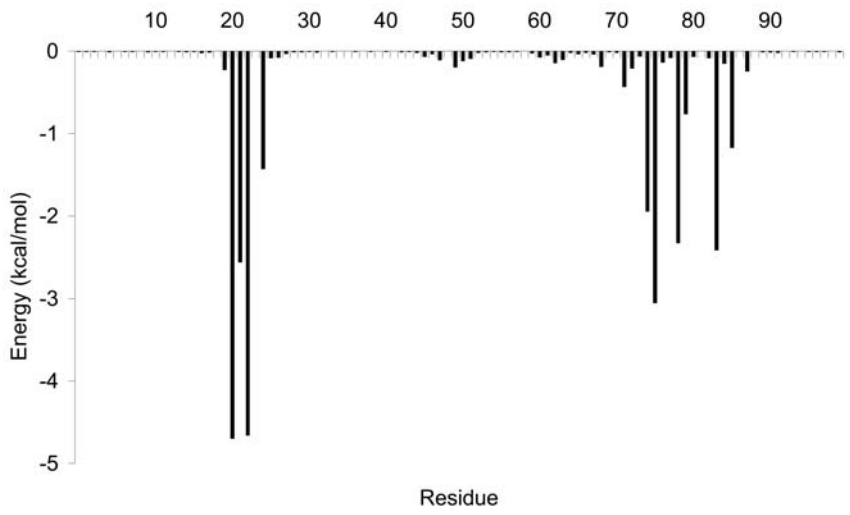


Fig. 3 VdW (vdW) interaction energy between every side chain of one monomer and the entire partner protein. Mean values accounting for the same residue in both monomers are represented.

Electrostatic interactions

Protein-protein electrostatic interaction energies throughout the production time were also analysed (Fig. 4a). Arginine and lysine residues had negative favourable values, while aspartic acid and glutamic acid residues had complex destabilising values. A single CMK monomer has 26 positively charged residues and 36 negatively charged residues. Therefore, one monomer has a high net negative charge that is translated into an unfavorable energy when analysing the electrostatic contribution of glutamic and aspartic acid residues. The net electrostatic interaction energy between monomers had a mean value of -27 kcal/mol (100 snapshots were analyzed), but it was strongly favored by the presence of intervening water molecules- the mean electrostatic interaction energy was computed to be -5537 kcal/mol when water molecules were incorporated in the calculation. The inclusion of ions or/and CDP-ME/AMP-PNP did not change the effect of the water molecules (data not shown). In order to see the effect of the solvent on residue electrostatic contribution, water molecules were included in a further energy computation. As seen in Fig. 4b, the electrostatic energies of glutamic acid and aspartic acid residues decreased to negative values. Thus, water molecules are essential to explain how protein-protein electrostatic interactions favour complex formation.

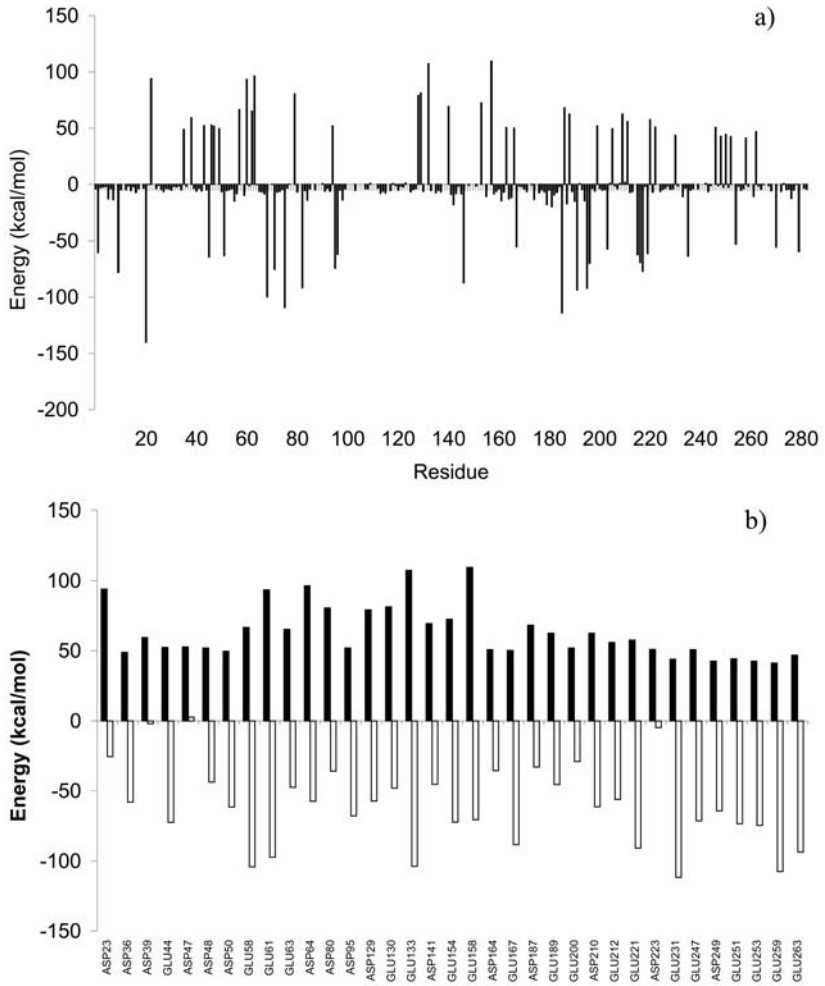


Fig. 4 a) Coulombic interaction energy between every side chain of one monomer and the entire partner protein. b) White bars: Same as a) for glutamic acid and aspartic acid residues but including water molecules in the partner monomer part. Black bars: same as a) for comparison. Mean values accounting for same residue in both monomers are represented.

Database search for compounds disrupting protein-protein interaction

Atoms involved in protein-protein interactions were monitored throughout the last 200 ps of the dynamics trajectory to extract interaction models, as described in *Materials and Methods*. One of those models, depicted in Fig. 5, was introduced as a hypothesis in Catalyst (Accelrys, Inc. USA). It is worth mentioning that one of the points of the model was deduced from a water molecule involved in a solvent-mediated interaction between monomers (between Asp^{23} and Lys^{76} in interface **A**). Points marked *vdW* were modified to *hydrophobic* when defining the pharmacophoric hypothesis in Catalyst. The search procedure was carried out in Available Chemical Database (ACD; 214 hits), Specs (512 hits) and ChemDiv databases (835 hits).

Hits were docked on CMK protein, as described in *Materials and Methods*, and poses were evaluated with the scoring function X-Score [40].

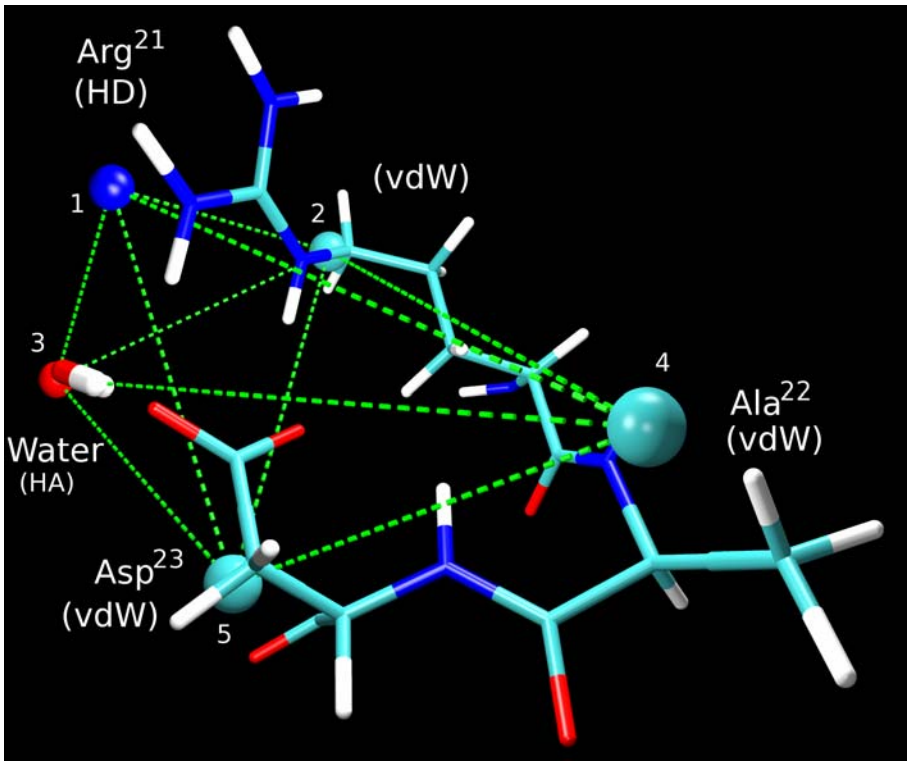


Fig. 5 Pharmacophoric points that form the interaction model used in molecule search. Points are derived from residues 21-23 of monomer 2 and a water molecule. Arg^{21} provided a hydrogen donor (HD) and a vdW pharmacophoric point; Ala^{22} and Asp^{23} provided vdW pharmacophoric points; the water molecule provided a hydrogen acceptor point. Distances between pair of points are: 1_2: 2.644 - 3.099 Å ; 1_3: 3.004 - 4.837 Å; 1_4: 6.965 - 8.907 Å; 1_5: 3.982 - 4.795 Å; 2_3: 3.737 - 5.753 Å; 2_4: 5.958 - 7.235 Å; 2_5: 3.755 - 4.644 Å; 3_4: 8.715 - 11.321 Å; 3_5: 3.111 - 5.839 Å; 4_5: 5.354 - 6.696 Å.

Final selection of compounds

A visual structure analysis was carried out on the 15 ligands with best docking from each database. Ranking was based on X-Score consensus values and the final selection took into account the compound structure.

Intensity-fading MALDI-TOF MS

Intensity-fading MALDI-TOF MS is a rapid and sensitive approach for detecting the formation of soluble complexes between proteins and small molecules [38]. This methodology has the advantages of a low cost and rapid performance together with the capability to check affinity properties of the analysed compounds. In this approach, detection of the complex is based on a reduction of the molecular ion intensities of the partners in a MALDI mass spectrum compared to the mass spectra of the partners alone.

In our study, we tested the complex formation between one of the final selected compounds (C1) and CMK. Docking of C1 onto CMK is shown in Fig. 6. RMSD after docking between pharmacophoric features on compound and pharmacophore from Fig. 5 was 1.65 Å, with a pK_d estimated value (X-Score) of 5.43. Because the complex is formed between a small molecule and an enzyme, the reduction of the small-molecule ion intensity can be associated directly to the formation of the complex.

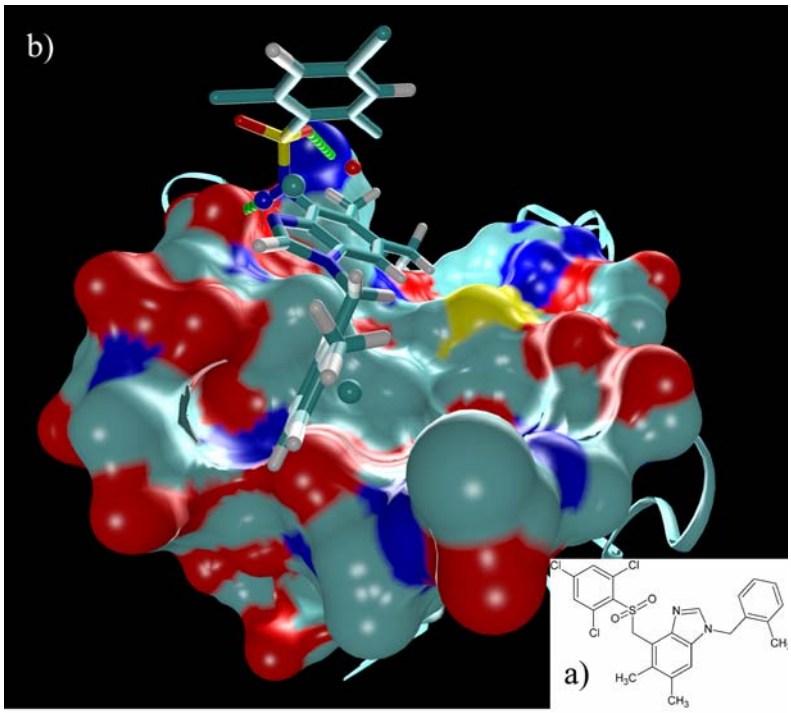


Fig. 6 a) 2-D structure of compound 1. b) Docking of C1 on CMK surface. CMK atoms closer than 5 Å were represented as a molecular surface using MSMS (Humphrey, Dalke et al. 1996). The rest of protein atoms were depicted following a ribbon pattern. Pharmacophoric points are represented as spheres.

The mass spectrum for C1 was acquired by adding 1µL of 1mM C1 to 1µL of CHCA matrix and by depositing 1µL of the mixture onto a MALDI-TOF plate. The MALDI-TOF mass spectrum of the C1+CHCA matrix showed two major peaks at m/z 508 and 510. The m/z 508 peak correlates with C1 molecular mass. In the subsequent series of experiments, 1µL of 1mM C1 was mixed with different concentrations of CMK and the corresponding MALDI-TOF mass spectrum was acquired. As shown in Fig. 7, all C1+CMK+CHCA mass spectra obtained showed C1 peaks at m/z 508 and 510. The intensities of these peaks showed a reduction (fading) in the presence of CMK that was CMK concentration-dependent. At the highest concentration assayed (0.2 mM, figure 7, top), the intensity-fading of the m/z 508 and 510 peaks was almost complete, making the peak intensities similar to the baseline noise of the spectrum. These results demonstrate the formation of a complex between CMK and C1.

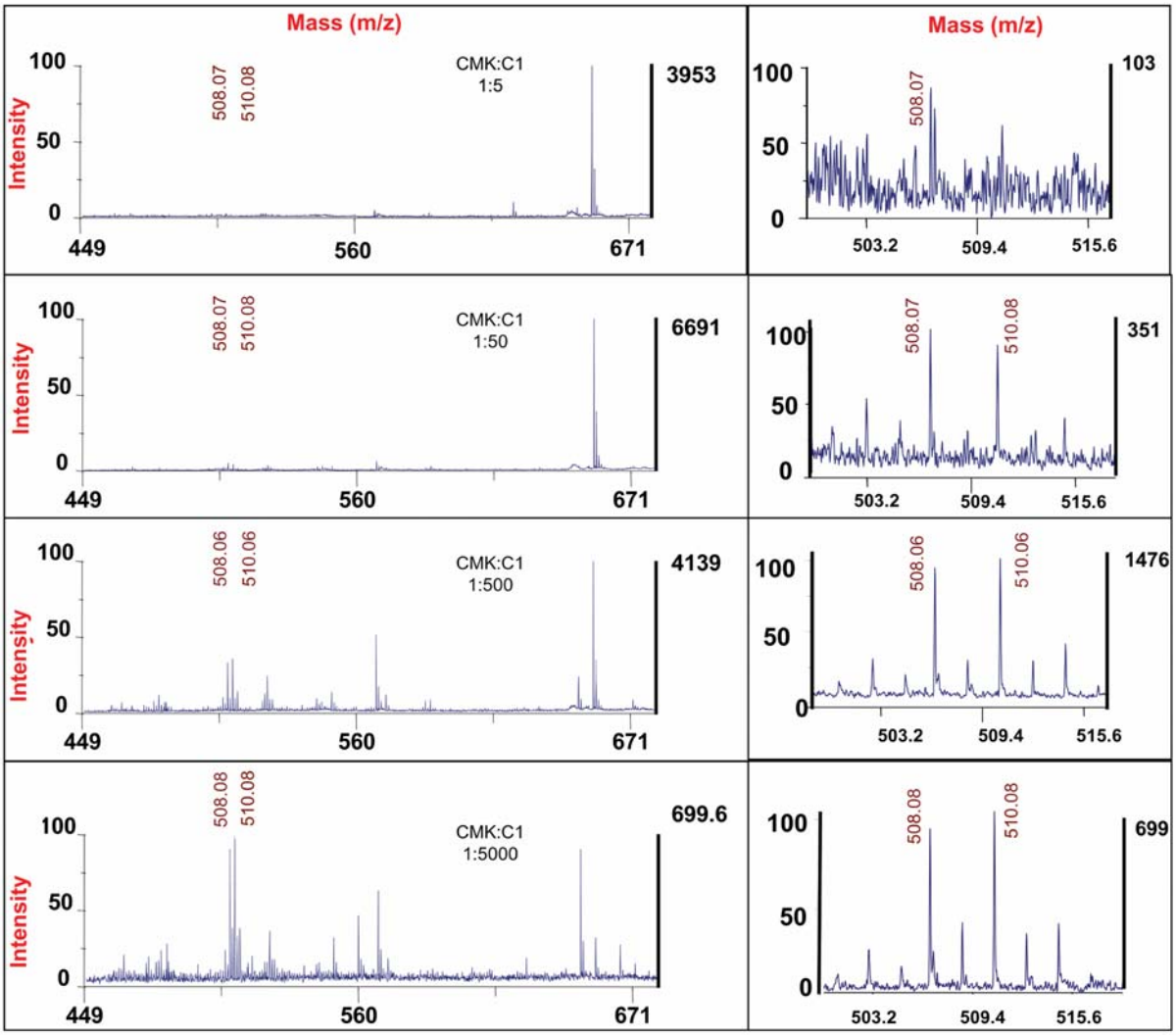


Fig. 7 Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) of the compound 1 (C1) molecule in the presence of increasing CMK concentrations. From top to bottom: 0.2, 0.02, 0.002 and 0.0002mM CMK, respectively; x-axis mass(m/z), y-axis relative intensity (left) and ion intensity (right). Peaks at m/z 508 and 510 correspond to C1. A detailed view is shown on the right panel. Note that in order to visualize C1 m/z 508 and 510 peaks spectra are presented at different intensity scales. The additional peaks correspond to adducts, corresponding to the α-cyano-4-hydroxycinnamic acid used as a matrix.

Conclusions

This paper describes a MD study carried out on a solvated dimer complex of CMK. The intermolecular hydrogen bonds elicited from the crystal structure were also observed in the dynamics trajectory but in a different structural pattern. Most direct protein-protein interactions in the crystal structure were observed in dynamics trajectory as being mediated by solvent through hydrogen bonds, a pattern already seen in the crystal structure of the complex. A deeper analysis of the solvent-mediated interactions indicates that four water-mediated protein-protein interactions remain along the MD trajectory, with one of them, in interface B, following a double-water motif. These *wet spots* [31], which likely are interfacial contact residues that interact through a water molecule, and double-water bonds, play an important role in the protein-protein interaction between contact surfaces, as it has been described in other protein models [32-34].

The study of intermolecular electrostatic interaction energies highlights the importance of water structure. Thus, the contribution of negatively charged residues was unfavorable when considering only protein atoms, but became instead stabilising when water molecules were included in the computation. These results indicate that water molecules are essential to explain how protein-protein electrostatic interactions favour complex formation and, together with the observations of solvent-mediated intermolecular interactions, reveal the importance of water for protein-protein interactions in the CMK dimer.

According to intermolecular interactions deduced from MD trajectory, some CMK residues could be *hotspots* for complex stabilization, and thus useful targets for protein-protein interaction disruption. Monitoring the positions of Arg^{21}, Ala^{22}, Asp^{23} and a water molecule during the MD trajectory a pharmacophore model was constructed where the first three residues of monomer 1 interacted favourably with monomer 2. The water molecules established a hydrogen bond network between Asp^{23} of monomer 1 and Lys^{76} of monomer 2. Compounds that matched the pharmacophore model were searched for in ACD, Specs and ChemDiv databases. After the docking procedure and visual analysis, four compounds were empirically tested and one of them exhibited binding to the recombinant CMK in a concentration-dependent manner according to intensity fading MALDI-TOF MS.

The data presented here indicate that a virtual screening approach can be used to identify candidate molecules that disrupt the CMK-CMK complex. This strategy can contribute to speeding up the discovery of new antimalarial, antibacterial and herbicidal compounds.

Acknowledgements

The excellent technical assistance of Drs. Eliandre de Oliveira and Maria Antonia Odena Caballol, from the Proteomics Platform, Barcelona Science Park, in the intensity-fading MALDI-TOF mass spectrometry experiments is gratefully acknowledged.

This work was financed in part with grants from the University of Barcelona (ACES-UB), the Spanish Ministerio de Ciencia y Tecnología (CTQ2006-06588/BQU, BIO2002-04419-C02-02 and BIO2008-01184) and the Generalitat de Catalunya (2005SGR00914).

Bibliography

1. Sacchettini JC, Poulter CD (1997) Science 277(5333):1788-1789
2. Chappell J (1995) Annu Rev Plant Physiol Plant Mol Biol 46:521-547
3. McGarvey DJ, Croteau R (1995) Plant Cell 7(7):1015-1026
4. Croteau R, Kutchan T, Lewis N (2000) Natural products (secondary metabolites). *In* Biochemistry and Molecular Biology of Plants. Buchanan B, Gruissem W, Jones R, editors. American Society of Plant Biologists, Rockville, MD. 1250–1268
5. Chappell J (2002) Curr Opin Plant Biol 2002 Apr; 5(2):151-7
6. Flesch G, Rohmer M (1988) Eur J Biochem 175(2):405-411
7. Rohmer M, Knani M, Simonin P, Sutter B, Sahm H (1993) Biochem J 295 (Pt 2):517-24
8. Schwarz MK (1994) Terpenbiosynthese in *Ginkgo biloba*: Eine überraschende geschichte. Ph. D. Thesis. *In* Eidgenossische Technische Hochschule. Eidgenossische Technische Hochschule, Zurich
9. Rodríguez-Concepción M, Boronat A (2002) Plant Physiol 130(3):1079-1089
10. Rohmer M, Grosdemange-Billiard C, Seemann M, Tritsch D (2004) Curr Opin Investig Drugs 5(2):154-162
11. Boucher Y, Doolittle WF (2000) Mol Microbiol 37(4):703-716
12. Lichtenthaler HK, Schwender J, Disch A, Rohmer M (1997) FEBS Lett 400:271-274
13. Eisenreich W, Schwarz M, Cartayrade A, Arigoni D, Zenk MH, Bacher A (1998) Chem Biol 5(9):R221-233
14. Lichtenthaler HK (1999) Annu Rev Plant Physiol Plant Mol Biol 50:47-65
15. Rohmer M (1999) Nat Prod Rep 16(5)(Oct):565-574
16. Kuzuyama T, Shimizu T, Takahashi S, Seto H (1998) Tetrahedron Letters 39(43):7913-7916
17. Jomaa H, Wiesner J, Sanderbrand S, Altincicek B, Weidemeyer C, Hintz M, Turbachova I, Eberl M, Zeidler J, Lichtenthaler HK, Soldati D, Beck E. (1999) Science 285(5433):1573-1576
18. Lichtenthaler HK (2000) Biochem Soc Trans 28: 785-789
19. Testa CA, Brown MJ (2003) Curr Pharm Biotechnol 4(4):248-259
20. Wiesner J, Borrmann S, Jomaa H (2003) Parasitol Res 90 (2):S71-76
21. Missinou MA, Borrmann S, Schindler A, Issifou S, Adegnika AA, Matsiegui PB, Binder R, Lell B, Wiesner J, Baranek T, Jomaa H, Kremsner PG (2002) Lancet 360(9349):1941-1942
22. Zeidler JG, Schwender J, Müller C, Wiesner J, Weidemeyer C, Beck E, Jomaa H, Lichtenthaler HK (1998) Zeitschrift für Naturforschung 53(c):980-986
23. Borrmann S, Issifou S, Esser G, Adegnika AA, Ramharter M, Matsiegui PB, Oyakhirome S, Mawili-Mboumba DP, Missinou MA, Kun JF, Jomaa H, Kremsner PG (2004) J Infect Dis 190(9):1534-1540
24. Miallau L, Alphey MS, Kemp LE, Leonard GA, McSweeney SM, Hecht S, Bacher A, Eisenreich W, Rohdich F, Hunter WN (2003) Proc Natl Acad Sci U S A 100(16):9173-9178
25. Gabrielsen M, Bond CS, Hallyburton I, Hecht S, Bacher A, Eisenreich W, Rohdich F, Hunter WN (2004) J Biol Chem 279(50):52753-52761
26. Wells JA, McClendon CL (2007) Nature 450(7172):1001-9
27. Sperandio O, Miteva MA, Segers K, Nicolaes GA, Villoutreix BO (2008) Open Biochem J 2:29-37
28. Smrcka AV, Lehmann DM, Dessal AL (2008) Comb Chem High Throughput Screen 11(5):382-95
29. Fletcher S, Hamilton AD (2007) Curr Top Med Chem (10):922-7
30. Whitty A, Kumaravel G (2006) Nat Chem Biol 2(3):112-8
31. Samsonov S, Teyra J, Pisabarro MT (2008) Proteins 73(2):515-25
32. Jiang L, Kuhlman B, Kortemme T, Baker D (2005) Proteins 58(4):893-904
33. Furukawa Y, Morishima I (2001) J Biol Chem 20; 276(16):12983-90
34. Langhorst U, Backmann J, Loris R, Steyaert J (2000) Biochemistry 39(22):6586-93
35. Janin J (1999) Structure 7(12):R277-9
36. van Dijk AD, Bonvin AM (2006) Bioinformatics 22(19):2340-7
37. Yanes O, Villanueva J, Querol E, Aviles FX (2005) Mol Cell Proteomics 4(10):1602-1613
38. Villanueva J, Yanes O, Querol E, Serrano L, Aviles FX. (2003) Anal Chem 75(14):3385-3395
39. Case DA, Pearlman DA, Caldwell JW, T. E. C. III, Wang J, Ross WS, Simmerling C, Darden T, Merz KM, Stanton RV, Cheng A, Vincent JJ, Crowley M, Tsui V, Gohlke H, Radmer R, Duan Y, Pitera J, Massova I, Seibel GL, Singh UC, Weiner P, Kollman PA (2002) AMBER7. University of California, San Francisco, CA

40. Wang R, Lai L, Wang S (2002) J Comp-Aid Mol Des 16:11.
41. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) J Comput Chem 25:1157
42. Jakalian A, Bush BL, Jack DB, Bayly CL (2000) J Comp Chem 21(2):132-146
43. Jorgensen WL, Chandrasekhar J, Madura JD (1983) J Chem Phys 79:926
44. Darden T, York D, Pedersen L (1993) J Chem Phys 98:10089
45. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) J Comp Phys 23:327
46. Berendsen HJC, Postman JPM, Van Gunsteren WF, DiNola A, Haak JR (1984) J Chem Phys 81:3684
47. Bernal C, Mendez E, Terencio J, Boronat A, Imperial S (2005) Anal Biochem 340(2):245-251
48. Bradford MM (1976) Anal Biochem 72:248-254
49. Karas M, Hillenkamp F (1988). Anal Chem 60(20):2299-2301
50. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) J Comput Chem 25(13):1605-1612
51. Humphrey W, Dalke A, Schulten K (1996) J Mol Graph 14(1): 33-38, 27-38
52. Levy Y, Onuchic JN (2006) Annu Rev Biophys Biomol Struct 35:389-415
53. Sreenivasan U, Axelsen PH (1992) Biochemistry 31(51):12785-12791

Diseño de una molécula pequeña y un péptido cíclico natural como ligandos de la CDP-metileritritol quinasa imitando interacciones proteína-proteína e interacciones mediadas por solvente.

Víctor Giménez-Oya[1], Óscar Villacañas[2$], Cristian Obiol-Pardo[2], Meritxell Antolin-Llovera[1], Jaime Rubio-Martinez[2] , Santiago Imperial[1]

[1] *Dept. de Bioquímica i Biología Molecular, Universitat de Barcelona (UB), Avda Diagonal 645, E-08028 Barcelona, Spain.*
[2] *Dept. de Química Física, Universitat de Barcelona (UB) and the Institut de Recerca en Química Teòrica i Computacional (IQTCUB), Martí i Franquès 1, E-08028 Barcelona, Spain.*
[$] *Current Address: Intelligent Pharma, S.L. Barcelona Science Park, C/ Baldiri Reixac 10, E-08028, Barcelona, Spain.*

Running title: Nuevos ligandos de la CDP-metileritritol quinasa

Palabras clave**:** CMK, MEP, MVA, interacciones mediadas por solvente, interacciones proteína-proteína, Dinámica molecular, Diseño de fármacos, Péptido cíclico, Moléculas pequeñas, Resonancia del Plasmón de superficie.

A la fecha de depósito de esta tesis este estudio se había enviado a la revista "Journal of Molecular Recognition" con respuesta de que algunos cambios son necesarios para la publicación del estudio.

Resumen

La ruta del metileritritol 4-fosfato (MEP) de biosíntesis de los precursores de isoprenoides (isopentenil difostafo y dimetilalil difosfato) existe en muchos de los patógenos humanos pero está ausente en animales, convirtiéndola en una prometedora diana para la obtención de nuevas drogas. Dos estrategias diferentes, una búsqueda virtual guiada por farmacóforo y un péptido cíclico que imita las interacciones proteína-proteína, se usaron para encontrar compuestos con la capacidad de unirse en la superficie de las zonas de interacción de la 4-(citidina 5-difosfo)-2*C*-metil-D-eritritol quinasa (CMK). La CMK cataliza el cuarto paso enzimático de la ruta del MEP. Una parte significativa de la hipótesis farmacofórica utilizada fue diseñada imitando interacciones proteína-proteína mediadas por puentes de agua importantes en la estabilización del homodímero de la CMK. Un derivado de la molécula 7H-furo[3,2-g]cromo-7-one y un péptido cíclico se encontraron con capacidad de unirse a la CMK según la técnica de resonancia del plasmón de superficie. Estudios de dinámica molecular sobre proteína-proteína, proteína-molécula pequeña y proteína-péptido cíclico se realizaron para complementar los ensayos experimentales.

Design of a small molecule and a natural cyclic peptide as ligands of CDP-methylerythritol kinase by mimicking direct protein-protein and solvent-mediated interactions

Victor Giménez-Oya[1], Óscar Villacañas[2$], Cristian Obiol-Pardo[2], Meritxell Antolin-Llovera[1], Jaime Rubio-Martinez[2] , Santiago Imperial[1]

[1] *Dept. de Bioquímica i Biología Molecular, Universitat de Barcelona (UB), Avda Diagonal 645, E-08028 Barcelona, Spain.*
[2] *Dept. de Química Física, Universitat de Barcelona (UB) and the Institut de Recerca en Química Teòrica i Computacional (IQTCUB), Martí i Franquès 1, E-08028 Barcelona, Spain.*
[$] *Current Address: Intelligent Pharma, S.L. Barcelona Science Park, C/ Baldiri Reixac 10, E-08028, Barcelona, Spain.*

Running title: New CDP-methylerythritol kinase ligands

Key words**:** CMK, MEP, MVA, solvent-mediated interactions, protein-protein interactions, molecular dynamics, drug design, cyclic peptide, small molecules, surface plasmon resonance.

ABSTRACT

The methylerythritol 4-phosphate (MEP) pathway for the biosynthesis of the isoprenoid universal building blocks (isopentenyl diphosphate and dimethylallyl diphosphate) is present in most of the human pathogens and is absent in animals, turning it into a promising therapeutic druggable pathway. Two different strategies, a pharmacophore-directed virtual screening and a protein-protein interaction(PPI)-mimicking cyclic peptide, were used to find compounds that bind to the PPI surface of the 4-(cytidine 5-diphospho)-2*C*-methyl-D-erythritol kinase (CMK), that catalyzes the fourth step of the MEP pathway. A significant part of the pharmacophore hypothesis used in this study was designed by mimicking water-mediated PPI relevant in the CMK homodimer complex stabilization. A 7H-furo[3,2-g]chromen-7-one derivative and a cyclic peptide were found to bind to CMK according to the surface plasmon resonance technology. Molecular dynamics (MD) studies of protein-protein, protein-small molecule and protein-peptide complexes were carried out to complement the experimental assays.

INTRODUCTION

Isoprenoids are one of the major families of compounds in nature comprising widespread functions (Sacchettini and Poulter 1997) in all kingdoms of life. This distribution covers from physiological processes to adaptative responses to the environment. Isoprenoid compounds are synthesized from two universal building blocks, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). Early works showed that the biosynthesis of the isoprenoid universal building blocks is carried out in yeast and animals through the mevalonic acid (MVA) pathway. For several years this fact was extrapolated to all the living organisms. Nevertheless, further studies carried out in bacteria and plants by Arigoni and Rohmer groups allowed the elucidation of a second biosynthetic pathway for IPP and DMAPP (for reviews, (Eisenreich, Schwarz et al. 1998; Lichtenthaler 1999; Rohmer 1999), namely the 2-*C*-methyl-D-erythritol 4-phosphate (MEP) pathway. This pathway produces the isoprenoid five-carbon building blocks from glyceraldehyde 3-phosphate and pyruvate, being MEP the first committed precursor. Nowadays, it is known that the MVA pathway is present in animals, yeast, and some bacteria while the MEP pathway is present in eubacteria, cyanobacteria and protozoa. Interestingly, both MVA and MEP pathways are present in plants, in which the MVA pathway is located in the cytoplasm whereas the MEP pathway is found in plasts. The presence of the MEP pathway in most of the human pathogens (including malaria and tuberculosis) has turned it into a promising druggable pathway.

The fourth step of the MEP pathway is an ATP-dependent phosphorylation of 4-(cytidine 5-diphospho)-2*C*-methyl-D-erythritol (CDP-ME) into CDP-ME 2-phosphate (Kuzuyama, Takahashi et al. 2000; Luttgen, Rohdich et al. 2000; Rohdich, Wungsintaweekul et al. 2000) through the action of the enzyme CDP-ME kinase (CMK). CMK belongs to the galactokinase, mevalonate kinase, homoserine kinase and phosphomevalonate kinase superfamily (previously named as GMHP superfamily) (Bork, Sander et al. 1993; Lange and Croteau 1999; Bonanno, Edo et al. 2001; Wada, Kuzuyama et al. 2003) which has three members in the MVA pathway. Although the sequence identity of proteins within this family is relatively low (10-20% among those that have structural information in the Protein Data Bank, PDB), they share a very similar structural scaffold (Cα RMSDs range from 2.6 to 4.0 Å) (Andreassi and Leyh 2004). Each of the seven different catalyst of this family of proteins is designed to transfer the γ-phosphoryl group to a different acceptor with the exception of the nonenzymatic Xol-1 protein.

The *Escherichia coli* CMK has been described as a homodimer complex in gel-filtration analysis, in matrix-assisted laser desorption time-of-flight MS (Miallau, Alphey et al. 2003) and in analytical centrifugation analysis (Gabrielsen, Bond et al. 2004; Gabrielsen, Rohdich et al. 2004), in corroboration with the crystallographic data (PDB=1jo4, (Miallau, Alphey et al. 2003). Only 4% of the CMK surface area was found to be involved in the formation of the homodimeric complex (Miallau, Alphey et al. 2003). This fact also converts CMK into an excellent target for searching small molecules capable of disrupting the interaction between the monomers in the CMK homodimer complex. For these reasons, we have focused on mimicking the CMK homodimer interface. The interaction zone between proteins is a promising objective in drug design but hardly achieved (Spencer 1998; Cochran 2000). Nowadays, the drug discovery process is focused on mimicking peptides and small molecules. Disrupting PPI interfaces with small “drug-like” molecules is a great challenge (Wells and McClendon 2007). Unlike peptides, small molecules are normally cheaper, have a better administration profile and cross the cell membrane more easily. In this study both strategies, a cyclic peptide and a small molecule

search, have been addressed (Fig. 1).

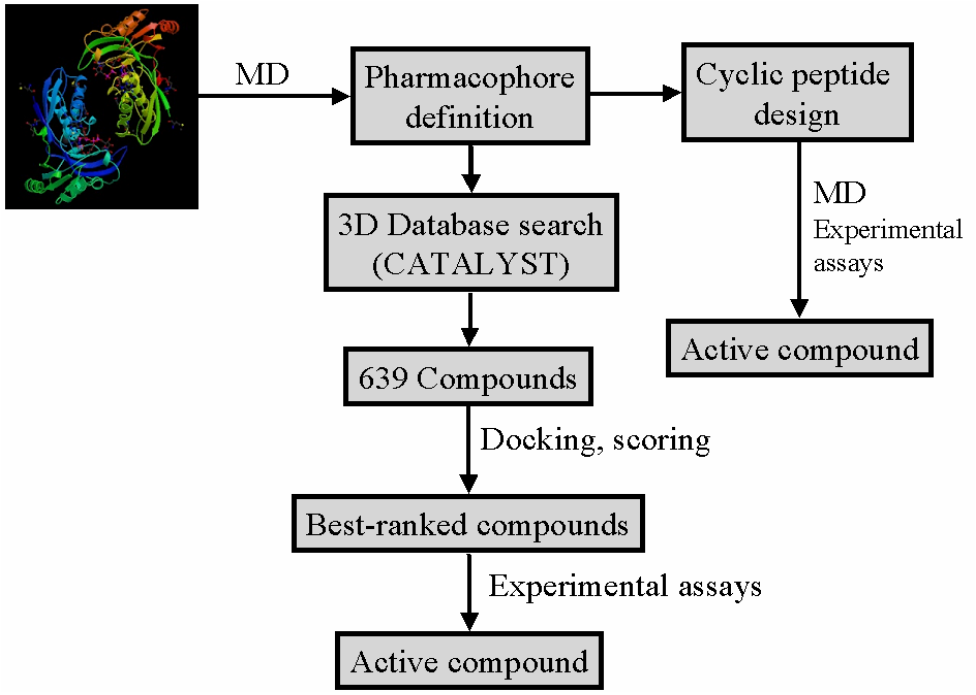


Fig.1. Scheme of the two strategies used in this study for both searching active small molecules and studying the effect of the cyclic peptide.

It is known that water is a crucial factor in protein folding, stability and in many other biological processes. Until recently, water molecules were usually ignored in molecular recognition and rational drug design studies. Fortunately, they have an increasing spreading influence in current structural studies. In *wet spots*, interfacial residues that interact through a water molecule (Samsonov, Teyra et al. 2008), water molecules contribute substantially to the free energy of the complex formation. As described in the literature (Lloyd, Garcia-Sosa et al. 2004; Leroux, Gresh et al. 2006), the inclusion of *wet spots* in the hypothesis of pharmacophore-directed high-throughput screenings has given successful results. Therefore, water molecules involved in PPI have been studied carefully in the present work.

Usually, the amino acid residues involved in the protein-protein interfaces are not located contiguously in the protein sequence; this is a problem if its natural sequence is used for designing a small peptide as complex disruptor. However, the particular characteristics of the CMK homodimer PPI surface, in which an isolated loop is buried into a partner's cleft (Fig. 2), make this protein an ideal starting point for the design of a small peptide as CMK complex disruptor, similarly to other protein models (Marsters, McDowell et al. 1994; Zobel, Wang et al. 2006).

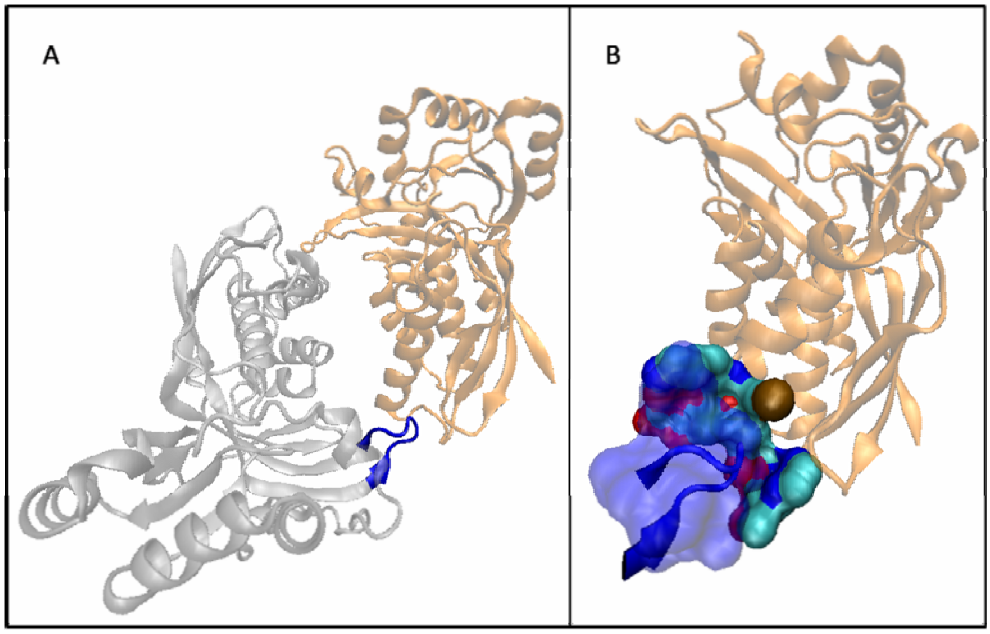


Fig.2.A. Figure of the CMK homodimer. Monomers are in transparent orange and transparent cyan. The loop involved in the interaction zone is marked in blue (only highlighted in one of the monomers). B. Figure of the interaction zone in which the loop of one monomer is inserted in the hole between one α-helix and two loops of the partner monomer. The latter monomer is in transparent orange with important residues for the interaction represented in surface, the mimicked loop is in blue and its surface in transparent blue. Figure rendered with VMD programme (Humphrey, Dalke et al. 1996).

Mimicking peptides are a reliable strategy to disrupt protein-protein interaction. However, most of them are not "drug-like" compounds because they normally have poor oral availability and high enzymatic degradation (Zang, Yu et al. 1998). Cyclization of mimicking peptides can reduce unwanted proteolysis and their flexibility in solution. Moreover, constraining conformational freedom and thus decreasing entropy lost in binding results in higher-affinity compounds compared to their linear counterparts. This entropy-mediated affinity gain is well documented in literature (Bach II, Eyermann et al. 1994; Morgan, Holland et al. 1994; Zang, Yu et al. 1998; Dumez, Snaith et al. 2002; Andrews, McInnes et al. 2004; Nam, Ye et al. 2004; Chang and Chu 2005). One interesting case is described in metalloproteinases (Dumez, Snaith et al. 2002) in which, additionally, it was postulated that this entropy-mediated gain could be identified in a lower free energy of the hydrolytic transition state of the complex compared to the cleaved products. Under these circumstances, the substrate may remain bound to the metalloproteinase and may inhibit the metalloproteinase activity. Unsatisfactorily, inhibitory compounds that could have corroborated their hypothesis were not found in the work.

The surface plasmon resonance (SPR) technology has proved to be a valuable tool for both quantitative and qualitative analysis of complex interaction. This technology allows the parameterization of the complex interaction at real time without the use of tags in the interactors (Cooper 2002; Karlsson 2004). The basis of this technology involves the immobilization of the receptor to a sensor chip and the direct binding of a ligand. The complex formation is assessed by changes in surface plasmon resonance which is associated to changes in the refractive index of the solution close to the immobilized receptor. The refractive index is directly associated with the mass concentration in the surface layer of the sensor chip, being increased when the complex is formed between the ligand and the receptor. This technology was used in this study to quantify the binding of the designed molecules to CMK. Furthermore, we complemented this study performing docking and molecular dynamics simulations to find new molecules disrupting PPI in the CMK homodimer.

MATERIALS AND METHODS

Molecular Dynamics Simulation of the CMK homodimer

As described in Gimenez-Oya and others 2009, minimizations and MD simulations were carried out with the AMBER7 package of programmes (Case, Cheatham et al. 2005) using *parm99* and *gaff* force fields (Wang, Lai et al. 2002; Wang, Wolf et al. 2004). The antechamber module of AMBER7 and the BCC-AM1 method (Jakalian, Jack et al. 2002) were used to assign the atom type and the atomic charges to the organic small molecules. Hydrogen atoms were added to the CMK X-ray structure (PDB=1jo4) (Miallau, Alphey et al. 2003) with the LEaP module of AMBER7. The system was neutralized with 31 Na^+ ions following a grid-shaped procedure for mapping electrostatic potential surface. A first minimization was carried out *in vacuum* with a water distance-dependent dielectric constant (*4r*). After this, a solvation cubic box of TIP3P waters (Jorgensen, Chandrasekhar et al. 1983) was created with a minimum distance between the protein and the edge of the box of 10 Å and waters closer than 1,8 Å to any of the molecules or the protein atoms were removed. Then the system was relaxed by a two-step procedure, optimizing the water molecules in the first step and optimizing the whole system in the last. Long-range electrostatic energy was computed in MD simulations using the PME summation method (Darden, York et al. 1993). SHAKE algorithm (Ryckaert, Ciccotti et al. 1977) was used to constrain the bonds involving the hydrogen atoms. The integration time was set to 2 fs and every 25 steps the non-bonded pair list was updated with a cutoff of 9 Å. Berendsen's algorithm (Berendsen, Postma et al. 1984) was used to introduce the regulation of the pressure and temperature with time coupling of 1.0 and 0.2 ps, respectively. First, the system was heated gradually to 300K increasing 30K each 10 ps, at constant volume. Then a pressure-constant equilibration process followed, in which a temperature coupling constant of 1.0ps was applied during the first 20 ps. Further 40 ps of equilibration were carried out raising the temperature coupling constant to 2.0 ps. During equilibration, ions and protein were controlled under harmonic restrains of 1 and 5 kcal mol^{-1} $Å^{-1}$ respectively. Then, 60 ps were used for removing restrains. The last stabilized 1ns with no restrains was used as production run for analysis.

Energy analysis of the CMK homodimer

Lennard-Jones and Coulombic energies were calculated between every side chain of the first monomer of CMK and the entire second monomer as previously described (Gimenez-Oya, Villacanas et al. 2009). These energy contributions were computed for snapshots at every picosecond of molecular dynamics throughout 1 ns of production time with the ANAL module of AMBER7 (Case, Cheatham et al. 2005).

Solvent-mediated interaction analysis of the CMK homodimer

Water molecules that established hydrogen bonds with both monomers were analyzed with a home-made programme throughout the last 200 ps (Fig. 3). Hydrogen bond distances were monitored and occupancy was defined as the percentage of time in which there existed at least one hydrogen bond to each monomer within a distance (hydrogen – acceptor atom) of 2.5 Å.

Design of the pharmacophore model from the MD trajectory

Interactions between monomers were monitored during the last 200ps to configure the pharmacophore model. If more than one atom contributed to the same interaction (e.g. both oxygen atoms of the carboxylate group of Glu), the centre of masses (CM) was calculated and it was used as a single interaction point. Maximum and minimum distances throughout the MD were obtained between every pair of points and were used to construct the pharmacophore for 3D database search. An interaction feature was assigned to each point; hydrogen donor/acceptor, hydrophobic and positive/negative charge. The basis of this new pharmacophore hypothesis was the previously described pharmacophore (Gimenez-Oya, Villacanas et al. 2009) but improved by adding two further hydrogen acceptor points corresponding to stable water positions which were found into wet spots during the simulation of the CMK homodimer.

Molecular modeling of the cyclic peptide/CMK complex

The cyclic peptide/CMK complex was constructed from the minimized CMK homodimer complex. The sequence of the peptide was chosen attempting to retain the most important interactions in the homodimer interface (amino acid residues from Gln^{20} to Thr^{27}). Then, the backbone of the peptide was closed by adding two glycines and creating the new amide bond with the LEaP programme of the AMBER7 package (Case, Cheatham et al. 2005) (Fig. 5B). The cofactors adenosine 5'-(β, γ-imido)triphosphate (AMP-PNP) and 4-(cytidine 5'diphospho)-2-C-methyl-D-erythritol (CDP-ME) were retained in the protein monomer. The system was solvated adding a cubic box of TIP3P waters (Jorgensen, Chandrasekhar et al. 1983), and neutralized by adding and appropriate number of counterions.

Database search and molecular modeling of the 7H-furo[3,2-g]chromen-7-one derivative/CMK complexes

The pharmacophore model was used in a 3D database search in order to find commercially available small molecules. Hits were then considered primary candidates as CMK ligands. For this purpose, CATALYST (Accelrys Inc.) programme was used. After constructing the pharmacophore and relaxing its distances by applying the maximum and minimum deviations observed in the MD simulation, search gave 639 hits by using the best flexible search. Databases included were: Mini Maybridge, ChemDiv NC, ChemDiv DC, Maybridge 2004, SPECS, Nat Diverse, ACD, IBS 2004, NCI, Aurora and Derwent. The general strategy can be seen in Fig. 1. Then, hits were docked considering one of the monomers of CMK as a rigid receptor and using our home-made programme of docking, Dock_Dyn (Rubio-Martinez, Pinto et al. 2005). An average number of 100 conformations per molecule were docked inside the receptor, and best poses were selected for further analysis. The quality of each pose was evaluated, in a first step, by measuring its pharmacophore RMS deviation with respect to the pharmacophore deduced from the protein-protein contacts. Thus, the lower the pharmacophore RMS deviation, the better is the docking pose. Moreover, in order to include the binding energy, the XSCORE (Wang, Lai et al. 2002) programme was used as the scoring method. Finally all docked compounds were ranked, using RMS and XSCORE criteria, and the best four were purchased from different chemical suppliers. A 7H-furo[3,2-g]chromen-7-one derivative, named in this text as compound (Fig. 5A), was shown to be active in

experimental assays. For further analysis, two docking structures of the compound/CMK complex were selected: the pose with the lowest RMS deviation (Fig. 6A) and the pose with the highest XSCORE pK_D (Fig. 6C), which were different in this case. Cofactors AMP-PNP and CDP-ME were retained in the systems.

Both systems were solvated by adding a cubic box of TIP3P waters, and neutralized by adding and appropriate number of counterions. Subsequent minimization and molecular dynamics of both systems were carried out in order to elucidate the most favourable binding pose of this molecule taking into account the protein flexibility and thus complementing the docking procedure.

Minimization and MD on the cyclic peptide/CMK complex and the compound/CMK complexes

A standard minimization protocol described already in the case of the CMK homodimer was performed to remove steric hindrance and, additionally, a similar molecular dynamics protocol was carried out to improve the initial structure of the complexes. After energetic convergence, 100 snapshots were extracted from the production dynamics and structural features, as well as the binding free energy, were analysed.

Binding Free Energy calculations using the MMGBSA approach

The Molecular Mechanics Generalized Born Surface Area (MMGBSA) protocol (Kollman, Massova et al. 2000) was applied to add quantitative binding free energies to the description of all CMK complexes. This protocol was used within the one-trajectory approximation and was evaluated extracting 100 time-equidistant snapshots from the molecular dynamics trajectory.

To solve the polar contribution of solvation free energy, the Generalized Born equation was applied using the sander programme of AMBER7 package (Case, Cheatham et al. 2005) and the parametrization of Tsui and Case (Tsui and Case 2000). To solve the non-polar contribution of solvation, a linear relationship with the solvent accessible surface area (SASA) was applied, using standard parameters and calculating the SASA by means of the Linear Combination of Pairwise Overlap (LCPO) method (Weiser, Shenkin et al. 1999). Entropic effects of the binding free energy were neglected in all cases, thus supposing that the cyclic peptide and the compound have a similar entropic penalty. Additionally an MMGBSA 'per residue' protocol was carried out to discern the most important residues that participate in the binding free energy of the complexes.

Experimental assays

Site-directed mutagenesis

Asp^{80} was described as an important residue involved in a direct interaction between the two monomers of the CMK homodimer (Miallau, Alphey et al. 2003). The point mutation Asp80Ala was introduced in the gene encoding CMK by site-directed mutagenesis using an oligonucleotide-directed strategy following protocols associated with the QuickChange site-directed mutagenesis procedures (Stratagene). To create pQE30-CMK D80A, pQE30-CMK wild type was used as a template and the following combination of primers were used in the amplification step: CMKEcoD80AF (5' AAAACTGCGCAG**CCA**GCGGGCGTCTT 3') and CMKEcoD80AR (5' AAGACGCCCGCTG**G**CTGCCGCAGTTTT 3'). Nucleotide changes are in boldface and

the mutated codons (amino acid changes) are underlined. The conditions of the PCR were 18 cycles of 30s, 95ºC; 60s, 65ºC; 504s, 72ºC. The PCR product was digested with *DpnI* enzyme and *Escherichia coli* $DH5_{\alpha}$ competent cells were transformed with the digestion reaction. Plasmids were extracted from isolated colonies and mutations were verified by DNA sequencing.

Purification of CMK and CMK D80A

Recombinant proteins were purified from the M15[pREP4] *Escherichia coli* cells harbouring the pQE30-cmk or pQE30-cmk D80A expression plasmid. Cells were grown in LB with ampicillin(100µg/ml) and kanamycin(25µg/ml) at 37ºC until DO_{600} of 0,4 and then, the culture was incubated with 1mM IPTG during 12 h at 25ºC. Cells were harvested and resuspended in lysis buffer (40mM Tris-HCl pH=8, 100mM NaCl, 0,5mM EDTA, lisozyme (1mg/mL) and complete mini-EDTA free tablets (Roche)). After an incubation of 20 min, cells were sonicated and the cell-free system was centrifuged to remove the cell debris. Then, the supernatant was incubated for 20 min with protamine sulphate (1.25mg/ml) and was centrifuged. The His6-CMK protein was retained on the resin of Hi-trap Chelating columns (GE Healthcare) and eluted using an imidazole gradient (10-500mM) in 40mM Tris-HCl buffer, pH 8.0 containing 100mM NaCl. CMK preparations were desalted by gel filtration on disposable Sephadex G-25 columns (PD10, GE Healthcare) equilibrated in 40mM Tris–HCl buffer with 0.1M NaCl, distributed into aliquots, snap-frozen with liquid N_2 and stored at -80ºC until use. Protein concentration was measured by the method of Bradford (Bradford 1976) using bovine serum albumin as a standard.

Synthesis of the cyclic peptide

The cyclic peptide, cyclo(GlyGlnArgAlaAspGlyTyrHisThrGly), was synthesized in an Abi433A synthesizer (Applied Biosystems) using the FastMoc protocol. Couplings were carried out in an Fmoc-Gly-2-chlorotrityl resin using TBTU/HOBt as a coupling agent. Linear peptide was eluted from the resin with TFA 1% in DCM and was cycled with PyBOP/HOAt in DCM/DMF (9:1). Side chains were deprotected using TFA/water/triethylsylane (95:2.5:2.5). Then, the cyclic peptide was purified by preparative high pressure liquid chromatography (HPLC) in a Waters DeltaPrep 4000 (Waters Symmetry C18 column) in a MeCN gradient (5-20%). The final product was characterized by matrix assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry in 4700 Proteomics Analyzer (Applied Biosystems) and by amino acid analysis through the HPLC AccQ.Tag method (Waters) in a Delta 600 (Waters).

Analysis of the interaction (surface plasmon resonance)

SPR experiments were carried out using BIAcore T100 (GE Healthcare). Recombinant CMK was purified and immobilized onto CM5 sensor chip using the standard amine coupling procedure recommended by the manufacturer (Johnsson, Lofas et al. 1991). The cyclic peptide was also immobilized onto CM5 sensor chip using the standard aldehyde coupling procedure. Binding experiments were performed using HBS-N buffer (10mM Hepes, 150mM NaCl, pH=7.2). Some binding assays with increased concentrations of the ionic force or with Tween20 0.05% were done. All experiments were done at a flow rate of 15µL/min. Other flow rates were tried to assess that no significant mass transport effect occurred. To remove potential artifacts, a standard method

of “double referencing” was used and the binding of peptide or ligand to CMK was expressed in resonance units (RU). Serial dilutions of molecules in HBS-N were done and were injected onto the surface of the chip. For monitoring the dissociation time of the complex, HBS-N buffer was flown over the surface during 240 s. The collected data were interpreted with the BIACORE T100 evaluation software, extracting the equilibrium dissociation constants (K_D), ratio of association (k_a) and ratio of dissociation (k_d).

RESULTS

Mimicking the interaction zone (I). The pharmacophoric hypothesis

Water molecules can play an important role in the stabilization of complexes (both protein-drug and protein-protein) and incorporation of waters in structure-based drug design has been described as a reliable strategy (Lloyd, Garcia-Sosa et al. 2004; Leroux, Gresh et al. 2006). As described in Gimenez-Oya and others 2009, water molecules are important in the PPI zone of the CMK homodimer. Water molecules increase the PPI zone of CMK, which is only the 4% of the surface area (Miallau, Alphey et al. 2003), through the presence of *wet spots* and double-water mediated interactions (Gimenez-Oya, Villacanas et al. 2009). Previously, we estimated that the electrostatic interaction energy between CMK monomers was more coherent when including water molecules in the calculation (Gimenez-Oya, Villacanas et al. 2009), most probably because the solvated medium buffered the negative charge of the CMK monomers. Based on these findings we surveyed for stable waters-mediated hydrogen bonds present in the interaction zone between the two monomers as described in Materials and Methods. Two water-mediated interactions were found close to the zone mimicked by the pharmacophoric hypothesis described in our previous work (Fig. 3). The first solvent-mediated interaction was found between the carbonyl group of Asp^{23} in monomer 1 and Gly^{49} (carbonyl group) and Arg^{72} (side chain) amino acid residues of monomer 2; the second one was established between Ala^{22} (carbonyl group) of monomer 1 and Gly^{49} (carbonyl group), Pro^{85} (carbonyl group), Gly^{87} (amide group) and Ser^{88} (carbonyl/amine groups) of monomer 2. The percentages of the running time in which any water molecule was within a distance of 2.5 **Å** from the amino acid residues were 94%/99% (to monomer 1 and to monomer 2, respectively) for the Asp^{23} related interaction and 83%/97% (to monomer 1 and to monomer 2, respectively) for the Ala^{22}-related interaction. The average minimum distances between amino acid residues and the water molecules were 1.98 Å (± 0.32 Å)/1.85 Å (± 0.13 Å) and 2.16 Å (± 0.64 Å)/ 1.94 Å (± 0.2 Å) (monomer 1/monomer 2), respectively for Asp^{23}-related interaction and for Ala^{22}-related interaction. The water molecules that mediated the interaction were not always the same throughout the MD trajectory. Up to 8 (for Asp^{23}-related interaction) and 5 (for Ala^{22}-related interaction) different water molecules took turn to fill the gap between residues and mediate the interaction among them. Positions of the water molecules in these solvent-mediated interactions were used to add two additional hydrogen acceptors points to the previous pharmacophoric hypothesis. The positions of the water molecules in these solvent-mediated interactions were used to add two additional hydrogen acceptors points to the previous pharmacophoric hypothesis.

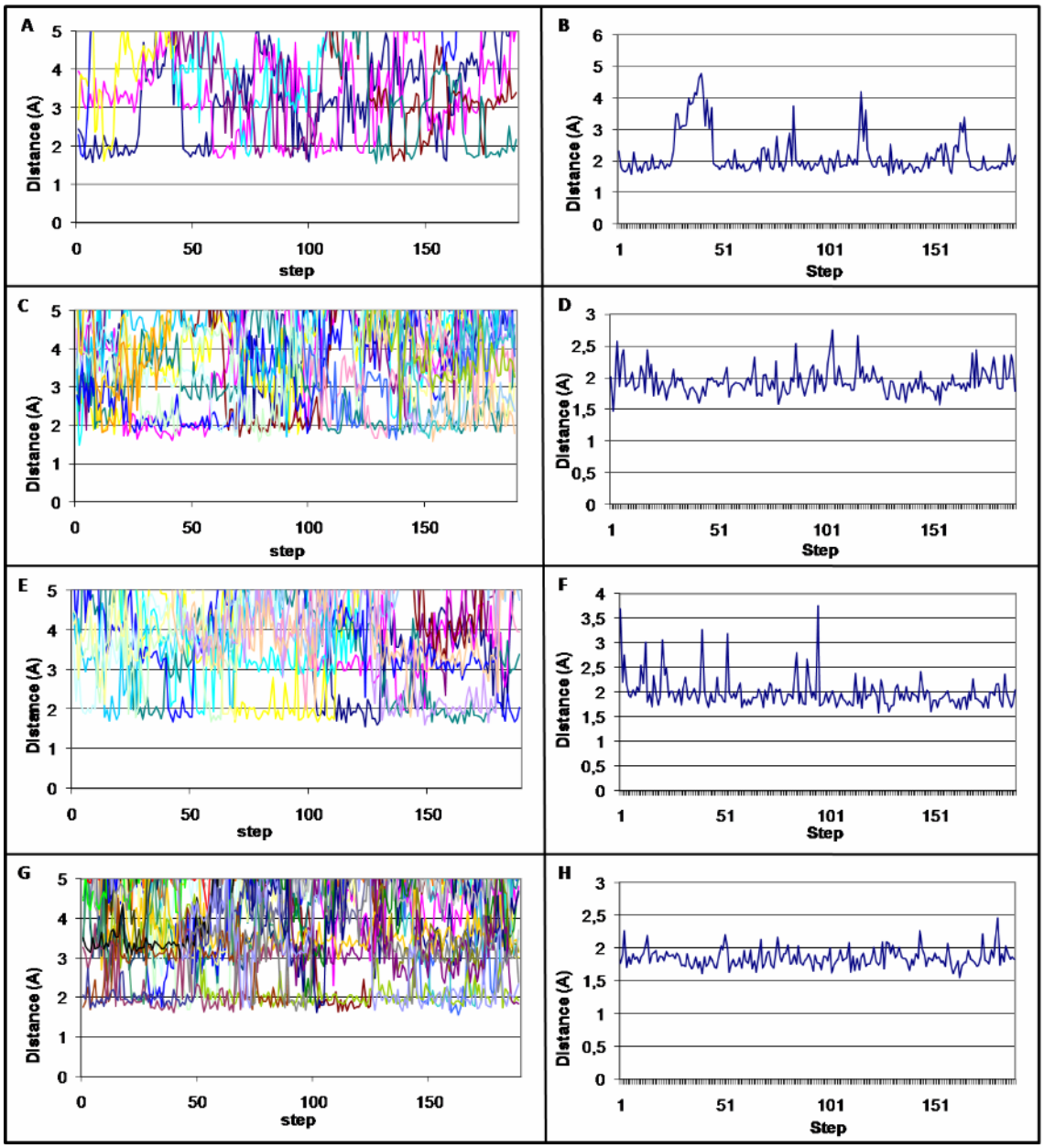


Fig.3. Plots of distances between water molecules and amino acid residues of the interaction zone during the production time of the molecular dynamic trajectory. Different colors represent different water molecules. A. Plot of distances between Ala^{22} residue and waters. B. Plot of the minimum distance between Ala^{22} and waters. C. Plot of distances between waters and residues of the other monomer. D. Plot of the minimum distance between waters and residues of the other monomer. E. Plot of distances between residue Asp^{23} and waters. F. Plot of the minimum distance between Asp^{23} residue and waters. G. Plot of distances between waters around Asp^{23} and the other monomer. H. Plot of the minimum distance between waters around Asp^{23} and the other monomer.

In summary, the new hypothesis mimics two *wet spots* (Samsonov, Teyra et al. 2008), including Asp^{23}, which is involved in two different protein-protein water-mediated hydrogen bonds, and Ala^{22}, a hydrogen donor that mimics Arg^{21} and three hydrophobic points. Maximum and minimum distances between every pair of points from MD were used to construct the new pharmacophore model for database search (Fig. 4). Database searches and molecular docking were carried out as described in Materials and Methods. Best ranked molecules were purchased and the interaction between these molecules and the protein was empirically assessed through the Surface Plasmon Resonance technology. A 7H-furo[3,2-g]chromen-7-one derivative (compound, Fig. 5A) was shown to be active in experimental assays.

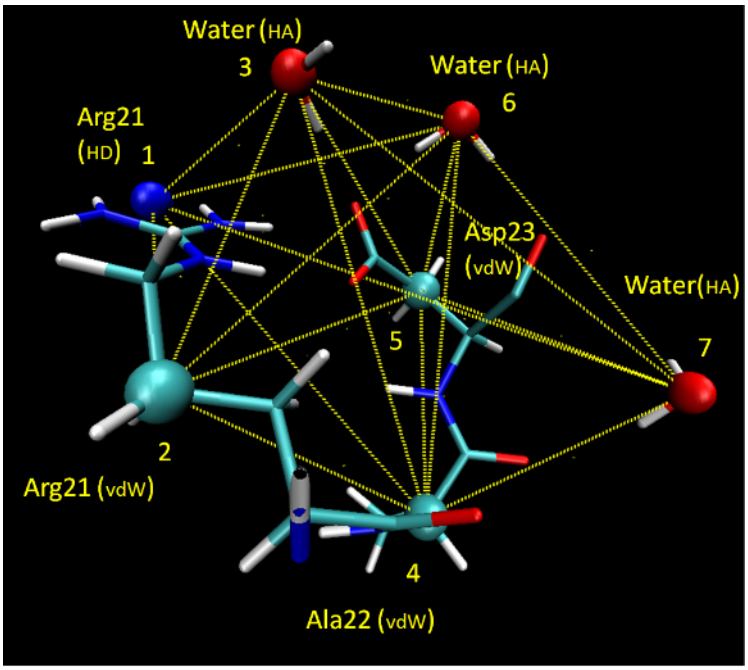


Fig.4. Pharmacophoric points forming the interaction model used in the database search. Points are derived from residues 21-23 of monomer 2 and water molecules. Arg[21] provided a hydrogen donor (HD) and a van der Waals (vdW) pharmacophoric point; Ala[22] and Asp[23] provided vdW pharmacophoric points; water molecules from three water bonds provided three hydrogen acceptors (HA). MD distances between pair of points are: 1-2: 2.6–3.1 Å; 1–3: 3.0–4.8 Å; 1–4: 7.0–8.9 Å; 1–5: 4.0–4.8 Å; 1-6: 8.1-9.5 Å; 1-7: 10.7-11.0; 2–3: 3.7–5.8 Å; 2–4: 6.0–7.2 Å; 2–5: 3.8–4.6 Å; 2-6: 8.0-9.4 Å; 2-7: 9.4-9.8 Å; 3–4: 8.7–11.3 Å; 3–5: 3.1–5.8 Å; 3-6: 5.9-6.8 Å; 3-7: 10.8-11.4 Å; 4–5: 5.4–6.7 Å; 4-6: 9.0-9.6 Å; 4-7: 4.4-4.6 Å; 5-6: 4.2-5.8 Å; 5-7: 7.4-7.5 Å; 6-7: 8.5-8.7 Å. Red points account for hydrogen acceptors, blue point accounts for hydrogen donors and cyan points account for van der Waals contacts. Figure rendered with VMD programme (Humphrey and others 1996).

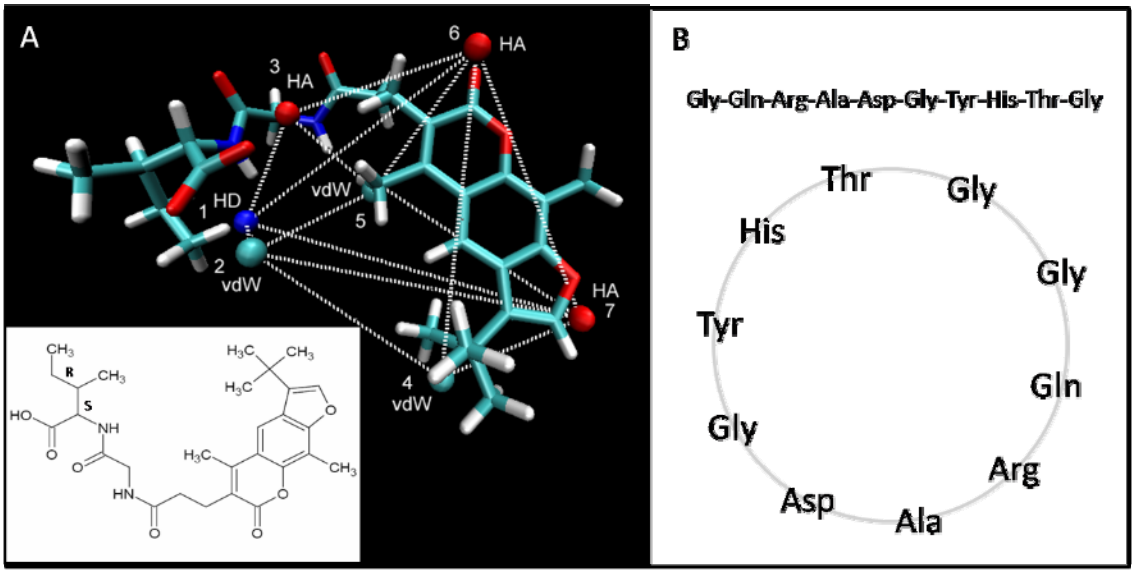


Fig.5. A. 3D structure of the compound (a 7H-furo[3,2-g]chromen-7-one derivative) showing the disposition of the pharmacophoric points. The inset gives 2D structure of the compound. B. Scheme and sequence of the cyclic peptide. 3D image rendered with VMD programme (Humphrey and others).

Description of the compound binding mode in complex with CMK

To better predict the binding mode of the compound in complex with CMK, two MD of 2ns were performed for the compound in complex with CMK monomer. The former departed from the best docking pose with the lowest pharmacophore RMS deviation (pose 1) and the latter from the docking pose with the highest XSCORE pK_d (pose 2).The combined use of docking and further MD was also applied in our previous work with good results (Obiol-Pardo, Granadino-Roldan et al. 2008) Both systems were equilibrated during the first nanosecond, and consequently, the last nanosecond was used to extract the structural information.

Fig. 6A shows the docking structure of the complex constructed with pose 1 and Fig. 6B shows the final structure of MD. In MD, important residues that participated in van der Waals recognition were Met^{75}, Lys^{76}, Ala^{79} and Leu^{84}. Particularly, a slight reorientation of the compound was noticed throughout MD to place the methyl of the central ring closer to Leu^{84} but decreasing van der Waals contacts with Met^{75}. As a consequence of this movement, the initial hydrogen bond formed between the carbonyl group of the compound and Arg^{72} was partially lost during MD (Fig. 6B). Protein flexibility and compound mobility added in the MD seemed to disfavour the electrostatic interaction between this polar residue and the compound. A similar behaviour was observed in the cyclic peptide/CMK MD, as will be described later.

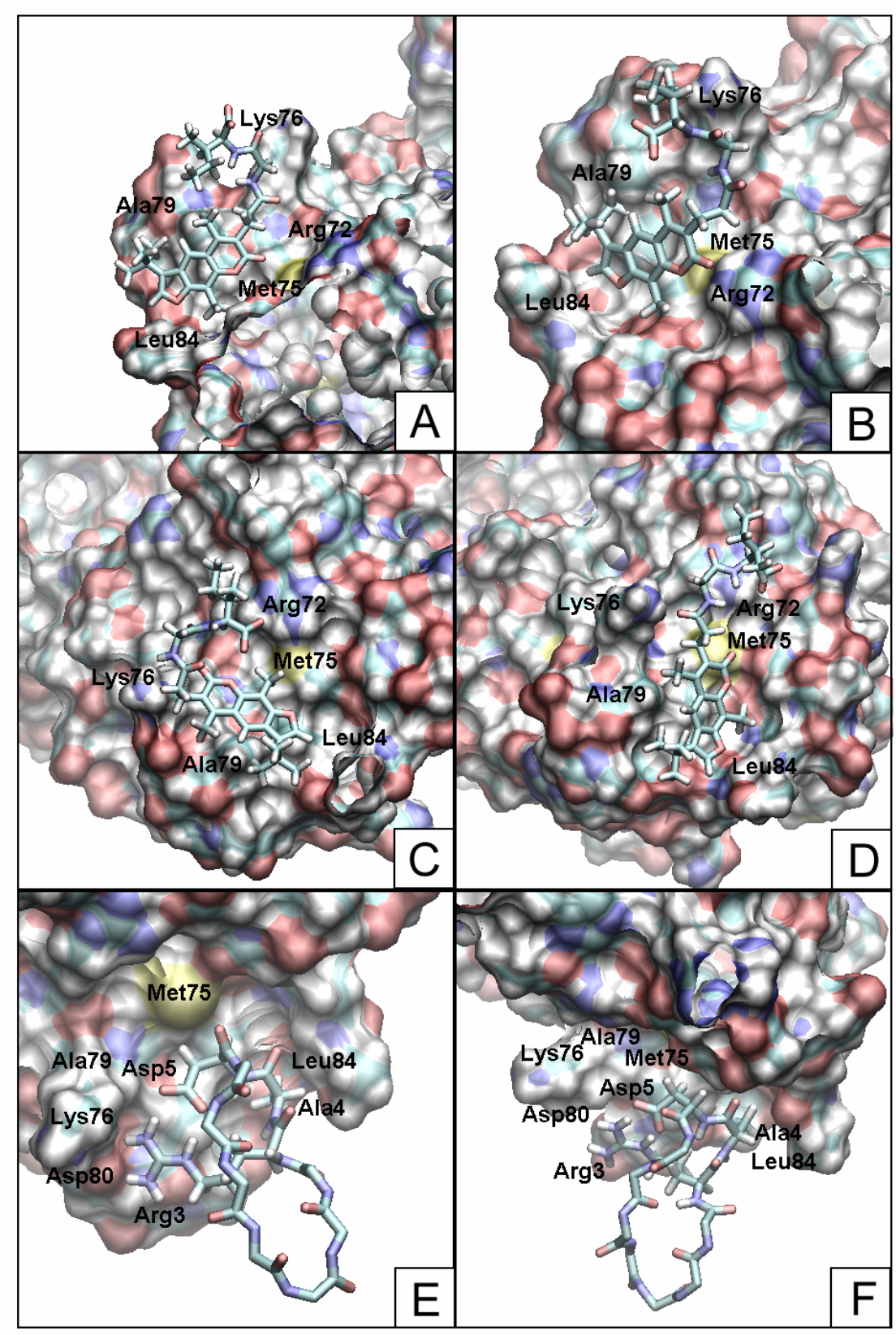


Fig.6. Figures of the complexes between either the compound or the cyclic peptide and CMK before (left) and after (right) MD. CMK protein is in surface, the compound and the cyclic peptide are in licorice and backbone of cyclic peptide is shown. A. Pose 1 (lowest RMS) of compound-CMK complex before MD. B. Pose 1 (lowest RMS) after MD. C. Pose 2 (highest XSCORE) of compound-CMK complex before MD. D. Pose 2 (highest XSCORE) after MD. E. Cyclic peptide-CMK complex before MD. F. Cyclic peptide-CMK complex after MD. Figure rendered with VMD programme (Humphrey and others 1996).

The predicted binding free energy using the MMGBSA protocol was -20.61 kcal mol^{-1} for pose 1, which was lower than the experimental value reported below in this work (-5.01 kcal mol^{-1}). Nevertheless, it is worth noting that the MMGBSA protocol used here is a crude estimation of the experimental value and the entropic contribution, which was not calculated, would decrease this estimated energy. A 'per residue' analysis throughout MD of pose 1 (Fig. 7A) was carried out and showed the most remarkable residues that participated in the complex stabilization. Thus, the compound contributed with -9.93 kcal mol^{-1} to the complex formation (last bar), whereas residues Met^{75}, Lys^{76}, Ala^{79}, Leu^{84} and Thr^{86} of the receptor contribute with -1.70 kcal mol^{-1}, -2.65 kcal mol^{-1}, -1.41 kcal mol^{-1}, -2.20 kcal mol^{-1} and -0.75 kcal mol^{-1}, respectively.

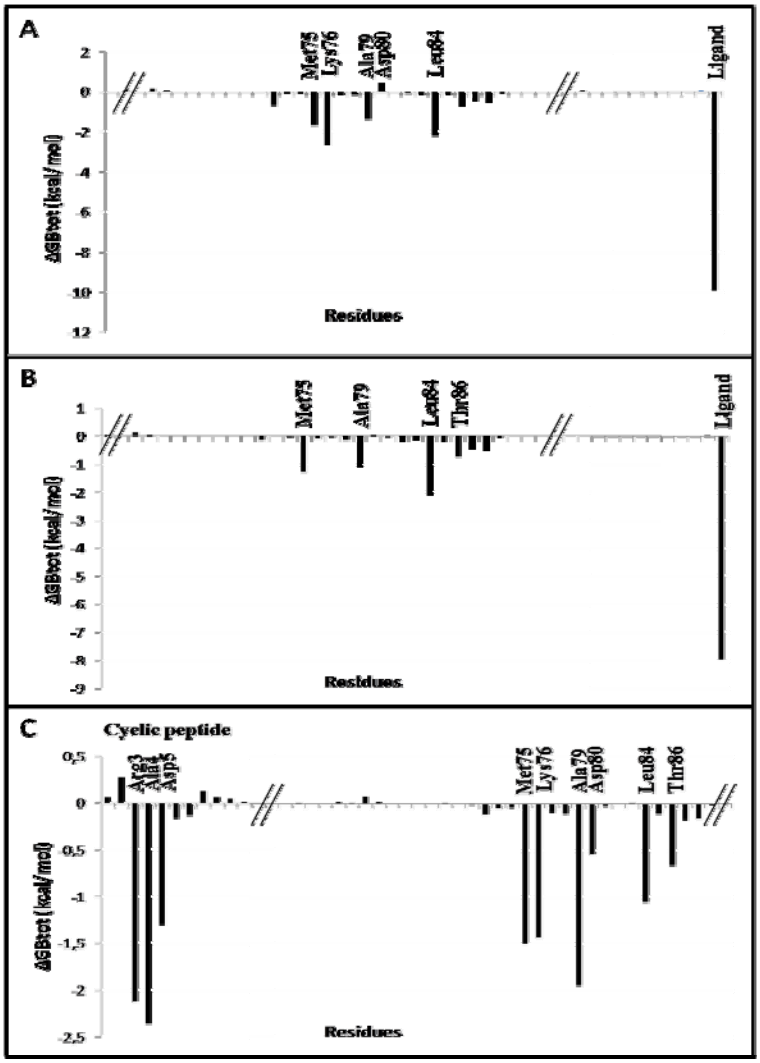


Fig.7. Plots of total binding free energy per residue extracted from the MMGBSA protocol and averaged over 100 snapshots. Important residues are marked. A. Plot of total energy per residue of Pose 1 complex (compound with the lowest RMS and CMK). B. Plot of total energy per residue of Pose 2 complex (compound with the highest XSCORE and CMK). C. Plot of total energy per residue in the complex between the cyclic peptide and CMK. The first 10 residues are the cyclic peptide amino acid residues and the other 283 amino acid residues are from the CMK monomer.

In addition, pose 2 was also evaluated by using the same protocol. The van der Waals recognition of this pose was very similar to pose 1. On the other hand, the carboxylate group of the compound now formed a hydrogen bond with Arg^{72} in the docking pose (Fig. 6C), which was also lost during MD, although an electrostatic interaction was maintained (Fig. 6D). Finally, the predicted binding free energy using the MMGBSA protocol was -14.94 kcal mol^{-1} for pose 2, higher than that obtained with pose 1. We suggest therefore that the most reliable pose for this molecule is achieved during the MD of pose 1. Additionally, Fig. 7B shows the MMGBSA 'per residue' protocol for pose 2 in complex with CMK. This analysis is in line with the interaction analysis already described, showing the strongest contribution in the compound (last bar, -7.95 kcal mol^{-1}). Among residues, Met^{75}, Ala^{79} and Leu^{84} showed the most remarkable binding free energy contribution being -1.31 kcal mol^{-1}, -1.1 kcal mol^{-1} and -2.13 kcal mol^{-1}, respectively. Note that Lys^{76} did not appear as relevant in this analysis with pose 2.

In both modelled binding modes, compound bound to the flat and extended CMK PPI surface, part of it is solvent exposed not participating in binding free energy. A further improvement of this hit compound should be obtained by minimizing these solvent exposed groups.

Mimicking the interaction zone (II). Using the loop sequence.

The interaction zone of the monomers of the CMK homodimer is characterized by the insertion of a loop, which is located between two beta-sheets of one monomer, into a cleft between an α-helix and two other loops from the partner monomer (Fig. 2A). This loop, which was the target of our previous efforts of mimicking with organic molecules, has the particular characteristic of being well isolated in a specific contiguous sequence. Moreover, the sequence Arg-X-Asp-Gly-Tyr-His-X-Leu-X-Thr-X-Phe is highly conserved in the loop between β2 and β3 among CMKs of different organisms (Wada, Kuzuyama et al. 2003). We decided to constrain this loop to assess the interaction properties. It is documented that, such as in a streptavidin-peptidyl ligand model (Giebel, Cass et al. 1995; Chang and Chu 2005), cyclic peptides are capable of binding more tightly than their linear counterparts. This entropy-mediated gain in binding affinity is associated to the conformational constrain (Khan, Parrish et al. 1998; Dumez, Snaith et al. 2002; Dathe, Nikolenko et al. 2004; Kang, Kim et al. 2004; Nam, Ye et al. 2004). As described in Materials and Methods, we constructed a cyclic peptide/CMK complex and, after MD, interaction energies of the complex were studied (Fig. 5B). The cyclic peptide was synthesized and the interaction between this molecule and the protein was empirically assessed through Surface Plasmon Resonance technology.

Description of the cyclic peptide binding mode in complex with CMK

Similarly to the compound, to better predict the binding mode of the cyclic peptide in complex with CMK, 2 ns of MD were calculated. Convergence was achieved after 0.3 ns. Fig. 6E shows the initial structure of this complex and Fig. 6F shows the final structure throughout the MD. The general binding mode was achieved with van der Waals contacts. Thus the non-polar parts of Lys^{76}, Leu^{84}, Met^{75}, Asp^{80}, Thr^{86} and Ala^{79} interact tightly with Arg^{3}, Ala^{4} and Asp^{5} of the cyclic peptide. These residues maintained the interaction during the MD as shown in Fig. 6F and 7C in corroboration with a strong hydrophobic interaction. Thus, considering only van der

Waals interactions, results are consistent with the fact that this peptide mimics the protein-protein contacts of the CMK homodimer interface. With respect to polar binding interactions, Asp^{80} forms a hydrogen bond with residue Arg^{3} of the cyclic peptide (Fig. 6E), although this contact was lost during MD. Nevertheless, the final structure (Fig. 6F) still showed an electrostatic interaction between these two residues. The predicted binding free energy using the MMGBSA protocol was -13.21 kcal mol^{-1}, which was also found lower than the experimental value reported in this work (-4.93 kcal mol^{-1}, see below) but in the same order as the one calculated for the compound in complex with CMK. With respect to the 'per residue' analysis, Fig. 7C shows the most remarkable residues that participate in the complex stabilization. Thus, residues Arg^{3}, Ala^{4} and Asp^{5} of the cyclic peptide contribute to the complex with -2.12 kcal mol^{-1}, -2.36 kcal mol^{-1} and -1.30 kcal/mol^{-1}, respectively, while residues Met^{75}, Lys^{76}, Ala^{79}, Asp^{80}, Leu^{84} and Thr^{86} of the protein contribute with -1.50 kcal mol^{-1}, -1.43 kcal mol^{-1}, -1.94 kcal mol^{-1}, -0.54 kcal mol^{-1}, -1.05 kcal mol^{-1} and -0.66 kcal mol^{-1}, respectively, to the binding free energy.

Binding free energy prediction of the CMK homodimer.

In the same manner as the compound/CMK and cyclic peptide/CMK complexes, the predicted binding free energy of the CMK homodimer complex was calculated using the MMGBSA protocol. The value obtained (-11.10 kcal mol^{-1}) was higher than both the compound/CMK complex (-20.61 kcal mol^{-1}) and the cyclic peptide/CMK complex (-13.21 kcal mol^{-1}). These results suggest a major affinity of both the compound and the cyclic peptide for CMK than the monomeric partner. However, MMGBSA values should be taken with caution, as they are crude estimations of binding energies.

Experimental binding analysis of the cyclic peptide/CMK complex through surface plasmon resonance

Recombinant CMK was immobilized into CM5 chip obtaining a final response of 910 resonance units (RU). To confirm the correctness of the immobilized chip, we qualitatively measured the binding of the CMK substrates into a chip: CDP-ME and ATP (data not shown). As shown in Fig. 8A, the cyclic peptide binds to CMK in a fast-on and fast-off binding system. The fast-on and fast-off binding mode was corroborated by an affinity analysis (insets of the Fig. 8). The K_D value deduced was 2.431 x 10^{-4}M, γ^2=0.588 RU^2 and the corresponding binding free energy of the cyclic peptide/CMK complex was therefore -4.93 kcal mol^{-1} (at 300 K).

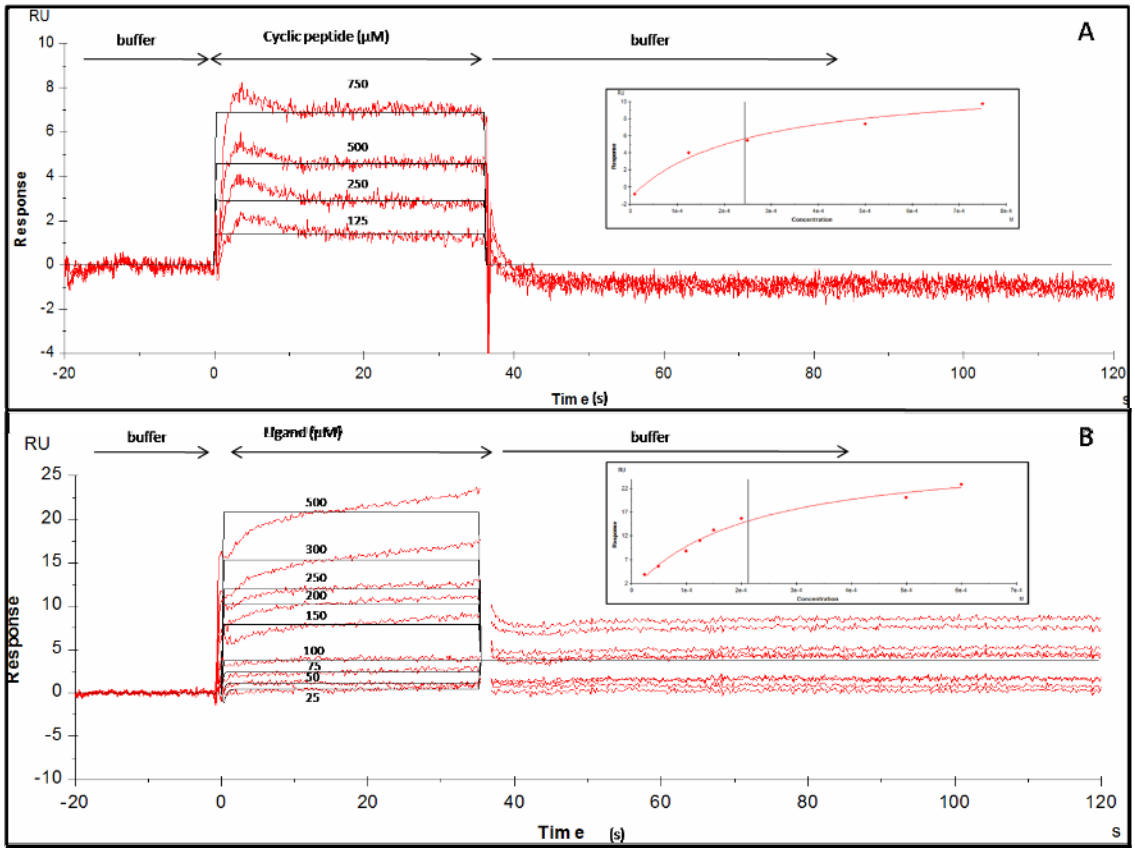


Fig. 8. A. Binding of the cyclic peptide to CMK immobilized into CM5 sensor chip. **B**. Binding of the compound to CMK immobilized into CM5 sensor chip. Representative sensorgrams (red lines) of the cyclic peptide or the ligand binding to immobilized CMK are overlaid with the calculated responses from 1:1 binding model (black lines). The inset gives the affinity analysis and the K_D values are shown.

In order to elucidate the specificity of the interaction of the cyclic peptide, we constructed a mutant of the CMK in which the Asp^{80} amino acid residue was replaced by alanine**.** A salt bridge between Arg^{21} and Asp^{80} were found as an important PPI in the crystal structure and in the MD trajectory of the homodimer. In fact, the cyclic peptide mimics the loop involving the Arg^{21} (Arg^{3} of the cyclic peptide). The mutant was also immobilized (final response 511.8 RU) into a CM5 chip and no interaction was found between the cyclic peptide and the recombinant mutated CMK Asp^{80}Ala. These results suggest a specific interaction between the cyclic peptide and the CMK and gives confidence to the proposed binding mode and binding site achieved throughout MD. It should be stressed that the per residue total binding free energies obtained from the cyclic peptide/CMK complex (Fig 7C) showed that this residue was not the most relevant one in the binding free energy of the complex, mainly because the salt bridge was broken during MD simulation. Experimental data suggested that either the binding free energy estimation for this residue was underestimated or the mutation of the Asp^{80} may distort the proper folding of the other residues involved in the interaction. However, these results suggest a specific interaction between the cyclic peptide and the CMK and gives confidence to the proposed binding site.

Furthermore, the cyclic peptide was also immobilized into a CM5 chip (final response 586RU), using an aldehyde coupling procedure recommended by the manufacturer, and the interaction was corroborated (K_D=1.14 x 10^{-5}M, γ^2=4.5, ΔG^o=-6.74 kcal mol^{-1}) (data not shown). This increase of affinity between the cyclic peptide and CMK may be attributed to the amine immobilizing protocol which binds the CMK in a random orientation (only associated to the presence of amine groups in the side chains of the residues) to the chip. This binding can

partially mask the interaction zone to the solvent, which may not happen when CMK was used as a ligand.

Experimental binding analysis of the compound/CMK complex through surface plasmon resonance

In a similar manner to the cyclic peptide, binding of the compound to CMK was assessed with CMK immobilized into a CM5 chip. The interaction was quantitatively measured at different concentrations of compound (Fig. 8B) using BIAcore100. BIAevaluation software gave values of the equilibrium dissociation constant based on the amount of ligand bound to the CMK surface sensorgrams of each ligand concentration by fitting the sensorgrams to a 1:1 Langmuir binding model. The K_D value deduced from the kinetic values was 2.121×10^{-4} M, γ^2=0.845 RU^2. Based on the K_D, the empiric binding free energy of the compound/CMK complex was found to be -5.01 kcal mol^{-1} (at 300 K).

Conclusions

In this paper, a pharmacophore-directed virtual screening protocol has been applied to discover small molecules able to bind to the PPI zone of CMK. Concretely, two additional hydrogen acceptors, which mimicked two water-mediated interactions involved in the CMK homodimer interaction, were introduced into our previous pharmacophoric hypothesis. On the whole, the new pharmacophore hypothesis was constructed by mimicking three solvent-mediated PPI, one direct hydrogen bond between monomers and three hydrophobic interactions. The addition of these water-mediated interactions (three out of the seven pharmacophoric points) highlights the importance of water molecules in the complex stabilization. This pharmacophore was introduced as hypothesis in CATALYST and a 3D database search was carried out. 639 hits were obtained and further studied with docking. In the end, a 7H-furo[3,2-g]chromen-7-one derivative was shown to be active, as assessed by SPR.

In addition, taking into account the special characteristics of the CMK homodimer PPI, a peptide corresponding to the sequence from Gln^{20} to Thr^{27} amino acid residues (the same sequence used to extract the pharmacophore) of CMK was cyclated. This cyclic peptide was also shown to bind to CMK in SPR assays.

MD studies were carried out on both the compound/CMK and the cyclic peptide/CMK complex. Predicted binding free energies, according to the MMGBSA protocol, as well as the important residues for the complex interaction, were described. Predicted binding free energies suggest a major affinity of the compounds (compound and cyclic peptide) than the CMK monomeric partner to the complex formation.

The data presented here indicate that both the constrained peptides and the pharmacophore-directed virtual screening approaches are reliable strategies for identifying molecules that bind to the CMK interface. Furthermore, this work also corroborates the possibility of including some wet spots into a pharmacophore hypothesis. These strategies can contribute to reduce time in discovering new drug compounds and point out the profitable interconnectivity between experimental assays and molecular modeling methods. Both the 7H-furo[3,2-g]chromen-7-one derivative and the designed cyclic peptide can be considered as the first step into designing drugs targeting CMK. We expect to expand this strategy to other protein of the MEP pathway of pathogenic organisms, such as malaria, tuberculosis and other infectious diseases.

Acknowledgements

The technical assistance of Marta Taulés, from the Cytometry unit, Barcelona Science Park, in the surface plasmon resonance experiments is grateful acknowledged. The Spanish Ministry of Science and Technology supported this work through the projects CTQ2006-06588/BQU and BIO2002-04419-C02-02. This work was also supported in part by the 'Generalitat de Catalunya' through the projects 2005SGR00914 and 2005SGR00204 and the 'Ministerio de Trabajo'. We are also grateful to the 'Departament d'Universitat, Recerca i Societat de la informació de la Generalitat de Catalunya i del Fons Social Europeu'.

Bibliography

Accelrys Inc., U. "CATALYST™."
Andreassi, J. L., 2nd and T. S. Leyh (2004). "Molecular functions of conserved aspects of the GHMP kinase family." Biochemistry **43**(46): 14594-601.
Andrews, M. J., C. McInnes, et al. (2004). "Design, synthesis, biological activity and structural analysis of cyclic peptide inhibitors targeting the substrate recruitment site of cyclin-dependent kinase complexes." Org Biomol Chem **2**(19): 2735-41.
Bach II, A., C. Eyermann, et al. (1994). "Structural studies of a Family of high affinity ligands for GPIIb/IIIa " J Am Chem Soc **116**(8): 3207-3219.
Berendsen, H., J. Postma, et al. (1984). "Molecular dynamics with coupling to an external bath." J Chem Phys **81**(8): 3684.
Bonanno, J. B., C. Edo, et al. (2001). "Structural genomics of enzymes involved in sterol/isoprenoid biosynthesis." Proc Natl Acad Sci U S A **98**(23): 12896-901.
Bork, P., C. Sander, et al. (1993). "Convergent evolution of similar enzymatic function on different protein folds: the hexokinase, ribokinase, and galactokinase families of sugar kinases." Protein Sci **2**(1): 31-40.
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**: 248-54.
Case, D. A., T. E. Cheatham, 3rd, et al. (2005). "The Amber biomolecular simulation programs." J Comput Chem **26**(16): 1668-88.
Cochran, A. G. (2000). "Antagonists of protein-protein interactions." Chem Biol **7**(4): R85-94.
Cooper, M. A. (2002). "Optical biosensors in drug discovery." Nat Rev Drug Discov **1**(7): 515-28.
Chang, Y. P. and Y. H. Chu (2005). "Using surface plasmon resonance to directly determine binding affinities of combinatorially selected cyclopeptides and their linear analogs to a streptavidin chip." Anal Biochem **340**(1): 74-9.
Darden, T., D. York, et al. (1993). "Particle mesh Ewald: An $N\cdot\log(N)$ method for Ewald sums in large systems." J Chem Phys **98**(12): 10089.
Dathe, M., H. Nikolenko, et al. (2004). "Cyclization increases the antimicrobial activity and selectivity of arginine- and tryptophan-containing hexapeptides." Biochemistry **43**(28): 9140-50.
Dumez, E., J. S. Snaith, et al. (2002). "Synthesis of macrocyclic, potential protease inhibitors using a generic scaffold." J Org Chem **67**(14): 4882-92.
Eisenreich, W., M. Schwarz, et al. (1998). "The deoxyxylulose phosphate pathway of terpenoid biosynthesis in plants and microorganisms." Chem Biol **5**(9): R221-33.
Gabrielsen, M., C. S. Bond, et al. (2004). "Hexameric assembly of the bifunctional methylerythritol 2,4-cyclodiphosphate synthase and protein-protein associations in the deoxy-xylulose-dependent pathway of isoprenoid precursor biosynthesis." J Biol Chem **279**(50): 52753-61.
Gabrielsen, M., F. Rohdich, et al. (2004). "Biosynthesis of isoprenoids: a bifunctional IspDF enzyme from Campylobacter jejuni." Eur J Biochem **271**(14): 3028-35.
Giebel, L. B., R. T. Cass, et al. (1995). "Screening of cyclic peptide phage libraries identifies ligands that bind streptavidin with high affinities." Biochemistry **34**(47): 15430-5.
Gimenez-Oya, V., O. Villacanas, et al. (2009). "Mimicking direct protein-protein and solvent-mediated interactions in the CDP-methylerythritol kinase homodimer: a pharmacophore-directed virtual screening approach." J Mol Model.
Humphrey, W., A. Dalke, et al. (1996). "VMD: visual molecular dynamics." J Mol Graph **14**(1): 33-8, 27-8.
Jakalian, A., D. B. Jack, et al. (2002). "Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation." J Comput Chem **23**(16): 1623-41.
Johnsson, B., S. Lofas, et al. (1991). "Immobilization of proteins to a carboxymethyldextran-modified gold surface for biospecific interaction analysis in surface plasmon resonance sensors." Anal Biochem **198**(2): 268-77.
Jorgensen, W., J. Chandrasekhar, et al. (1983). "Comparison of simple potential functions for simulating liquid water." J Chem Phys **79**: 926.
Kang, J. H., S. Y. Kim, et al. (2004). "Macrocyclic diacylglycerol-bis-lactones as conformationally constrained analogues of diacylglycerol-lactones. Interactions with protein kinase C." J Med Chem **47**(16): 4000-7.
Karlsson, R. (2004). "SPR for molecular interaction analysis: a review of emerging application areas." J Mol Recognit **17**(3): 151-61.
Khan, A. R., J. C. Parrish, et al. (1998). "Lowering the entropic barrier for binding conformationally flexible inhibitors to enzymes." Biochemistry **37**(48): 16839-45.
Kollman, P. A., I. Massova, et al. (2000). "Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models." Acc Chem Res **33**(12): 889-97.

Kuzuyama, T., S. Takahashi, et al. (2000). "Characterization of 1-deoxy-D-xylulose 5-phosphate reductoisomerase, an enzyme involved in isopentenyl diphosphate biosynthesis, and identification of its catalytic amino acid residues." J Biol Chem **275**(26): 19928-32.

Lange, B. M. and R. Croteau (1999). "Isopentenyl diphosphate biosynthesis via a mevalonate-independent pathway: isopentenyl monophosphate kinase catalyzes the terminal enzymatic step." Proc Natl Acad Sci U S A **96**(24): 13714-9.

Leroux, V., N. Gresh, et al. (2006). "Role of water molecules for binding inhibitors in the SH2 domain of Grb2: A molecular dynamics study." teochem **806**(1-3): 51-66.

Lichtenthaler, H. K. (1999). "The 1-Deoxy-D-Xylulose-5-Phosphate Pathway Of Isoprenoid Biosynthesis In Plants." Annu Rev Plant Physiol Plant Mol Biol **50**: 47-65.

Luttgen, H., F. Rohdich, et al. (2000). "Biosynthesis of terpenoids: YchB protein of Escherichia coli phosphorylates the 2-hydroxy group of 4-diphosphocytidyl-2C-methyl-D-erythritol." Proc Natl Acad Sci U S A **97**(3): 1062-7.

Lloyd, D. G., A. T. Garcia-Sosa, et al. (2004). "The effect of tightly bound water molecules on the structural interpretation of ligand-derived pharmacophore models." J Comput Aided Mol Des **18**(2): 89-100.

Marsters, J. C., Jr., R. S. McDowell, et al. (1994). "Benzodiazepine peptidomimetic inhibitors of farnesyltransferase." Bioorg Med Chem **2**(9): 949-57.

Miallau, L., M. S. Alphey, et al. (2003). "Biosynthesis of isoprenoids: crystal structure of 4-diphosphocytidyl-2C-methyl-D-erythritol kinase." Proc Natl Acad Sci U S A **100**(16): 9173-8.

Morgan, B., D. Holland, et al. (1994). "Structure-based design of an inhibitor of the zinc peptidase Thermolysin." J Am Chem Soc **116**(8): 3251-3260.

Nam, N. H., G. Ye, et al. (2004). "Conformationally constrained peptide analogues of pTyr-Glu-Glu-Ile as inhibitors of the Src SH2 domain binding." J Med Chem **47**(12): 3131-41.

Obiol-Pardo, C., J. M. Granadino-Roldan, et al. (2008). "Protein-protein recognition as a first step towards the inhibition of XIAP and Survivin anti-apoptotic proteins." J Mol Recognit **21**(3): 190-204.

Rohdich, F., J. Wungsintaweekul, et al. (2000). "Biosynthesis of terpenoids: 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase from tomato." Proc Natl Acad Sci U S A **97**(15): 8251-6.

Rohmer, M. (1999). "The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants." Nat Prod Rep **16**(5): 565-74.

Rubio-Martinez, J., M. Pinto, et al. (2005). "Dock_dyn: a program for fast molecular docking using molecular dynamics information ".

Ryckaert, J., G. Ciccotti, et al. (1977). "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes." J Comp Phys **23**(3): 327-341.

Sacchettini, J. C. and C. D. Poulter (1997). "Creating isoprenoid diversity." Science **277**(5333): 1788-9.

Samsonov, S., J. Teyra, et al. (2008). "A molecular dynamics approach to study the importance of solvent in protein interactions." Proteins **73**(2): 515-25.

Spencer, R. W. (1998). "High-throughput screening of historic collections: observations on file size, biological targets, and file diversity." Biotechnol Bioeng **61**(1): 61-7.

Tsui, V. and D. A. Case (2000). "Theory and applications of the generalized Born solvation model in macromolecular simulations." Biopolymers **56**(4): 275-91.

Wada, T., T. Kuzuyama, et al. (2003). "Crystal structure of 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase, an enzyme in the non-mevalonate pathway of isoprenoid synthesis." J Biol Chem **278**(32): 30022-7.

Wang, J., R. M. Wolf, et al. (2004). "Development and testing of a general amber force field." J Comput Chem **25**(9): 1157-74.

Wang, R., L. Lai, et al. (2002). "Further development and validation of empirical scoring functions for structure-based binding affinity prediction." J Comput Aided Mol Des **16**(1): 11-26.

Weiser, J., P. Shenkin, et al. (1999). "Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO)" J Comp Chem **20**: 217-230.

Wells, J. A. and C. L. McClendon (2007). "Reaching for high-hanging fruit in drug discovery at protein-protein interfaces." Nature **450**(7172): 1001-9.

Zang, X., Z. Yu, et al. (1998). "Tight-binding streptavidin ligands from a cyclic peptide library." Bioorg Med Chem Lett **8**(17): 2327-32.

Zobel, K., L. Wang, et al. (2006). "Design, synthesis, and biological activity of a potent Smac mimetic that sensitizes cancer cells to apoptosis by antagonizing IAPs." ACS Chem Biol **1**(8): 525-33.

Análisis del estado oligomérico de la CDP-ME quinasa de *E.coli*: la formación del homodímero no es necesaria para la actividad enzimática

Victor Giménez-Oya[1], Meritxell Antolin-Llovera[1], Santiago Imperial[1#]

[1] *Dept. de Bioquímica i Biología Molecular, Universitat de Barcelona (UB), Avda Diagonal 645, E-08028 Barcelona, Spain.*
[2] *Dept. de Química Física, Universitat de Barcelona (UB) and the Institut de Recerca en Química Teòrica i Computacional (IQTCUB), Martí i Franquès 1, E-08028 Barcelona, Spain.*
Current Address: Intelligent Pharma, S.L. Barcelona Science Park, C/ Baldiri Reixac 10, E-08028, Barcelona, Spain.

Running title: Oligomerización de la CDP-ME quinasa de *E.coli*

Palabras clave**:** CDP-ME quinasa, péptido cíclico, diseño de drogas, dispersion dinámica de la luz, formación del homodímero, metileritritol fosfato, mevalonato, estado oligomérico, cromatografía de exclusion molecular.

Resumen

El cuarto paso de la ruta MEP de biosíntesis de isoprenoides está catalizado por la CDP-ME quinasa (CMK) que pertenece a la familía GHMP de proteínas. Esta enzima ha sido aislada y caracterizada de un gran número de especies bacterianas y, a pesar de ello, su estado oligomérico presenta algunas dudas. La información cristalográfica proveniente de diferentes especies presenta controversia ya que ha sido caracterizada como monomérica en bacterias termofílicas y dimérica en *E.coli*. En este trabajo se ha reinvestigado el estado oligomérico de la CMK recombinante de *E.coli*, la cual, presentó una mezcla de especies oligoméricas siendo la monomérica la predominante según ensayos de cromatografía de exclusión molecular, geles nativos de poliacrilamida y dispersión dinámica de luz. Estas especies oligoméricas mostraron un equilibrio dinámico según experimentos de cromatografía de exclusión molecular y unión química cruzada. Las especies monoméricas de la CMK de *E.coli* presentaron actividad enzimática en unos valores similares a los descritos para la mezcla de especies. Además, moléculas especialmente diseñadas para impedir las interacciones proteína-proteína en la formación del dímero de la CMK de *E.coli* no mostraron un efecto significativo sobre la actividad enzimática. Todos estos resultados sugieren que las especies monoméricas de la CMK de *E.coli* son enzimáticamente activas y que su estado dimérico no es necesario para la actividad enzimática *in vitro*.

Analysis of the oligomeric state of *E.coli* CDP-ME kinase: Homodimer formation is not necessary for enzyme activity

Victor Giménez-Oya[1], Meritxell Antolin-Llovera[1], Santiago Imperial[1#]

[1] *Dept. de Bioquímica i Biología Molecular, Universitat de Barcelona (UB), Avda Diagonal 645, E-08028 Barcelona, Spain.*

Running title: Oligomerization of the *E.coli* CDP-ME kinase

Key words**:** CDP-ME kinase, cyclic peptide, drug design, dynamic light scattering, homodimer formation, methylerythritol phosphate, mevalonate, oligomeric state, size-exclusion chromatography.

ABSTRACT

The fourth step of the MEP pathway of isoprenoid biosynthesis is catalyzed by CDP-ME kinase (CMK), a member of the GHMP kinase family of proteins. Although the enzyme has been isolated from a number of bacterial species and charactherized its oligomeric state is still ill-defined. Crystallographic data for the enzyme from different organisms are controversial with CMK from thermophilic bacteria being reported as monomeric and *E.coli* CMK described as dimeric enzyme. In this paper we have re-investigated the oligomeric state of the recombinant CMK from *E.coli* and showed that it exists as a mixture of oligomeric species being the monomeric the most prevalent according to size-exclusion chromatography, native-PAGE and dynamic light scattering analysis. According to size-exclusion chromatography and chemical cross-linking experiments a dynamic equilibrium exists among the different oligomeric species of the enzyme. In addition, the monomeric *E.coli* CMK was shown to be active and exhibited an enzyme activity similar to the previously described for the enzyme. Molecules that were specifically designed to disrupt the protein-protein interaction of the *E.coli* CMK homodimer displayed no inhibitory effects on the enzyme activity. These results suggest that *E.coli* CMK monomeric species are enzymatically active and that a homodimeric state is not necessary for the enzyme activity.

INTRODUCTION

Isoprenoids are one of the largest families of compounds in nature (Sacchettini and Poulter 1997) comprising widespread biological functions. Isoprenoids are derived from two universal five-carbon building blocks, isopentenyl diphosphate (IPP) and dymethylallyl diphosphate (DMAPP). These five-carbon building blocks are synthesized from the mevalonate (MVA) pathway or the 2-*C*-methyl-D-erythritol 4-phosphate (MEP). The MVA pathway is present in animals, yeast and some bacteria, while the MEP pathway is present in eubacteria, cyanobacteria and protozoa. Interestingly, both MVA and MEP are present in plants (for reviews, (Eisenreich, Schwarz et al. 1998; Lichtenthaler 1999; Rohmer 1999)). The MEP pathway, starting from glyceraldehyde 3-phosphate and pyruvate, is used by many pathogens such as *Mycobacterium tuberculosis* (Boucher and Doolittle 2000) and *Plasmodium falciparum* (Jomaa, Wiesner et al. 1999) and is absent in humans. Efforts to combat pathogenic diseases are hampered with the emergence of multidrug resistance (MDR) in pathogenic organisms. The identification of new targets for drug-design is a mandatory way to overcome the MDR problem and, given the charactheristics of the MEP pathway, it becomes an attractive druggable pathway.

The fourth step of the MEP pathway is an ATP-dependent phosphorylation of the 2-OH group of 4-(cytidine 5-diphospho)-2*C*-methyl-D-erythritol (CDP-ME) into CDP-ME 2-phosphate (Kuzuyama, Takahashi et al. 2000; Luttgen, Rohdich et al. 2000; Rohdich, Wungsintaweekul et al. 2000) through the action of the CDP-ME kinase (CMK or *IspE*). CMK is a Mg-cation dependent kinase belonging to the GMHP kinase superfamily (Bork, Sander et al. 1992; Bork, Sander et al. 1993) (namely for galactokinase, mevalonate kinase, homoserine kinase and phosphomevalonate kinase members). The fold of this group of kinases consist of two α+β domains, with the active site located between the two domains. In this group, the N-terminal domain contains two β sheets and four α helices and the C-terminal domain is a ferredoxin-like core with four additional α-helices (Cheek, Zhang et al. 2002). This family of kinases has three conserved motifs (Fig. 1). Motif I is located in the N terminus and contains seven well conserved amino acid residues. Motif II (or P-loop) is involved in the nucleotide binding and has a conserved PXXXGSSAA sequence, which is structurally different from the prototypical Walker A motif and displays a different orientation of the ATP. Motif III is a small glycine-rich loop which interact with the substrate that is phosphorylated (Andreassi and Leyh 2004). Although the sequence identity of proteins within this family of kinases is relatively low (10-20% among those for which structural information is available in the Protein Data Bank, PDB), they share a very similar structural scaffold (Cα RMSDs range from 2.6 to 4.0 Å) (Andreassi and Leyh 2004), suggesting the same evolutionary origin from a small kinase protein.

Crystal structure of *Escherichia coli* CMK (EcCMK) reveals a dimeric protein with only 4% of the surface area involved in the dimer interface (Miallau, Alphey et al. 2003). The dimer is stabilized by hydrophobic interactions between the side chains of Tyr25 and Lys76 of the partner, with hidrogen bonds donated from the amide of Gly87 to the carbonyl of Ala22 of the partner, salt bridges between Arg21 of one subunit with Asp-80 of the partner (Miallau, Alphey et al. 2003) and numerous solvated-mediated interactions (Gimenez-Oya, Villacanas et al. 2009). Furthermore, the homodimeric state of the EcCMK was reported by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, size-exclusion chromatography (Miallau, Alphey et al. 2003) and in analytical centrifugation analysis(Gabrielsen, Rohdich et al. 2004). In this study we have re-examined the oligomeric state of the recombinant EcCMK by different experimental methods. We aport evidence that EcCMK homodimerization is not required for enzymatic activity.

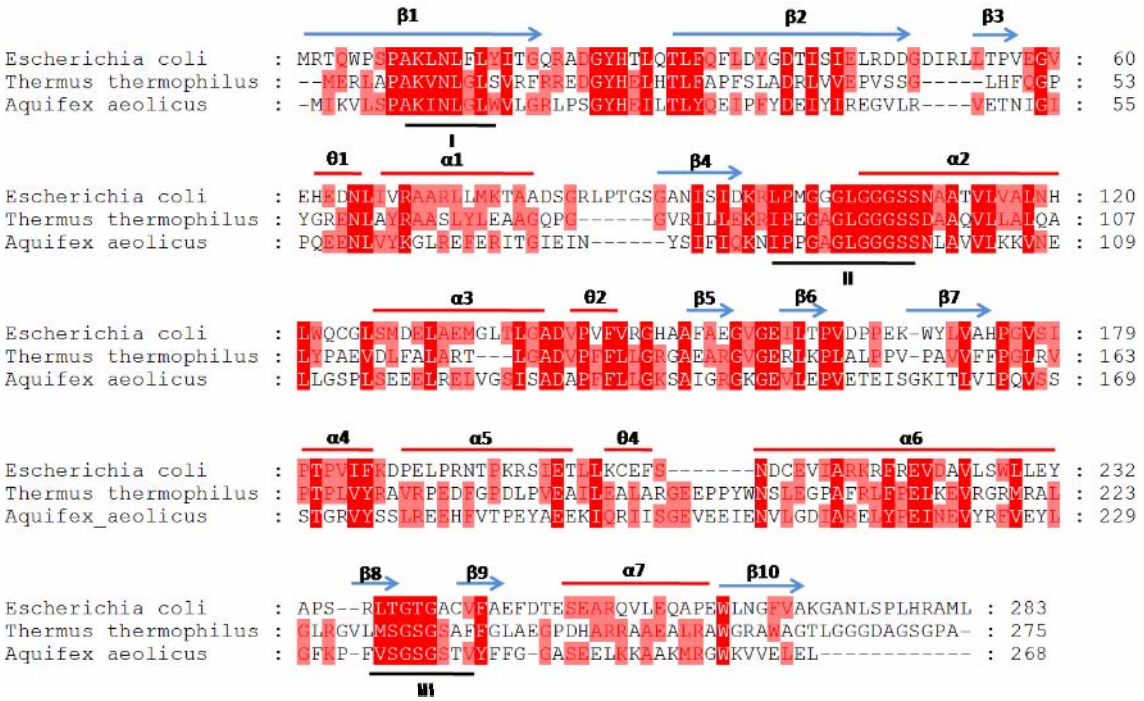


Fig 1. Amino acid sequence alignment of CMK from *Escherichia coli*, *Aquifex aeolicus* and *Thermus thermophilus* with elements of the secondary structure indicated (α-helices, β-sheets and 3_{10}-helices). The three conserved GHMP motifs are underlined. Alignments were performed with the ClustalW (http://www.ebi.ac.uk/Tools/clustalw2/index.html) and arranged using the GeneDoc program.

Materials and Methods

Materials.

All chemicals were purchased from Sigma unless stated otherwise. MEP was from Echelon Biosciences. CDP-ME was synthesized in our laboratory as previously described (Bernal, Mendez et al. 2005). Plasmid pQE30 expression vector was from Qiagen. The *E. coli* DH5α and M15 [pREP4] strains were purchased from Novagen. Superdex 75, Superdex 200 and HiTrap chelating HP colums were from GE Healthcare.

Cloning of the *ispE* gene from *E. coli*

The coding region of the *IspE* gene (815bp) from *E.coli* genome was amplified by PCR and subcloned into the pQE30 plasmid as previously described (Bernal, Mendez et al. 2005). The construct obtained pQE30-*EcIspE* allows the production of a recombinant protein in *E. coli* with a N-terminal extension including six histidine amino acid residues.

Preparation of recombinant *E. coli* CDP-ME kinase

The pQE30-*EcIspE* construct was used to transform competent *E. coli* M15 [pREP4] cells. Bacterial cells were grown in LB medium supplemented with 100 μg/ml ampicillin and 25 μg/ml kanamycin. The culture was incubated at 37ºC with vigorous shaking. At an optical density (600 nm) of 0.5~0.6, IPTG was added to a final concentration of 1mM, and the culture was further incubated at 25ºC for 12 h. Cells were harvested by centrifugation and cell pellets were thawed and re-suspended in lysis buffer (40 mM Tris-HCl, pH 8.0,100 mM NaCl,10 mM imidazole) containing 1 mg/ml lysozyme. The mixture was incubated at 4ºC for 20 min and subsequently sonicated (5 x 30s bursts with 1min cooling in melting ice). The crude lysate was centrifuged at

13,000g at 4ºC for 20 min to remove cellular debris. The supernatant was loaded on a HiTrap chelating HP column (GE healthcare) pre-equilibrated with lysis buffer. The recombinant His6-CDP-ME kinase was eluted with an imidazole concentration gradient (from 10 mM to 500mM). The fractions containing the recombinant enzyme were 95% pure as judged by SDS–polyacrylamide gel electrophoresis (SDS-PAGE) followed by Coomassie brilliant blue R-250 staining. CMK preparations were desalted by filtration on disposable Sephadex G-25 columns (GE Healthcare) equibrated in 40mM Tris–HCl buffer, pH 8.0, containing 0.1M NaCl, distributed into aliquots, snap-frozen with liquid N_2 and stored at -80ºC until use. Protein concentration was measured by the method of Bradford (Bradford 1976) using bovine serum albumin as standard.

MALDI-TOF mass spectrometry

Samples were mixed with a CHCA matrix (1:1) (α-ciano-4-hydroxycinnamic acid, Aldrich) solution (10mg/mL) in acetonitrile/H20 (1/1, v:v). One µl of the mix was overlaid into the MALDI-TOF plaques and were dried using the dried doplet method (Karas and Hillenkamp 1988). MALDI-TOF mass spectra were recorded in a 4700 Proteomics Analyzer instrument (Applied Biosystems). Acquisition of mass spectra was performed in the MS reflector positive-ion mode. Typical parameters were set to source and grid voltages 20 and 14kV, respectively, power laser from 5200 to 5800, signal/noise threshold 5, and noise windows width of 50.

Native gel electrophoresis

The native molecular mass of EcCMK was evaluated using nondenaturing electrophoresis. Protein standards were α-lactalbumin (14.2 kDa), carbonic anhydrase from bovine erythrocytes (29 kDa), chicken egg albumin (45 kDa), bovine serum albumin (monomer, 66 kDa, and dimer, 132 kDa) and urease from Jack bean (trimer, 272kDa, hexamer, 542kDa). Proteins were electrophoresed on a set of nondenaturing gels with polyacrylamide concentrations of 5, 6, 7, 8, 9, or 10%, respectively. The tracking dye was xylidine Ponceau. Proteins were stained with Coomassie blue R-250. The relative mobility (Rf) of the protein was determined by the distance of the protein migration relative to the tracking dye migration. A value of 100 [log (Rf × 100)] for each protein was plotted against the percent concentration of polyacrylamide (Shi, Feng et al. 2007). The line of these values was determined obtaining the slope for each protein (mass). The negative slope obtained for each protein was then plotted against the native molecular masses of protein standards to produce a linear log/log plot from which the molecular mass of EcCMK was extrapolated.

Size exclusion chromatography

The oligomeric state of the recombinant EcCMK was studied by analytical size exclusion chromatography (Superdex 75 and Superdex 200 columns (10 x 300 mm), GE healthcare) using a FPLC system (AKTA system, GE healthcare) at a flow rate of 0.4 ml/min at room temperature. Recombinant EcCMK was loaded on both Superdex 75 and Superdex 200 size exclusion chromatography columns (GE healthcare) pre-equilibrated with buffer (40 mM Tris-HCl, pH 8.0, 100 mM NaCl) and calibrated with aldolase (158 kDa), conalbumin (75kDa), ovalbumin (45 kDa), carbonic anhydrase (29 kDa) and ribonuclease (13.5 kDa) as standard proteins. The relative amount of each oligomer was measured by integration of the area under the curve (absorbance at 280 nm).

Dynamic light scattering

The oligomerization state of EcCMK was also analyzed by Dynamic Light Scattering (DLS) experiment using a Zetasizer Nano-S instrument with a termostated cell at 4ºC (Malvern Instruments). DLS is used to measure hydrodynamic sizes, polydispersion and aggregation effects of protein samples.

Formaldehyde cross-linking experiments

The buffer of the purified EcCMK was exchanged for a non-nucleophile buffer (50mM Hepes pH 8.0, 100mM NaCl) by the protein desalting spin columns (Pierce). Formaldehyde was added to reach a final concentration of 50mM and incubated for various times at room temperature. Aliquots were mixed with SDS-PAGE loading buffer to stop the cross-linking reaction and boiled for 5 min. Samples were analysed by 10% SDS-PAGE.

Enzyme activity determinations

CMK catalyzes the phosphorylation of CDP-ME to form CDP-MEP using ATP as a phosphate donor. Enzyme activity was assayed by using a previously described spectrophotometric assay (Bernal, Mendez et al. 2005). In the method, the product ADP is used as a substrate by a coupled enzyme system formed by pyruvate kinase and lactate dehydrogenase (PK/LDH). Reaction progress can be followed by a decrease in absorbance at 340nm due to the oxidation of NADH in the last coupled reaction step. Enzyme assays were conducted at 25ºC in a 96-well plate system The standard assay system contained 100 mM Tris-HCl, pH 8.5, 20 mM KCl, 10 mM $MgCl_2$, 0.45 mM NADH, 4 mM phosphoenolpyruvate, 1 mM CDP-ME, 8 U pyruvate kinase, 9 U lactate dehydrogenase and 1 μg EcCMK. The assay mixtures were incubated at 25ºC, the reaction was started adding ATP to 2mM and monitored photometrically at 340 nm using a microtiter plate reader (BenchMark plus, Bio-Rad).

To test the inhibitory effect of different compounds on the EcCMK activity, assay mixtures were prepared as described above and the compounds were added to the assay solutions (final concentration varied from 10-1000 μM). The reaction was started by addition of ATP and monitored at 340 nm. Five and six replicates were done at each concentration for each compound (two small molecules and a cyclic peptide).

Drawings

Figures depicting complex structures were constructed with the Visual Molecular Dynamics (VMD) molecular graphics software (Humphrey, Dalke et al. 1996). Cα superimpositions were done with the root mean square deviation (RMSD) calculator tool implemented in VMD.

Results

Overexpression in *E. coli* and purification of EcCMK.

The recombinant plasmid pQE30-*EcIspE* was used for the overexpression of EcCMK in *E. coli* M15 [pREP4] cells. To obtain high levels of soluble EcCMK induction was carried out with 1mM IPTG at 25ºC for 12 h. The N-terminal fusion of a six-histidine tag allowed the purification of the recombinant EcCMK to homogeneity in one step by nickel affinity chromatography. The overall yield of the purified protein was 21mg from 1 L culture medium. On SDS-PAGE, the purified EcCMK appeared as a single band with an apparent molecular mass of 31kDa (Fig. 2 left). The molecular mass was more accurately determined by mass spectrometry (Fig. 2 right). The measured mass is 34267.7 Da, which is consistent with the theoretical mass prediction of the protein (the calculated mass, 30973 Da plus the N terminal tag that includes the 6 histidine amino acid residues).

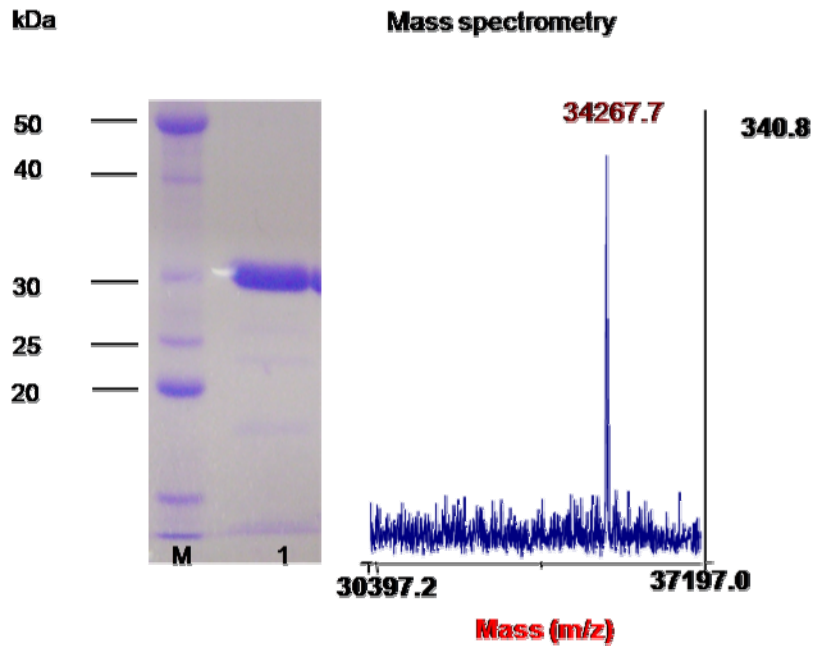


Fig 2. Determination of molecular mass by SDS-PAGE (left) or mass spectrometry (right). Lane M, molecular mass marker; lane 1, purified recombinant EcCMK.

Determination of the oligomerization state of EcCMK.

Native gel electrophoresis

The quaternary structure of EcCMK was analysed by nondenaturing polyacrylamide gel electrophoresis. As shown in Fig 3A, EcCMK preparations showed by Native-PAGE one major band with a slope of 5.42 in the set of nondenaturing gels, which corresponds to a molecular mass of 31 kDa. Only weak traces of upper higher molecular weight bands were found in 10% polyacrilamide nondenaturing gels. Except this, nondenaturing gels indicated that the most of the recombinant EcCMK migrated like a monomeric protein.

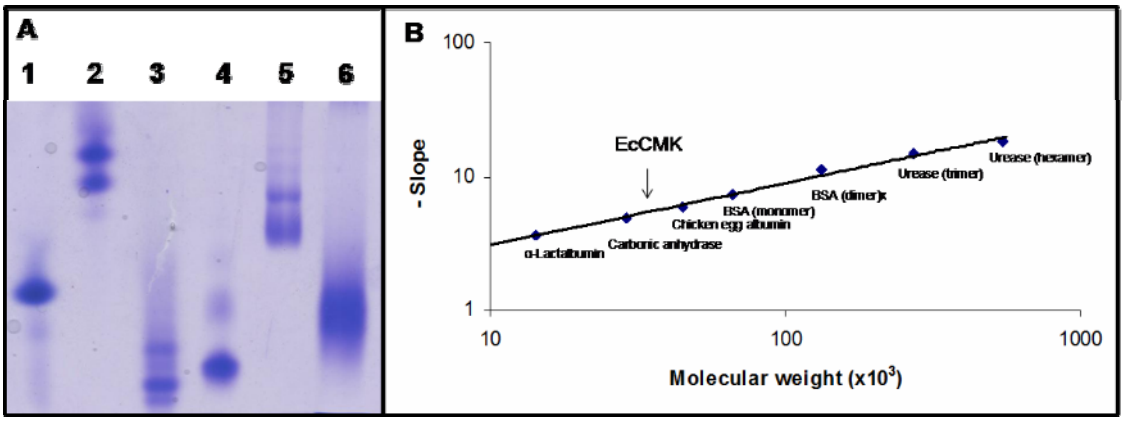


Fig 3. Estimation of the native molecular weight of EcCMK. (A) EcCMK and standard proteins were analyzed by polyacrylamide gel electrophoresis under nondenaturing conditions in gels of different polyacrylamide concentration. A nondenaturing electrophoresis gel with 6% polyacrylamide is presented. Lane 1 α-lactalbumin (14.2 kDa), lane 2 carbonic anhydrase (29 kDa), lane 3 chicken egg albumin (45 kDa), lane 4 bovine serum albumin (monomer, 66 kDa, and dimer, 132 kDa), lane 5 urease (trimer, 272kDa, hexamer, 542kDa) and lane 6, EcCMK. (B) A plot of the negative slopes versus the known molecular weights of the standards, from which the molecular weight of EcCMK was determined to be 31kDa.

Size-exclusion chromatography

Recombinant EcCMK was also analysed by size-exclusion chromatography through Superdex 200 and Superdex 75. Fig. 4.A shows a representative chromatogram obtained by Superdex 200 chromatography. Two major peaks eluting at 15.4 and 13.5 ml are shown which correspond to molecular masses of 33.5 and 76 kDa and which are consistent with monomeric and dimeric species, respectively. Interestingly, the major peak corresponded to the monomeric form of the enzyme. In all chromatographic separations performed peaks that would correspond to higher oligomeric species were detected though at low amounts. The proportion of these minor peaks was shown to inversely correlate with the protein concentration of the EcCMK loaded to the column. The relative amount of each oligomer was measured and the monomer/dimer ratio was shown to vary from 2.5 to 7.5. When fractions of the monomeric species were collected, kept at 4ºC overnight and reloaded onto a Superdex200 column, a substantial amount of the protein was shown to elute as a dimeric species (Fig. 4B). Similarly, a fraction of the dimeric EcCMK samples also eluted as a monomeric species after being collected, kept at 4ºC overnight and reloaded (Fig.4C). These results indicate that monomeric and dimeric species are interchangeable and suggest that they are in a dynamic equilibrium as described for other proteins (D'Autreaux, Pecqueur et al. 2007).

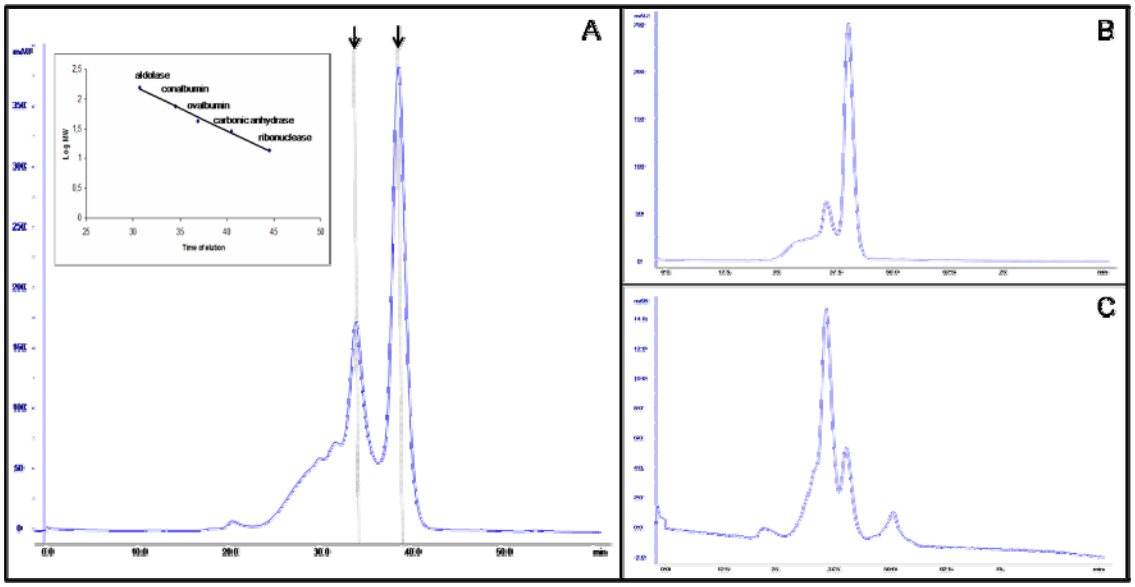


Fig 4. A.Analysis of EcCMK oligomeric state by size-exclusion chromatography. Two major peaks are indicated by arrows in the chromatogram. Recombinant EcCMK was loaded onto a Superdex 200 (10 x 300mm) pre-equilibrated with 40 mM Tris-HCl, pH 8.0, buffer containing 100 mM NaCl at a flow rate of 0,4ml/min. The inset shows the plot of molecular weight versus elution time of the following protein standards: aldolase (158 kDa), conalbumin (75kDa), ovalbumin (45 kDa), carbonic anhydrase (29 kDa) and ribonuclease (13.5 kDa). B. Fractions of the peak corresponding to the monomeric EcCMK were kept at 4ºC overnight and reloaded onto a Superdex 200. The chromatogram shows the appearance of a peak corresponding to the dimeric EcCMK size. C. Fractions of the peak corresponding to the dimeric EcCMK kept at 4ºC overnight and reloaded onto a Superdex 200. The chromatogram shows the appearance of a peak corresponding to a monomeric state of the enzyme.

Dynamic light scattering

In dynamic light scattering (DLS) experiments most of the recombinant EcCMK present in the purified preparations was shown to be in monomeric form although samples showed a high polidispersity. After separation by size-exclusion chromatography fractions corresponding to the monomeric EcCMK (15.4ml-peak) appeared as monomeric and monodisperse whereas fractions from the 13.5ml-peak presented polidispersity.

Cross-linking experiments

Chemical cross-linking carried out with formaldehyde and subsequent SDS-PAGE was used to assess the oligomeric state of the EcCMK in solution (Fig.5). In the absence of the cross-linking reagent, EcCMK appears as a major band at 34 kDa. After 15 minutes of reaction, a thin band of about 70 kDa, corresponding to the dimeric size of the EcCMK, was observed although most of the protein still remained as a 34 kDa species. As the cross-linking reaction was progressing other higher oligomeric species appeared although their cross-linking rate was found relatively slow.

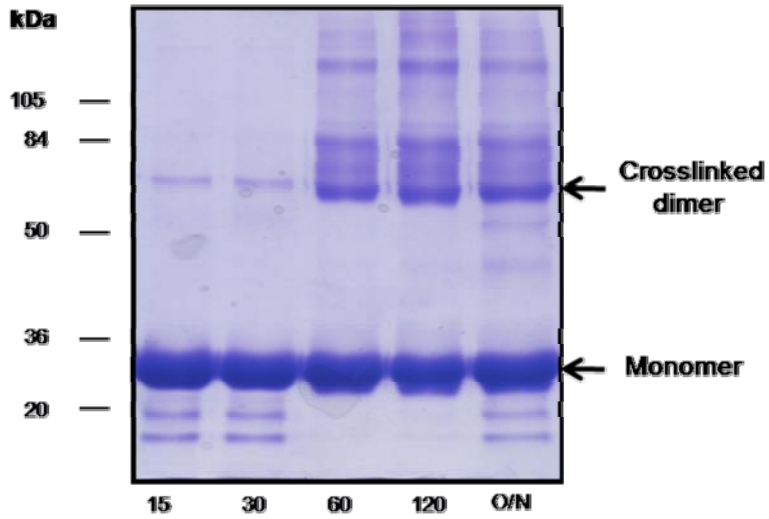


Fig 5. Formaldehyde cross-linking experiment. EcCMK was cross-linked with 50mM formaldehide for 15, 30, 60, 120 minutes and overnight, and then analyzed by SDS-PAGE. Comassie staining reveals a progressive transition of monomeric to dimeric or higher-oligomeric species.

Enzymatic activity determinations

Determination of the CDP-ME kinase activity of the monomeric species of EcCMK

Purified EcCMK preparations were separated into monomeric and dimeric species by Superdex75 size-exclusion chromatography and assayed for enzyme activity. Two peaks corresponding to a molecular mass of 36.4 and 62.8 kDa, respectively, were eluted. The major peak corresponded to the monomeric species and accounted for 83% of the total protein. Protein concentration from fractions corresponding to the monomeric species were measured by the method of Bradford (Bradford 1976). The CDP-ME kinase activity of the major peak was measured spectrophotometrically using a PK/LDH coupled assay (Bernal, Mendez et al. 2005). An activity of 3.592 (±0.008) µmoles min^{-1} mg^{-1} was obtained for these monomeric enzyme preparations which was similar to the specific activity previously reported for the purified recombinant EcCMK described by Bernal et al 2005 containing all the different oligomeric species of the enzyme. It is important to note that the fractions used to determine enzyme activity of the monomeric EcCMK were also analyzed by DLS at equal timepoints and were shown to be monomeric and monodisperse by this thecnique. These results demonstrate that monomeric EcCMK is enzymatically active and suggest that no oligomeric forms are needed for enzymatic activity.

Inhibitory assays of molecules designed to disrupt the protein-protein interactions of the EcCMK homodimer complex

The effect of small molecules designed to disrupt the protein-protein interaction zone of the EcCMK homodimer complex on the CMK enzyme activity was also studied. Two small molecules derived from a pharmacophore-directed virtual screening approach and one cyclic peptide which mimicked the protein-protein interaction (PPI) zone of the EcCMK homodimer were assayed (Fig 6). These compounds were designed to disrupt the protein-protein interactions of the EcCMK homodimer. Compound A is a small molecule and was obtained through a virtual screening approach including one protein-protein interaction, three hydrophobic interactions and one solvent-mediated interaction into the pharmacophore hypothesis. Compound B is also a small molecule and was obtained improving the previous pharmacophoric hypothesis to include two more solvent-mediated interactions.

Compound C is a cyclic peptide that mimicks the loop between amino acid residues 20 and 27 of the EcCMK implicated in the interaction of the homodimer complex (Miallau, Alphey et al. 2003).

These three molecules were shown to bind to recombinant EcCMK according to intensity-fading MALDI-TOF MS (Compound A) (Gimenez-Oya, Villacanas et al. 2009) or surface plasmon resonance (Compound B and Compound C) (Gimenez-Oya 2009). The predicted binding free energies for the complex of either compound B or C were lower than the predicted value for the CMK homodimer complex. Nevertheless, as shown in fig.7, none of these compounds showed an inhibitory effect on the CMK activity. These results reinforce the hypothesis that EcCMK homodimerization is not required for enzymatic activity.

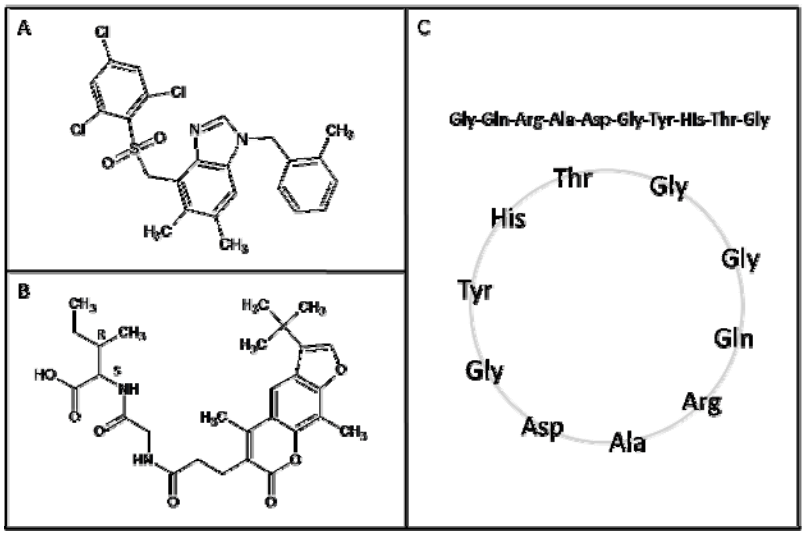


Fig 6. A. 2D structure of compound A. B. 2D structure of compound B. C. Scheme and sequence of compound C.

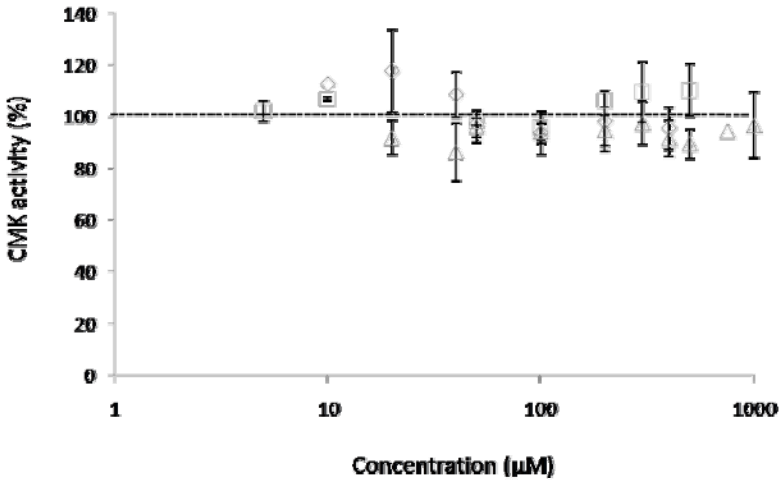


Fig 7. CMK activity of EcCMK in presence of compounds designed to disrupt EcCMK subunit interactions. The activity is represented in % values compared to a control without any compound. Average activities in presence of the compound A, compound B and compound C are represented with diamonds, squares and triangles, respectively. Standard deviations of compound A, compound B and compound C are represented with lines.

Discussion

Even though EcCMK has been described as a dimeric protein (Gabrielsen, Bond et al. 2004), the size-exclusion chromatography experiments carried out in this study with purified preparations of the enzyme showed that they contain a mixture of species being the monomeric species the most abundant forms. The presence of monomeric and oligomeric species, in addition to results obtained by native-PAGE and dynamic light scattering experiments reinforce the idea of a dynamic oligomeric equilibrium. The oligomeric status may depend on diverse factors. One factor that can affect the oligomeric status is the concentration of salt. Size-exclusion chromatography experiments carried out in this study were performed at mild salt concentration (0.1M NaCl) while in other papers where EcCMK was described as dimeric (Gabrielsen, Rohdich et al. 2004) a higher concentration of salt (0.3M) was used. This idea is consistent with the characterization of the homologue CMK of *Thermus thermophilus* as monomeric protein in solution in size-exclusion chromatography experiments performed in columns equilibrated in a buffer containing 0.15M of NaCl (Wada, Kuzuyama et al. 2003). Despite CMK of *Aquifex Aolicus* was also described as a monomeric protein in solution no information about NaCl used in size-exclusion chromatography experiments was provided (Sgraja, Alphey et al. 2008). The fact that other experimental conditions can influence the oligomerization of the CMK cannot be ruled out as has been reported for other enzymes.

Previous work on EcCMK using molecular mechanics generalized born surface area (MMGBSA) determined a high value of the binding free energy of the homodimer complex compared to other dimeric proteins (Gimenez-Oya 2009). This high value attributed to the lower surface of interaction between the EcCMK monomers suggests a labile homodimer interaction and is consistent with the dynamic equilibrium that was observed in the size-exclusion chromatography experiments carried out in this study.

Only minor differences were found among the X-ray structure from three homologues: *Escherichia coli* (Miallau, Alphey et al. 2003), *Thermus thermophilus* (TtCMK) (Wada, Kuzuyama et al. 2003) and *Aquifex aeolicus* (Sgraja, Alphey et al. 2008). Superpositions of C_α between EcCMK and AaCMK (Fig. 8) showed a RMSD value of 1.6A and a Z-value of DALI server (Holm and Sander 1996) of 29.1 (Sgraja, Alphey et al. 2008). Amino acid residues similar to those involved in the salt bridge-stabilizing protein-protein interaction in EcCMK were also found in AaCMK and in TtCMK. One possibility is that the orientation of these residues do not support the dimeric conformation. No allosteric changes were highlighted in the dimeric EcCMK compared to the monomeric TtCMK.

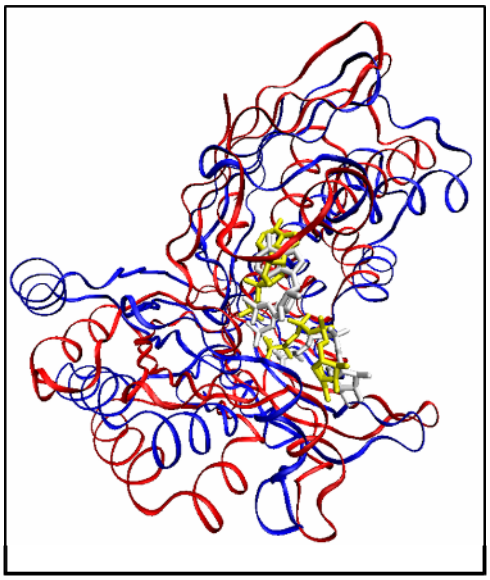


Fig 8. Cα overlay of AaCMK (PDB number= 2V8P) and EcCMK (PDB number= 1jo4) marked in new ribbons mode implemented in VMD visualisation software (Humphrey et al 2005). AaCMK is shown in blue and EcCMK is shown in red. CDP-ME and ATP are shown in yellow and white from AaCMK and EcCMK, respectively.

EcCMK monomeric species displayed similar activity values to the ones reported for the mixture of EcCMK oligomeric species. This result suggests that monomers are enzymatically active and that both monomeric and dimeric species can contribute to the CMK activity. Furthermore, no inhibitory effects were found when CMK activity was assayed in the presence of compounds designed to disrupt the protein-protein interface of the EcCMK homodimer.

The EcCMK catalytic center is located in a deep cavity near the interface of the cofactor and substrate-binding domains for EcCMK (Miallau, Alphey et al. 2003). Lys10 and Asp141 amino acid residues are involved in the hydrogen-bonding interactions which polarizes the hydroxyl group of the substrate, in the base to generate the O2M nucleophile and stabilize the transition state. All amino acid residues involved in the catalytic site or in the ATP or CDP-ME binding domains were described within the same monomer, indicating that monomeric conformation is sufficient for enzymatic activity. This is consistent with the monomeric status of some CMK homologues described to date.

Members of the GHMP superfamily share the characteristic two α+β fold domains with three conserved motifs (Cheek, Zhang et al. 2002). These three conserved motifs are involved in the interaction with the substrates and the cofactor. Interactions between substrates-residues involved in the conserved motifs are described from a single polypeptide chain in GHMP proteins (even for those reported as dimeric). Furthermore, the catalytic center is distant from the dimer interface, suggesting no cross-interaction within partners in the enzymatic activity. Oligomeric status of the GHMP proteins is a controversial issue that can also be observed surveying X-ray structures of GHMP members involved in isoprenoid five-carbon building blocks biosynthesis; mevalonate kinases (MVK), phosphomevalonate kinases (PMVK) and mevalonate 5-diphosphate decarboxylases (MDD). Despite S*treptococcus pneumoniae*, *Rattus norvergicus* and *Homo sapiens* MVK were described as homodimer complexes in solution and in their corresponding X-ray structure (Fu, Wang et al. 2002; Andreassi and Leyh 2004; Fu, Voynova et al. 2008), other homologues such as *Leishmania major* MVK was charactherized as a monomer in solution (Sgraja, Smith et al. 2007). On the other hand, *Methanococcus jannaschii* MVK was described as a dimer in solution but was found to be a monomer in the X-ray structure (Yang, Shipman et al.

2002).

Crystal structures of *Trypanosoma brucei*, *Sthapylococcus aureus* and *Saccharomyces cerevisiae* MDD were found to be dimeric, however, in size exclusion chromatography and sedimentation assays (Byres, Alphey et al. 2007) *T. brucei* MDD was reported as a monomeric enzyme and *S. aureus* MDD was found to be dimeric. Interestingly, *S. cerevisiae* MDD was found to be dimeric in sedimentation assays and in crystal structure, but as described in this paper for EcCMK, displayed a mixture of dimeric and monomeric species in size-exclusion experiments (Byres, Alphey et al. 2007). In both MDDs and EcCMK, the dimeric interface includes hydrophobic clusters of residues with two stabilizing salt-bridges and a numerous solvent-mediated interactions. The stability of the MDDs dimer depends on the surface area of the dimeric interface as well as the stabilizing salt-bridges.

Other members of the GMHP superfamily are galactokinase and homoserine kinase. The *Lactococcus lactis* galactokinase was described as a monomer in solution which was in accordance with the X-ray structure (Thoden and Holden 2003). The *H. sapiens* and *S. cerevisiae* galactokinases were reported to be monomeric in solution (Timson, Ross et al. 2002; Timson and Reece 2003), but were described as homodimers when their corresponding structures were resolved (Thoden, Sellick et al. 2005; Thoden, Timson et al. 2005). This difference in the oligomeric state was postulated to be due either to a possible crystalline packing or the unknown *in vivo* state of the protein (Thoden, Timson et al. 2005).

Variability in the oligomeric state is common in the GHMP kinase superfamily (table 1), however, dimerization does not to appear to influence catalysis. It is in accordance with this, that the active site remains distant from the dimer interface and is composed from a single monomer. Furthermore, no evidence for allosteric regulation has been obtained for the members of the GHMP kinase family. Despite the possibility that the oligomeric state of the EcCMK could be affected by the salt concentration in solution, other possibilities should be explored such as *in vivo* protein instability of the different oligomers. Interestingly, the MVK I268T mutant described in a hyperimmunoglobulinemia D/periodic fever syndrome (HIDS) patient showed <50% of MVK protein than controls, suggesting that mutations in important residues implicated in the dimer formation are causative of the protein instability (Hinson, Ross et al. 1999; Fu, Wang et al. 2002). Other possibility is that oligomeric CMK could facilitate the metabolic flux by interacting with other enzymes of the MEP pathway. Even though EcCMK has been reported to interact with EcIspF, data regarding formation of metabolite channeling in the MEP pathway are also controversial (Gabrielsen, Bond et al. 2004; Lherbet, Pojer et al. 2006).

Conclusion

Despite EcCMK has been previously described as homodimeric protein, the experiments described in this paper clearly demonstrate that purified preparations of the recombinant enzyme exist as a mixture of oligomeric and monomeric species being the last the most prevalent. This has been shown by three different experimental methodologies: size-exclusion chromatography, native-PAGE and dynamic light scattering (DLS). Under the experimental conditions used the EcCMK monomeric species were active and exhibited similar CMK activity values than the previously reported for purified mixture of species. Molecules that were designed to disrupt the protein-protein interaction of the CMK homodimer displayed no inhibitory effects. These results indicate that as described for other enzymes of the GHMP kinase superfamily EcCMK homodimerization is not required for enzymatic activy.

Protein	Organism	X-ray structure[1]	Solution[2]
CDP-ME kinase	*Escherichia coli*	dimeric (Miallau 2003, 1oj4)	dimeric (Analytical ultracentrifugation (Gabrielsen 2004), Size-exclusion (Miallau 2003)), mixture of monomeric and dimeric species (Size-exclusion (in this paper))
	Thermus thermophilus	monomeric (Wada 2003, 1uek)	monomeric (Size-exclusion (Wada2003))
	Aquifex aeolicus	monomeric (Sgraja 2008, 2v2z)	monomeric (Analytical ultracentrifugation, Size-exclusion (Sgraja2008))
Mevalonate kinase	*Streptococcus pneumoniae*	dimeric (Andreassi 2007, 2oi2)	dimeric (Size-exclusion (Andreassi 2004))
	Rattus norvegicus	dimeric (Fu2002, 2r42)	dimeric (Fu 2002)
	Homo sapiens	dimeric (Fu 2008, 2r3v)	dimeric (Fu 2008)
	Methanococcus jannaschii	monomeric (Yang 2002, 1kkh)	dimeric (Size-exclusion (Huang 1999))
	Leishmania major	dimeric (Sgraja 2007, 2hfu)	monomeric (Analytical ultracentrifugation, Size-exclusion (Sgraja2007))
Phosphomevalonate kinase	*Streptococcus pneumoniae*	monomeric (romanowski 2002, 1k47)	monomeric (Size-exclusion (romanowski 2002)
Mevalonate5-diphosphate decarboxylase	*Trypanosoma brucei*	dimeric (Byres 2007, 2hke)	monomeric (Analytical ultracentrifugation, Size-exclusion (Byres2007))
	Sthapylococcus aureus	dimeric (Byres 2007, 2hk2)	dimeric (Analytical ultracentrifugation, Size-exclusion (Byres2007))
	Saccharomyces cerevisiae	dimeric (Bonnano 2001, 1fi4)	dimeric (Analytical ultracentrifugation(Byres2007)), mixture of monomeric and dimeric species (Size-exclusion (Byres2007))
	Homo sapiens	dimeric (Voynova 2008, 3d4j)	
	Rattus norvegicus		tetrameric (Bhat 1980)
Galactokinase	*Lactococcus lactis*	monomeric (Thoden 2003, 1pie)	
	Homo sapiens	dimeric (Thoden 2005, 1wuu)	monomeric (Size-exclusion (Timson2003))
	Saccharomyces cerevisiae	dimeric (Thoden 2005, 2aj4)	monomeric (Size-exclusion (Timson2002))
	Pyrococcus furiosus	monomeric (nine independent chains, Hartley 2004, 1s4e)	monomeric (Hartley 2004)
	Escherichia coli		monomeric (Analytical ultracentrifugation (Wilson 1969)
	Vicia faba sedes		dimeric (Size-exclusion (Dey 1983)
Homoserine kinase	*Methanococcus jannaschii*	monomeric (four independent monomers in monoclinic crystals and one monomer in tetragonal crystals (Khrisna 2001), 1h72)	
	Saccharomyces cerevisiae		dimeric (Size-exclusion (Manhaupt 1990))
	Escherichia coli		dimeric (Size-exclusion (Burr 1976))
XOL-1	*Caenorhabditis elegans*	dimeric or tetrameric (Luz 2003, 1mg7)	monomeric (Size-exclusion (Luz 2003))
Shikimate kinase	*Methanococcus jannaschii*		dimeric (Size-exclusion (Daugherty 2000)

Table 1. Oligomeric characterization of proteins of the GHMP superfamily. Results were obtained either from crystallographic data(1) or in solution (2) from purified preparations. X-ray structure schedule shows oligomeric state, bibliographic reference and PDB id. In solution schedule shows oligomeric state, technique used and reference.

Acknowledgements

The excellent assistance of Dr. Francisco J. Fernandez and Barbara Calisto, Institut de Biologia Molecular de Barcelona – CSIC in the DLS determinations and chromatographic experiments is gratefully acknowledged. The Spanish Ministry of Science and Technology supported this work through the projects BIO2002-04419-C02-02. The 'Generalitat de Catalunya' also supported this work through the grant 2005SGR00914.

Bibliography

Andreassi, J. L., 2nd, P. W. Bilder, et al. (2007). "Crystal structure of the Streptococcus pneumoniae mevalonate kinase in complex with diphosphomevalonate." Protein Sci **16**(5): 983-9.

Andreassi, J. L., 2nd and T. S. Leyh (2004). "Molecular functions of conserved aspects of the GHMP kinase family." Biochemistry **43**(46): 14594-601.

Bernal, C., E. Mendez, et al. (2005). "A spectrophotometric assay for the determination of 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase activity." Anal Biochem **340**(2): 245-51.

Bhat, S. C. and T. Ramasarma (1980). "Purification & properties of mevalonate pyrophosphate decarboxylase of rat liver." Indian J Biochem Biophys **17**(4): 249-54.

Bonanno, J. B., C. Edo, et al. (2001). "Structural genomics of enzymes involved in sterol/isoprenoid biosynthesis." Proc Natl Acad Sci U S A **98**(23): 12896-901.

Bork, P., C. Sander, et al. (1992). "An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins." Proc Natl Acad Sci U S A **89**(16): 7290-4.

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

Boucher, Y. and W. F. Doolittle (2000). "The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways." Mol Microbiol **37**(4): 703-16.

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**: 248-54.

Burr, B., J. Walker, et al. (1976). "Homoserine kinase from Escherichia coli K12." Eur J Biochem **62**(3): 519-26.

Byres, E., M. S. Alphey, et al. (2007). "Crystal structures of Trypanosoma brucei and Staphylococcus aureus mevalonate diphosphate decarboxylase inform on the determinants of specificity and reactivity." J Mol Biol **371**(2): 540-53.

Cheek, S., H. Zhang, et al. (2002). "Sequence and structure classification of kinases." J Mol Biol **320**(4): 855-81.

D'Autreaux, B., L. Pecqueur, et al. (2007). "Reversible redox- and zinc-dependent dimerization of the Escherichia coli fur protein." Biochemistry **46**(5): 1329-42.

Daugherty, M., V. Vonstein, et al. (2001). "Archaeal shikimate kinase, a new member of the GHMP-kinase family." J Bacteriol **183**(1): 292-300.

Dey, P. M. (1983). "Galactokinase of Vicia faba seeds." Eur J Biochem **136**(1): 155-9.

Eisenreich, W., M. Schwarz, et al. (1998). "The deoxyxylulose phosphate pathway of terpenoid biosynthesis in plants and microorganisms." Chem Biol **5**(9): R221-33.

Fu, Z., N. E. Voynova, et al. (2008). "Biochemical and structural basis for feedback inhibition of mevalonate kinase and isoprenoid metabolism." Biochemistry **47**(12): 3715-24.

Fu, Z., M. Wang, et al. (2002). "The structure of a binary complex between a mammalian mevalonate kinase and ATP: insights into the reaction mechanism and human inherited disease." J Biol Chem **277**(20): 18134-42.

Gabrielsen, M., C. S. Bond, et al. (2004). "Hexameric assembly of the bifunctional methylerythritol 2,4-cyclodiphosphate synthase and protein-protein associations in the deoxy-xylulose-dependent pathway of isoprenoid precursor biosynthesis." J Biol Chem **279**(50): 52753-61.

Gabrielsen, M., F. Rohdich, et al. (2004). "Biosynthesis of isoprenoids: a bifunctional IspDF enzyme from Campylobacter jejuni." Eur J Biochem **271**(14): 3028-35.

Gimenez-Oya, V., O. Villacanas, et al. (2009). "Mimicking direct protein-protein and solvent-mediated interactions in the CDP-methylerythritol kinase homodimer: a pharmacophore-directed virtual screening approach." J Mol Model.

Hartley, A., S. E. Glynn, et al. (2004). "Substrate specificity and mechanism from the structure of Pyrococcus furiosus galactokinase." J Mol Biol **337**(2): 387-98.

Hinson, D. D., R. M. Ross, et al. (1999). "Identification of a mutation cluster in mevalonate kinase deficiency, including a new mutation in a patient of Mennonite ancestry." Am J Hum Genet **65**(2): 327-35.

Holm, L. and C. Sander (1996). "Alignment of three-dimensional protein structures: network server for database searching." Methods Enzymol **266**: 653-62.

Huang, K. X., A. I. Scott, et al. (1999). "Overexpression, purification, and characterization of the thermostable mevalonate kinase from Methanococcus jannaschii." Protein Expr Purif **17**(1): 33-40.

Humphrey, W., A. Dalke, et al. (1996). "VMD: visual molecular dynamics." J Mol Graph **14**(1): 33-8, 27-8.

Jomaa, H., J. Wiesner, et al. (1999). "Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis

as antimalarial drugs." Science **285**(5433): 1573-6.
Karas, M. and F. Hillenkamp (1988). "Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons." Anal Chem **60**(20): 2299-301.
Krishna, S. S., T. Zhou, et al. (2001). "Structural basis for the catalysis and substrate specificity of homoserine kinase." Biochemistry **40**(36): 10810-8.
Kuzuyama, T., S. Takahashi, et al. (2000). "Characterization of 1-deoxy-D-xylulose 5-phosphate reductoisomerase, an enzyme involved in isopentenyl diphosphate biosynthesis, and identification of its catalytic amino acid residues." J Biol Chem **275**(26): 19928-32.
Lherbet, C., F. Pojer, et al. (2006). "Absence of substrate channeling between active sites in the Agrobacterium tumefaciens IspDF and IspE enzymes of the methyl erythritol phosphate pathway." Biochemistry **45**(11): 3548-53.
Lichtenthaler, H. K. (1999). "The 1-Deoxy-D-Xylulose-5-Phosphate Pathway Of Isoprenoid Biosynthesis In Plants." Annu Rev Plant Physiol Plant Mol Biol **50**: 47-65.
Luttgen, H., F. Rohdich, et al. (2000). "Biosynthesis of terpenoids: YchB protein of Escherichia coli phosphorylates the 2-hydroxy group of 4-diphosphocytidyl-2C-methyl-D-erythritol." Proc Natl Acad Sci U S A **97**(3): 1062-7.
Luz, J. G., C. A. Hassig, et al. (2003). "XOL-1, primary determinant of sexual fate in C. elegans, is a GHMP kinase family member and a structural prototype for a class of developmental regulators." Genes Dev **17**(8): 977-90.
Mannhaupt, G., H. D. Pohlenz, et al. (1990). "Yeast homoserine kinase. Characteristics of the corresponding gene, THR1, and the purified enzyme, and evolutionary relationships with other enzymes of threonine metabolism." Eur J Biochem **191**(1): 115-22.
Miallau, L., M. S. Alphey, et al. (2003). "Biosynthesis of isoprenoids: crystal structure of 4-diphosphocytidyl-2C-methyl-D-erythritol kinase." Proc Natl Acad Sci U S A **100**(16): 9173-8.
Rohdich, F., J. Wungsintaweekul, et al. (2000). "Biosynthesis of terpenoids: 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase from tomato." Proc Natl Acad Sci U S A **97**(15): 8251-6.
Rohmer, M. (1999). "The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants." Nat Prod Rep **16**(5): 565-74.
Romanowski, M. J., J. B. Bonanno, et al. (2002). "Crystal structure of the Streptococcus pneumoniae phosphomevalonate kinase, a member of the GHMP kinase superfamily." Proteins **47**(4): 568-71.
Sacchettini, J. C. and C. D. Poulter (1997). "Creating isoprenoid diversity." Science **277**(5333): 1788-9.
Sgraja, T., M. S. Alphey, et al. (2008). "Characterization of Aquifex aeolicus 4-diphosphocytidyl-2C-methyl-d-erythritol kinase - ligand recognition in a template for antimicrobial drug discovery." Febs J **275**(11): 2779-94.
Sgraja, T., T. K. Smith, et al. (2007). "Structure, substrate recognition and reactivity of Leishmania major mevalonate kinase." BMC Struct Biol **7**: 20.
Shi, W., J. Feng, et al. (2007). "Biosynthesis of isoprenoids: characterization of a functionally active recombinant 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (IspD) from Mycobacterium tuberculosis H37Rv." J Biochem Mol Biol **40**(6): 911-20.
Thoden, J. B. and H. M. Holden (2003). "Molecular structure of galactokinase." J Biol Chem **278**(35): 33305-11.
Thoden, J. B., C. A. Sellick, et al. (2005). "Molecular structure of Saccharomyces cerevisiae Gal1p, a bifunctional galactokinase and transcriptional inducer." J Biol Chem **280**(44): 36905-11.
Thoden, J. B., D. J. Timson, et al. (2005). "Molecular structure of human galactokinase: implications for type II galactosemia." J Biol Chem **280**(10): 9662-70.
Timson, D. J. and R. J. Reece (2003). "Functional analysis of disease-causing mutations in human galactokinase." Eur J Biochem **270**(8): 1767-74.
Timson, D. J., H. C. Ross, et al. (2002). "Gal3p and Gal1p interact with the transcriptional repressor Gal80p to form a complex of 1:1 stoichiometry." Biochem J **363**(Pt 3): 515-20.
Voynova, N. E., Z. Fu, et al. (2008). "Human mevalonate diphosphate decarboxylase: characterization, investigation of the mevalonate diphosphate binding site, and crystal structure." Arch Biochem Biophys **480**(1): 58-67.
Wada, T., T. Kuzuyama, et al. (2003). "Crystal structure of 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase, an enzyme in the non-mevalonate pathway of isoprenoid synthesis." J Biol Chem **278**(32): 30022-7.
Wilson, D. B. and D. S. Hogness (1969). "The enzymes of the galactose operon in Escherichia coli. 3. The size and composition of galactokinase." J Biol Chem **244**(8): 2137-42.
Yang, D., L. W. Shipman, et al. (2002). "Structure of the Methanococcus jannaschii mevalonate kinase, a member of the GHMP kinase superfamily." J Biol Chem **277**(11): 9462-7.

Clonación molecular y caracterización de la 1-desoxi-D-xilulosa 5-fosfato de la bacteria termofílica *Thermotoga maritima*

Victor Giménez-Oya, Albert Boronat , Santiago Imperial

Dept. de Bioquímica i Biología Molecular, Facultat de Biologia, Universitat de Barcelona,, Avda Diagonal 645, E-08028 Barcelona, Spain.

Running title: 1-desoxi-D-xilulosa 5-fosfato sintasa de *Thermotoga maritima*

Key words**:** 1-desoxi-D-xilulosa 5-fosfato sintasa, metileritritol fosfato, *Thermotoga maritima*, isoprenoid biosynthesis, thiamine

Resumen:

La enzima 1-desoxi-D-xilulosa 5-fosfato sintasa (DXS) cataliza el primer paso limitante en la ruta del metileritritol fosfato de biosíntesis de isoprenoides y, a su vez, es un paso crucial en la ruta de síntesis de tiamina en bacterias. Se ha identificado el marco abierto de lectura TM_1770 como posible gen *dxs*. La transformación de una cepa de *E.coli* defectiva para el gen de la *dxs*, con vectores de expresion que codifican para TM_1770 resultó en un rescate de esta auxotrofia. La proteína recombinante conjugada a una cola carboxiterminal de 6 His mostró una temperatura óptima para la actividad DXS a 70ºC con una cinética típica de Michaelis-Mentem. La composición aminoacídica de TmDXS fue comparada con otras DXS de organismos termofílicos y mesofílicos para poder identificar los rasgos que sustentan el carácter termofílico de la enzima. Uno de ellos se presenta mediante el incremento de residuos cargados en detrimento de pequeños residuos polares no cargados presente en otras DXS termofílicas. Todos los residuos caracterizados en otras DXS importantes para la estabilización, unión al sustrato y catálisis se han encontrado conservados en TmDXS. Un ejemplo es el residuo His en la posición 35 que, por mutagénesis dirigida, se ha caracterizado como crucial en la actividad enzimática de TmDXS de igual manera que el residuo homólogo His 49 de E.coli DXS que participa en la transferencia de protones durante la catálisis enzimática.

Molecular cloning and Characterization of 1-deoxy-D-xylulose 5-phosphate synthase from the thermophilic bacterium *Thermotoga maritima*

Victor Giménez-Oya, Albert Boronat , Santiago Imperial

Dept. de Bioquímica i Biología Molecular, Facultat de Biologia, Universitat de Barcelona,, Avda Diagonal 645, E-08028 Barcelona, Spain.

Running title: 1-deoxy-D-xylulose 5-phosphate synthase from *Thermotoga maritima*

Key words**:** 1-deoxy-D-xylulose 5-phosphate synthase, methylerythritol phosphate, Thermotoga maritima, isoprenoid biosynthesis, thiamine

Abstract

1-deoxyxylulose-5-phosphate synthase (DXS) catalyzes the first, rate-limiting step in the methylerythritol phosphate pathway of isoprenoid biosynthesis, as well as a crucial step in the "de novo" thiamine biosynthesis in most bacteria. The *Thermotoga maritima* open reading frame (ORF) TM_1770 was identified by comparative sequence analysis as a putative *dxs* gene. Transformation of *E.coli* cells blocked in the *dxs* gene with expression vectors harboring TM_1770 resulted in a rescue of the auxotrophy showing that the gene is able to replace the endogenous *dxs E. coli* gene. These results were confirmed by producing the recombinant protein in *E. coli* and carrying out functional assays. The recombinant protein, TmDXS comprising 608 amino acid residues conjugated with a C-terminal 6 His tag, displayed an optimum DXS activity at 70ºC and showed a normal Michaelis-Mentem kinetics. The amino acid composition of TmDXS was compared with those of other thermophillic and mesophillic microorganisms in order to identify thermostability traits. The ratio of charged versus polar non-charged amino acid residues were higher in TmDXS than in DXS from mesophilic bacteria and similar to the ratios obtained for the enzyme from other thermophilic organisms. All the amino acid residues described in other DXS to be involved in coenzyme stabilization, substrate binding and catalysis are highly conserved in TmDXS. As an example the histidine residue at position 35 in TmDXS was shown by site-directed mutagenesis to be a crucial residue as was found for the homologue histidine 49 of *E. coli* DXS which is involved in the proton transfer during catalysis.

Introduction

Isoprenoids are the largest family of natural products that arise over than 35000 different compounds (Penuelas and Munne-Bosch 2005). All isoprenoids are derived from the same five-carbon precursors, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). Isoprenoids are ubiquitous distributed among all life kingdoms and have widespread functions, that go from structural components of the membranes, plant hormones (like gibberelins, abscisic acid..), structural carriers, to secondary metabolism functions associated to defense or communication in adaptative responses to the environment. For a long time, the synthesis of isoprenoid precursors was assumed in all the organisms to occur through the classical mevalonate (MVA) pathway from acetyl-CoA (Spurgeon and Porter 1981). Nowadays, it is known that IPP and DMAPP can be synthesized through a different set of reactions, the methylerythritol 4-phosphate (MEP) or MVA-independent pathway. The MEP pathway is present in most eubacteria and some protozoa and is absent in animals and in most of the archaea. Both biosynthetic pathways are present in plants, although in different cell compartments, while the MVA pathway is present in the cytosol, the MEP pathway is located in plastids. Moreover, 1-deoxy-D-xylulose 5-phosphate synthase (DXS) was found as one of the rate-limiting steps of the MEP biosynthetic pathway and the isoprenoid production in *Arabidopsis thaliana* (Estevez, Cantero et al. 2001)**.**

DXS catalyzes the first reaction step of the MEP pathway, a transketolase-type condensation involving pyruvate and glyceraldehyde 3-phosphate to form 1-deoxy-D-xylulose 5-phosphate (DXP) (Lange, Wildung et al. 1998). This reaction is a thiamine pyrophosphate-dependent acycloin condensation involving C atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to produce DXP (Sprenger, Schorken et al. 1997). The catalysis of this type of reaction is due to a secondary activity of the thiamine pyrophosphate cofactor as has been documented on pyruvate decarboxylating enzymes like the E1 component of the pyruvate decarboxylase complex, which could transfer the acetaldehyde group from pyruvate to glyceraldehyde 3-phosphate (Sprenger, Schorken et al. 1997). DXP is the first precursor of the MEP biosynthesis pathway. Furthermore, 1-deoxy-D-xylulose was found to be a precursor of the "de novo" biosynthesis of the coenzymes thiamine (Sprenger, Schorken et al. 1997) and pyridoxal (Cane, Hsiung et al. 1998) in *Escherichia coli*. Thiamine (vitamin B1) and pyridoxol (vitamin B6) are essential cofactors involved in the catalytic activity of several enzymes.

Since thiamine diphosphate (TPP) was discovered as an active thiamine cofactor diverse metabolic roles were associated to TPP. Mainly, TPP assists in making and breaking bonds between carbon and sulfur, oxygen, hydrogen, and nitrogen atoms; and most remarkably, the breaking and making of carbon-carbon bonds (Frank, Leeper et al. 2007). Although the catalytic activity of the cofactor is well known (Settembre, Begley et al. 2003), its regulation presents a very intricate and interesting model (Frank, Leeper et al. 2007). DXS is one of the first enzymes involved in the formation of the TPP hydroxyethylthiazole ring (Frank, Leeper et al. 2007) and it is itself a TPP-dependent enzyme (Sprenger, Schorken et al. 1997), suggesting a positive feed-back regulation. At the RNA level, the regulation takes place in the presence of "riboswitches" (Thore, Leibundgut et al. 2006). Findings in the TPP-dependent enzymes like the mode of regulation (Sudarsan, Barrick et al. 2003), mechanistic catalytic similarities, and the conservation of the few amino acids that bind to TPP in the TPP-dependent proteins suggested an early evolutionary origin for this molecule (Frank, Leeper et al. 2007).

Vitamin B6 and all of its vitamers (pyridoxine, pyridoxol, pyridoxine 5-phosphate, pyridoxol 5-phosphate) have also an essential role in the catalytic activity of many of enzymes including those catalyzing transamination, decarboxylation, racemization, C_α-C_β cleavage and α,β-elimination reactions (Giebel, Cass et al. 1995; Fitzpatrick,

Amrhein et al. 2007). In *E. coli*, pyridoxal 5-phosphate (PLP) is synthesized through the condensation of 4-(phosphohydroxy)-L-threonine and DXP (Laber, Maurer et al. 1999), but in the last years it has been demonstrated that most organisms (bacteria, fungi and plants) synthesize PLP by DXP-independent pathway (Garrido-Franco 2003) and the DXP-dependent pathway being present only in a small subset of γ-proteobacteria (Fitzpatrick, Amrhein et al. 2007). The DXP-independent pathway was elucidated in *Saccharomyces cerevisiae* (Bean, Dvorachek et al. 2001; Ehrenshaft and Daub 2001), in *Cercospora nicotianae* (Ehrenshaft, Bilski et al. 1999; Osmani, May et al. 1999), and in the model plant *Arabidopsis thaliana* (Tambasco-Studart, Titiz et al. 2005; Titiz, Tambasco-Studart et al. 2006; Wagner, Bernhardt et al. 2006) (for recent review see (Webb, Marquet et al. 2007). The two pathways for PLP synthesis have been found to be self-exclusive in the organisms reported to date (Garrido-Franco 2003; Tambasco-Studart, Titiz et al. 2005) although, in fact, they are highly similar in structural and mechanistic terms (Fitzpatrick, Amrhein et al. 2007).

All these characteristics show that DXS can play a crucial role in essential biosynthetic pathways in different organisms. To date, genes encoding DXS have been cloned and characterized from eubacteria (Sprenger, Schorken et al. 1997; Lois, Campos et al. 1998; Harker and Bramley 1999; Miller, Heuser et al. 1999; Kuzuyama, Takahashi et al. 2000; Bailey, Mahapatra et al. 2002; Eubanks and Poulter 2003; Lee, Oh et al. 2007), from plants (Mandel, Feldmann et al. 1996; Bouvier, d'Harlingue et al. 1998; Estevez, Cantero et al. 2000; Lois, Rodriguez-Concepcion et al. 2000; Kim, Kim et al. 2005; Alos, Cercos et al. 2006; Gong, Liao et al. 2006; Munoz-Bertomeu, Arrillaga et al. 2006; Zhang, Li et al. 2008), and highlighted from protozoa (Cassera, Gozzo et al. 2004). Recently, the X-ray DXS structure from *E. coli* and *Deinococcus radiodurans* have been characterized (Xiang, Usunow et al. 2007). Despite the relatively high number of DXS genes characterized, no one from a thermophilic microorganism has been reported to date.

Thermotoga maritima is a Gram negative non-spore-forming, rod-shaped bacterium from the order of Thermotogales. *Thermotoga maritima* was originally isolated from the Vulcano's (Italy) geothermal heated marine sediments (Huber, Langworthy et al. 1986)), and has an optimum growth at a temperature of 80ºC. This bacterium has an evolutionary significance because its phylogeny shows one of the deepest and slowly evolving lineages in Eubacteria (Achenbach-Richter, Gupta et al. 1987) which has the highest percentage (24%) of genes that are most similar to archaeal genes (Nelson, Clayton et al. 1999). Furthermore, the gene order conservation of many of the clustered regions proposes that lateral gene transfer may have occurred between these two kingdoms. *Thermotoga maritima* grow with an ability through a fermentative metabolism, as well as a respiratory organism generating energy in the presence of Fe(III) (Vargas, Kashefi et al. 1998) in the presence of low concentrations of oxygen of up to 0.5% (v/v) (Le Fourn, Fardeau et al. 2008).

The crucial role in essential biosynthetic pathways reported for DXS and the evolutionary importance of *Thermotoga maritima* and its thermophilic characteristics led us to consider the characterization of the enzyme DXS from *T. maritima* (TmDXS). Moreover, this enzyme offers the opportunity by understanding its divergence from homolog sequences to study the mechanisms underlying enzyme thermostability.

Results and discussion

The genome database of *Thermotoga maritima* was surveyed and a putative gene encoding DXS in TM_1770 (Q9X291) was identified. Other putative genes for the MEP pathway of biosynthesis of isoprenoids precursors (TmDXR TM_0889, TmIspD TM_1393, TmIspE TM_1383, TmIspF TM_0647, TmIspG TM_0891, TmIspH TM_1444) and for "de novo" biosynthesis of thiamine were also identified. The putative genes for the biosynthesis of pyridoxol showed high similarity with the DXP-independent pathway enzymes pDX1 and pDX2 (TmPdxS or YaaD TM_0473 and TmPdxT or YaaE TM_0472). The structure of the enzymes YaaD and YaaE has recently been elucidated (Zein, Zhang et al. 2006) corroborating the presence of the DXP-independent pathway for the pyridoxol biosynthesis in *T. maritima*.

Cloning of the ORF encoding Thermotoga maritima DXS and complementation assay in a DXS-deficient *Escherichia coli* strain

In order to characterize the enzyme DXS from *Thermotoga maritima*, the TM_1770 ORF was cloned into a series of *E.coli* expression vectors. Using the published sequence of the ORF TM_1770 (Expasy accession number Q9X291, GenBank accession number AE000512), sets of flanking primers (see table 1) were designed and used to isolate the full length open reading frame by PCR from the MSB8 strain genomic DNA. ORF TM_1770 contains 1827 bp and encodes a putative 608 amino acid DXS. TmDXS has 40.8% of identity and 56.5% similarity to the *E.coli* DXS(EcDXS). The putative *TmDXS* gene was subcloned into three different expression vectors (pEQ30, pT7-7 and pET23).

EcAB4-2 cells (*Escherichia coli* MG1655 dxs::CAT harboring MVA+ operon (Campos, Rodriguez-Concepcion et al. 2001)) transformed with the expression constructs were able to grow in media without mevalonate (Fig1). Therefore, the recombinant TmDXS was expressed from the introduced constructs and showed DXS activity, rescuing the DXP auxotrophy of the *EcAB4-2* cells. *EcAB4-2* cells transformed with the pET23 (b) expression vector could not grow under these conditions in the absence of exogenous mevalonate.

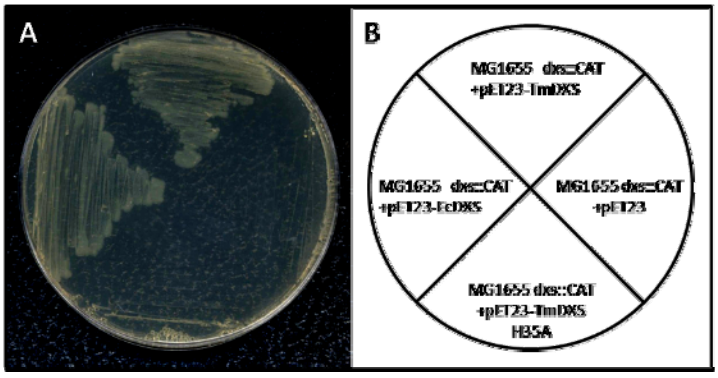


Fig. 1 TmDXS complementation assays in DXS-deficient *EcAB4-2* cells. A, *EcAB4-2* cells (MG1655 (dxs:CAT +MVAoperon)) harboring plasmid (pET23) and expression constructs (pET23-EcDXS, pET23-TmDXS and pET23-TmDXS H35A) streaked on a plate without mevalonate. The pET23-EcDXS and pET23-TmDXS expression constructs are able to rescue the *EcAB4-2* strain DXP auxotrophy. The pET23 (negative control) and pET23-TmDXS H35A constructs did not complemented the auxotrophy. B, Scheme of the complementation assays.

Expression and purification of TmDXS in *E.coli* BL21 Rosetta2

Low levels of recombinant protein were found when M15[pREP4] cells harboring pQE30-TmDXS, BL21[DE3] or BL21 [DE3] pLysS cells harboring pT7-TmDXS or pET23-TmDXS were induced with IPTG under different

conditions (IPTG concentrations and growth temperatures). The phylogenetic position of *Thermotoga maritima* between two kingdoms of life, archaea and eubacteria, implies a different codon usage from eubacteria. A successful lateral gene transfer would require a similar codon usage between foreign genes (archaeal genes in this case) and recipient genomes (Medrano-Soto, Moreno-Hagelsieb et al. 2004). Despite this, it must be noted that in the case of DXS, lateral gene transfer from archaea is very unlikely due to the absence of the MEP pathway in this group. Analysis of the nucleotide sequence of TmDXS showed that 24% of codons were found to be rare codons in *Escherichia coli* (less than 0.2 of the ratio from codon abundance to all the codons for the same amino acid abundance). In an attempt to improve expression levels BL21 [DE3] Rosetta2 cells were used. These cells contain the pRARE2 plasmid that encodes seven tRNAs which are rarely used in *E.coli*. BL21 [DE3] Rosetta2 cells harboring the pET23-TmDXS allowed the expression of the TmDXS protein in sufficient amount to be purified. The TmDXS recombinant protein was purified by a procedure described in the Methods section. The 6His-tag recombinant protein, which has a theoretical molecular weight of 68219.79 Da, showed an apparent molecular mass of aprox 70KDa in a SDS-polyacrilamide gel electrophoresis (Fig.2). This size is very close to the predicted size of TmDXS. A total of 2.5 mg of recombinant protein was obtained from 1L of culture media. Some low molecular weight bands were also shown in the SDS-PAGE gels. These bands were TmDXS-related as confirmed by western blot analysis with anti-his antibody (data not shown).

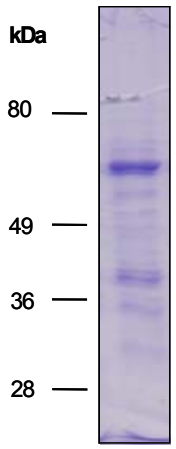


Fig.2 SDS-PAGE analysis of purified preparations of TmDXS. SDS-PAGE was carried out in 10% polyacrylamide gels.Sample: eluates obtained after immobilized metal affinity chromatography (IMAC). ´Gels were stained with Coomassie brilliant blue R-250.

Kinetic parameters of TmDXS

Most studies on the functional characterization of DXS from different organisms have been carried out using an spectrophotometric assay which uses DXR as coupled enzyme (Jomaa, Wiesner et al. 1999). In order to determine DXS at different temperatures using a non-termophilic DXR that assay was modified into a two-step procedure. The reaction catalyzed by TmDXS was carried out at different temperatures and for a fixed time and then the DXP formed was quantified spectrophotometrcially at room temperature. On this assay, TmDXS exhibited an optimum of activity at 70ºC. At 75ºC, the enzyme showed an activity similar than the optimum (94% of activity) (Fig.3A).

An Arrhenius plot of activity versus the reciprocal of temperature (Fig 3A, inset) failed to reveal any break in the 37 to 70ºC temperature range that might be a sign of a drastic phase transition. From the plot, an activation energy of 34.312 KJ mol^{-1} K^{-1} was calculated. The optimum of activity has been reported in *Thermotoga maritima* enzymes in a range from 60ºC in a EST53 carboxylesterase (Kakugawa, Fushinobu et al. 2007) to 90ºC in a Endo-β-1,4-D-galactanase (Yang, Ichinose et al. 2006). Our results are in agreement with this range of optimum temperatures. TmDXS also showed a normal Michaelis-Mentem kinetics (Fig 3B) with a K_M of 75.729µM and V_{max} 4.553 µmol min^{-1} mg^{-1} for acid pyruvic, K_{cat} of 5.156 s^{-1} and K_{cat}/K_M of 6.8348 x $10e^4$ $M^{-1}s^{-1}$, measured at 70ºC and pH 7.8.

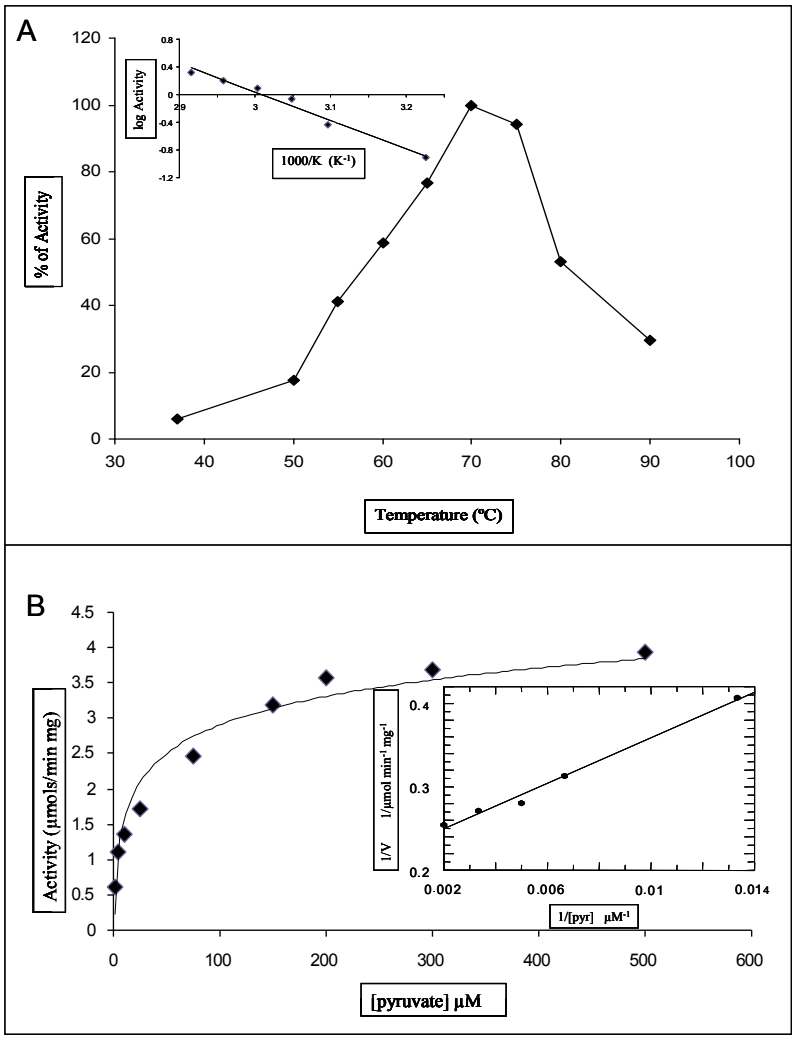


Fig.3A Effect of the temperature on the activity of *Thermotoga maritima* DXS. Temperature influence on the TmDXS enzyme activity (given as percentage of the activity at 70ºC). The inset gives the Arrhenius plot for the activity of the TmDXS in the enzyme range 37ºC to 70ºC. Fig.3B. Pyruvate substrate saturation curve for TmDXS determined in the presence of a fixed concentration of glyceraldehide 3-phosphate utilizing a coupled enzyme assay measured at 70ºC. The inset shows the Lineweaver–Burk plot of initial velocity vs. concentration of pyruvate.

Site-directed mutagenesis of the conserved histidine residue (H35A)

Escherichia coli DXS His49 (Querol, Rodriguez-Concepcion et al. 2001) and *Agrobacterium tumefaciens* DXS His48 (Lee, Oh et al. 2007) are homolog residues playing an important role in the proton transfer for catalysis. This residue belongs to a cluster of histidines and arginines involved in the binding with the diphosphate group of TPP, proton transfer in the catalysis and the stabilization of the covalent intermediate between the substrate and the TPP (Lindqvist, Schneider et al. 1992). Comparative analysis revealed that TmDXS presents this conserved histidine residue at position 35 (Fig.4). To evaluate the importance of this amino acid residue in the TmDXS enzyme activity, we constructed a mutant in which the histidine residue was replaced by an Alanine (named pET23-TmDXS H35A, see Materials and Methods section). As shown in figure 1A, *EcAB4-2* cells harboring pET23-TmDXS H35A were not able to rescue the DXP auxotrophy indicating that this enzyme variant (with the point mutation) did not show enough DXS activity to allow cell growth.

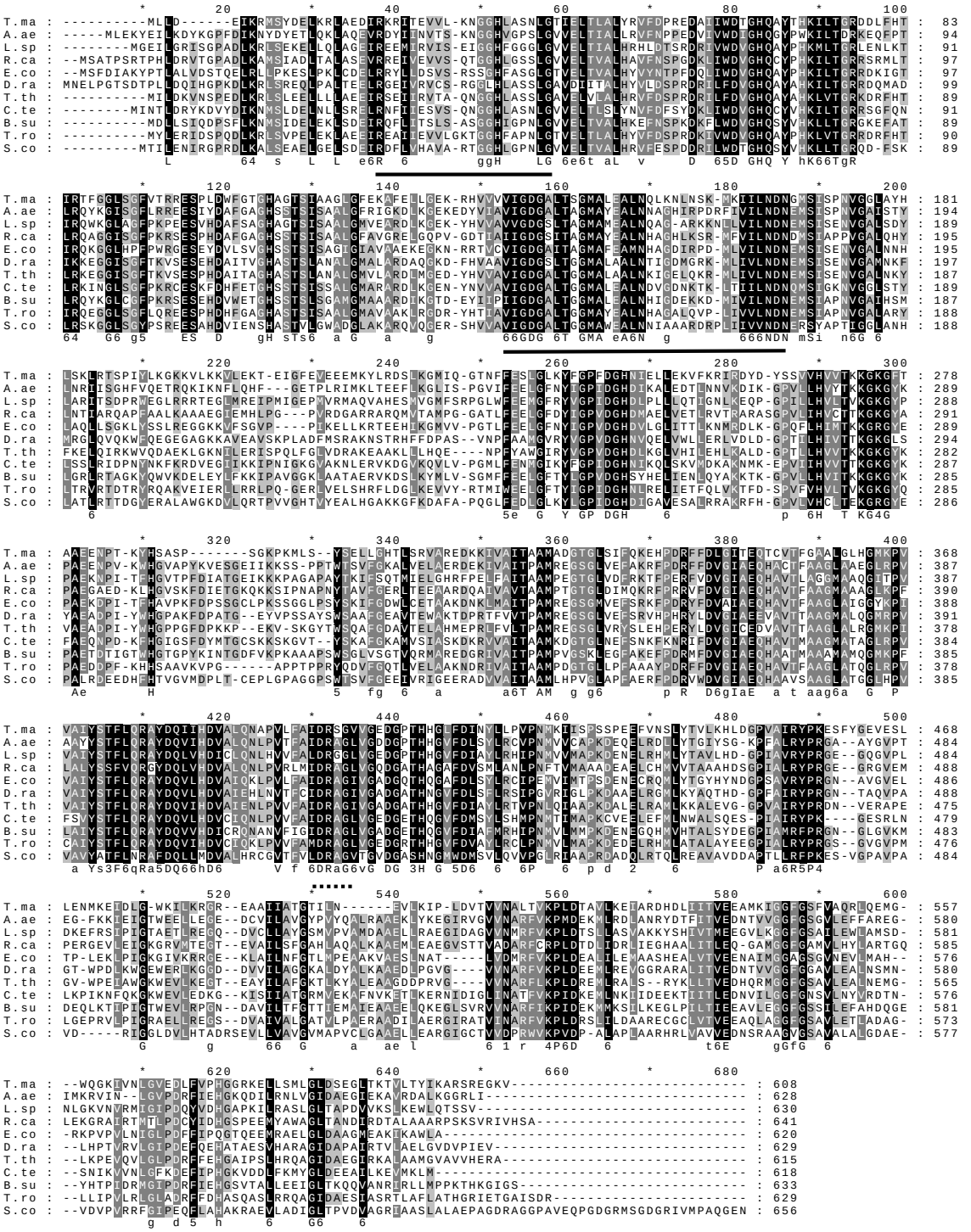


Fig.4 Multiple aligment of DXS proteins from the principal phylla of eubacteria. DXS protein sequences from *Thermotoga maritima* (T.ma, expasy entry number Q9X291, Thermotogae), *Aquifex aeolicus* (A.ae, O67036, Aquificae), *Leptospirum sp.* (L.sp, A3ERK3, Nitrospirae), *Escherichia coli* (E.co, P77488, Gamma-proteobacteria), *Rhodobacter capsulatus* (R.ca, P26242, Alpha-proteobacteria), *Deinococcus radiodurans* (D.ra, Q9RUB5, Deinococcus-thermus), *Thermus thermophilus* (T.th, Q72H81, Deinococcus-thermus), *Clostridium tetani* (C.te, Q894H0, Firmicutes), *Bacillus subtilis* (B.su, P54523, Firmicutes), *Thermomicrobium roseum* (T.ro, B9L1L6, Thermomicrobia) and *Streptomyces coelicolor* (S.co, Q9X7W3, Actinobacteria) were aligned with the Clustal W program and refined with the Genedoc software. TPP-binding domain and the transketolase-like N-terminal domain are underlined. The DRAG signature motif is discontinous underlined.

Phylogenetic tree and amino acid sequence analysis of TmDXS

As shown in figure 5, the phylogenetic position of TmDXS is close to the clustery group of of the Gram+ eubacteria enzymes, and especially to Clostridiales. TmDXS sequence showed a higher identity and similarity to *Clostridium acetobutylicum* (44.6% and 64%, respectively) than to the enzyme of any other Gram– organism such as *Escherichia coli* (40.8% and 56.5%, respectively). All the *Thermotoga maritima* MEP pathway enzymes showed high similarity, as well too other thermophilic enzymes or organisms within the same phyllum, to enzymes of Firmicutes (especially with the orders of Clostridiales and Bacillales). Similarly analysis of 16S rRNA 7-methylguanosine methyltransferase showed high degree of identity/similarity of *Thermotoga maritima* with Firmicutes (especially with Clostridiales and Lactobacillales). These results suggest evolution proximity rather than a lateral gene transfer process.

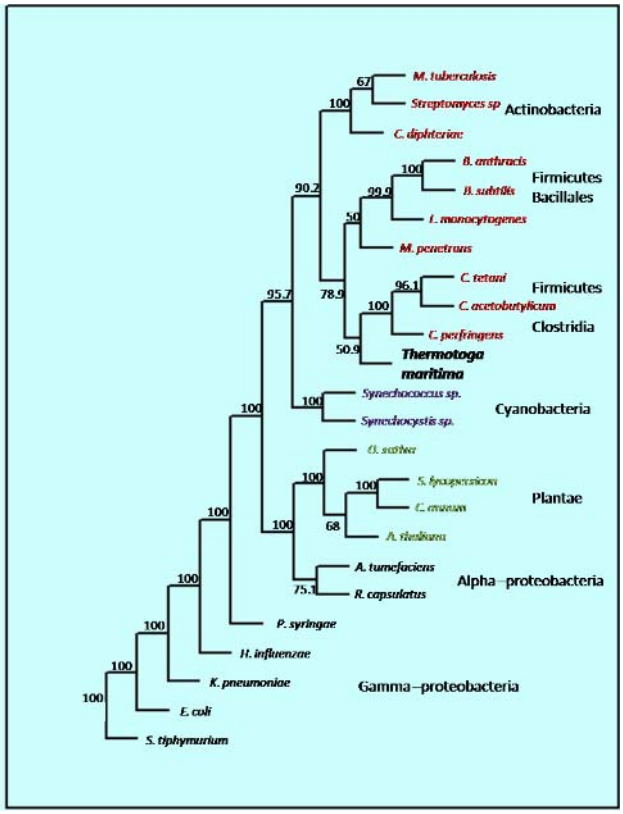


Fig.5 Unrooted phylogenetic tree of DXS proteins extracted by the neighbor-joining method with the Jones-Taylor-Thornton evolutionary model. One thousand bootstrap replicates were performed and percentage bootstrap supports are given. The tree was generated using the Phylyp package of programs and some taxonomic clusters are marked. Gram positives are in red, Gram negative are in black, Cyanobacteria are in blue and Plantae in green. The following sequences were included: *Escherichia coli* (P77488), *Rhodobacter capsulatus* (P26242), *Streptomyces sp*. (Q9X7W3), *Mycobacterium tuberculosis* (P0A554), *Bacillus subtilis* (P54523), *Agrobacterium tumesfaciens* (Q8UHD7), *Synechococcus sp.* (Q2JTX2), *Synechocystis sp*. (P73067), *Arabidopsis thaliana* (Q38854), *Solanum lycopersicom* (Q9XH50), *Capsicum annum* (O78328), *Oryza sativa* (O22567), *Salmonella tiphymurium* (Q8ZRD1), *Thermotoga maritima* (Q9X291), *Bacillus anthracis* (Q81M54), *Clostridium acetobutylicum* (Q97HD3), *Clostridium perfringens* (Q8XJE1), *Clostridium tetani* (Q894HO), *Corynebacterium diphtheriae* (Q6NGV3), *Haemophilus influenzae* (Q4QKG6), *Klebsiella pneumoniae* (A6T5F3), *Listeria monocytogenes* (Q8Y7C1), *Mycoplasma penetrans* (Q8EWX7) and *Pseudomona syringae* (Q48NXO).

One of the important parameters in determining protein thermostability is the amino acid composition. Thermostable proteins contain an increased ratio of charged amino acid residues in detriment of polar non-charged residues. These charged amino acid residues confer more rigidity and stability to the protein by minimizing deamination and backbone cleavages (Fukuchi and Nishikawa 2001). We analyzed the ratios of charged and polar non-charged amino acid residues in DXS from different organisms. TmDXS ratios were reasonably higher than those corresponding to DXS from mesophilic bacteria and similar ratios to other thermophilic organism (Fig.6). The increase of charged residues would result in a higher number of hydrogen bonds achieving more thermostability. Despite that, the increase was small in percentage, but the evolution forces may act to specific zones increasing the stabilizing hydrogen bond networks. Moreover, the DXS from organisms belonging to the order of clostridiales contained similar levels of charged residues as shown for termophilic organisms, highlighting another reason for the proximity shown in the phylogenetic analysis. However, in clostridiales the ratio of polar non-charged residues was even higher indicating that the increase of charged residues was not at expense of non-charged polar residues. Despite the increased ratio of charged/polar non-charged amino acid residues in TmDXS, other mechanisms such as hydrophobic interactions that can favor thermostability cannot be discarded.

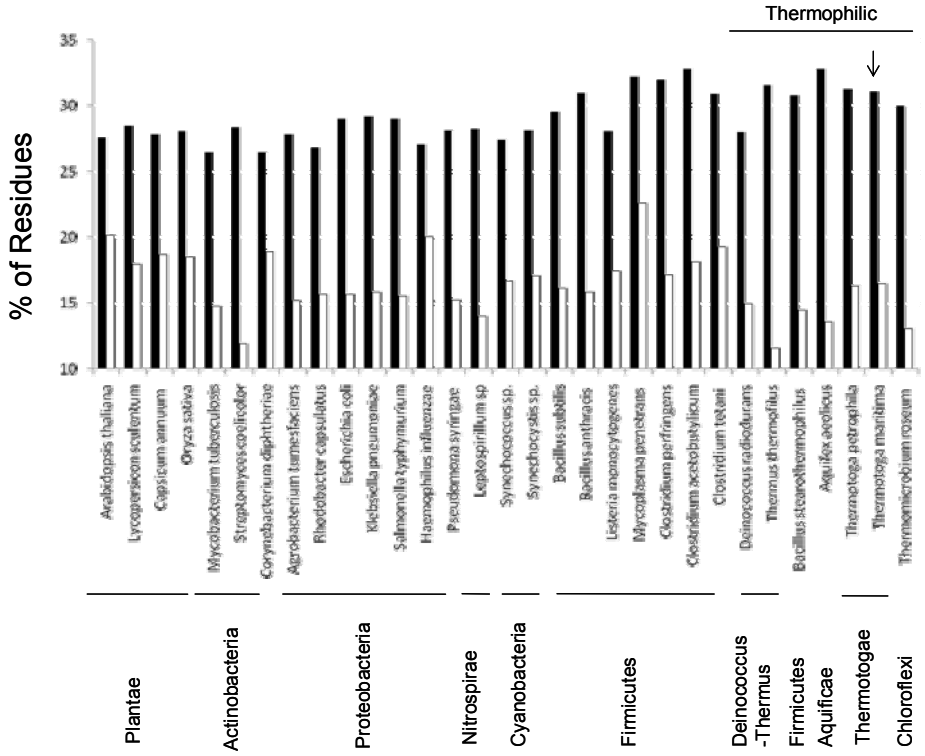


Fig.6 Ratio of charged residues to total residues (in black) and the ratio of polar non-charged residues to total residues (in white) of DXS from different organisms. TmDXS is marked with an arrow.

Comparative sequence analysis of the TmDXS (Fig. 4) showed a sequence motif common to all known TPP-dependent enzymes which have been suggested to embody the TPP-binding site (Reynen and Sahm 1988; Hawkins, Borges et al. 1989). The consensus binding motif $GDGX_{25\text{-}30}N$ begins with the highly conserved Gly-Asp-Gly (GDG) located in the Gly137 residue of the TmDXS, continues with 28 amino acid residues with less

homology and finalizes with an asparagine residue (Asn167) preceded by some hydrophobic residues (Schenk, Duggleby et al. 1998). Furthermore, all the amino acid residues described in EcDXS (or DrDXS) to contribute to the stabilization of TPP (Xiang, Usunow et al. 2007) have been found conserved in the TmDXS sequence (EcDXS Lys284, Phe393, Ile368, Ile185 and Glu371/ TmDXS Lys 273, Phe 375, Ile 348, Ile171 and Glu350). TmDXS also shows a good match with the N-terminal transketolase motif described in EcDXS, R-$(X)_4$-[D,E,Q]-$(X)_{5,6}$-[G,S,T]-G-H-[P,L,I,V,M]-[G,S,T]-(X_3)-[G,S,T] (Querol, Rodriguez-Concepcion et al. 2001), including the histidine residue crucial for the proton transfer in catalysis (EcDXS His49, TmDXS His35). Residues that could have interactions with glyceraldehyde 3-phosphate in EcDXS/DrDXS also have been found conserved in TmDXS (EcDXS His299, Tyr391, Arg420, Asp427 and Arg475/TmDXS His288, Tyr372, Arg400, Asp407 and Arg455). Finally, in the proposed signature motif for DXS, DRAG (Bailey, Mahapatra et al. 2002), the alanine residue has been modified to a serine (DRSG) in TmDXS.

Despite the relatively high number of DXS genes characterized, no one from a thermophilic microorganism has been reported to date. We have cloned the ORF TM_1770 that encodes the 1-deoxy-D-xylulose 5-phosphate synthase (DXS) in order to facilitate the characterization of the DXS from *T. maritima*. Using the expression construct pET23-TmDXS and the E. coli expression strain BL21 [DE3] Rosetta2, the recombinant enzyme was overexpressed in sufficient amount to obtain 2.5mg of pure and active protein from 1L of culture. Kinetic parameters were determined from the recombinant enzyme and, by site-directed mutagenesis, an important residue for the activity described in *E. coli* and *Agrobacterium tumefaciens* DXS was confirmed to be essential in TmDXS as well.

Materials and Methods

Materials

Biochemical reagents for molecular procedures were obtained from New England biolabs and Promega. Pyruvate and DL-glyceraldehide 3-phosphate was purchased from Sigma and isopropyl β-D-thiogalactoside (IPTG) was purchased from Applichem

Bacterial strains, plasmids and media

Genomic DNA from *Thermotoga maritima* MSB8 strain was kindly provided by Dr J. Querol (Barcelona Science Park-University of Barcelona). *Escherichia coli* $DH5_\alpha$ strain was used for cloning procedures and *Escherichia coli* M15 [pREP4], BL21 [DE3], BL21 [DE3] pLysS and BL21 [DE3] Rosetta2 (Novagen) were used for protein expression. The plasmids used were pQE30 (Qiagen), pT7-7 and pET23(b) (Novagen). Cells were grown in Luria-Bertoni (LB) media (10 g/L triptone, 5 g/1 NaCl and 5 g/l yeast extract) and the suitable antibiotics.

Gene Cloning

Putative TmDXS (TM_1770, ExPaSy number entry Q9X291) was amplified from genomic DNA of MSB8 strain of *Thermotoga maritima* by PCR with *Pfu* polymerase (Promega) and different sets of oligonucleotide primers. The sets of primers were designed based on the expression vector used for the final expression constructs. The gene was amplified for pQE30 (Qiagen) with the primers TMdxsF and TMdxsR (see table 1), for pT7-7 with the primers TMdxsF2 and TMdxsR and for pET23 (Novagen) with the primers TMdxsFNdeI(2) and TMdxsRXhoI. The conditions of all PCR reactions consisted of 33 cycles of 40s, 94ºC; 30s, 60ºC; 230s, 72ºC. The desired 1,8Kb

DNA fragment was electrophoresed in a 1% agarose gel and recovered using a QIAquick gel purification kit (Qiagen). The PCR reactions were digested with restriction enzymes (*BamHI/HindIII* for pQE30, *NdeI/HindIII* for pT7-7 and *NdeI/XhoI* for pET23) and were subsequently cloned into the vectors. Expression plasmid constructs (pQE30-TmDXS, pT7-TmDXS and pET23-TmDXS) were subsequently verified by DNA sequencing.

Site-directed mutagenesis

The H35A point mutation was introduced in the gene encoding TmDXS by site-directed mutagenesis using an oligonucleotide-directed strategy following protocols associated with the QuickChange site-directed mutagenesis procedures (Stratagene). To create the pET23-TmDXS H35A construct, pET23-TmDXS wild type was used as a template and the MutdxstmH35AF and MutdxstmH35AR primers (see Table1) were used in the amplification step. The conditions of the PCR were 18 cycles of 30s, 95ºC; 60s, 63ºC; 650s, 72ºC. The PCR product was digested with *DpnI* enzyme and *E. coli* DH5α competent cells were transformed with the digestion reaction. Plasmids were extracted from isolated colonies and mutations were verified by DNA sequencing. Finally, a construct harboring the point mutation was digested with the same restriction enzymes and subcloned into a digested pET23 vector.

***Escherichia coli* complementation assay**

EcAB4-2 strain cells (derived from *E. coli* MG1655 background with dxs::CAT), harboring a disruption in the *dxs* gene through the chloramphenicol acetyltransferase (CAT) gene insertion and with the MVA+ operon inserted in the chromosome (Campos 2001). The strain is capable to rescue the DXP auxotrophy in the presence of mevalonate through the MVA+ operon. *E. coli EcAB4-2* cells were transformed with the expression plasmids pET23, pET23-EcDXS, pET23-TmDXS and pET23-TmDXS H35A. Transformants were streaked on LB plates supplemented with 1mM isopropyl β-D-thiogalactoside (IPTG), 1mM mevalonate (MVA), 0,02% (w/v) arabinose, 100µg/mL ampicillin, 34µg/mL cloramphenicol and 24 µg/mL kanamycin.

Production and purification of *T.maritima* DXS

Escherichia coli BL21 [DE3] Rosetta 2 cells harboring the appropriate expression vector were grown at 37ºC in six 2-L. Erlenmeyer flasks each containing 500 mL of LB supplemented with 100µg/mL ampicillin and 34µg/mL cloramphenicol. At OD_{600} 0.6, IPTG and thiamine were added to 1mM. After 16h of incubation at 22ºC and 180 r.p.m, cells were harvested by centrifugation at 1500g for 15 min at 4ºC, resuspended in 60mL of lysis buffer (40mM Tris-HCl pH 8.0, 100mM NaCl, 10mM imidazole, 1mg/mL lysozime, 0.5mM EDTA and complete mini-EDTA-free protease-inhibitor cocktail (Roche)) and incubated for 20 minutes at 4ºC. The resuspended cells were sonicated (5 x 30s bursts with 1min cooling in melting ice) and were centrifuged at 11,000g for 30 min at 4ºC; the supernatant was incubated with protamine sulphate (1,25mg/mL) for 15 min, and then centrifuged at 13,000g for 30min at 4ºC. Supernatants were transferred to clean centrifuge tubes, immersed in 70ºC water bath for 10 min and further centrifuged for 10 min at 13,000g. TmDXS was purified from the heat-treated supernatant by immobilized metal affinity chromatography (IMAC) using Hi-trap Chelating columns (GE Healthcare) attached to an FPLC system (BioLogic Workstation, BIORAD). The resin was washed with 40mM Tris–HCl buffer, pH 8.0, containing 0.1M NaCl, 10mM imidazole and 4mM β-mercaptoethanol. The His6-TmDXS was retained on the resin and eluted using an imidazole gradient (10–500mM) in 40mM Tris–HCl buffer, pH 8.0, containing 0.1M NaCl buffer. The fractions containing TmDXS were resolved by SDS–PAGE and stained with Coomassie brilliant blue R-250. TmDXS preparations were desalted by gel filtration on disposable Sephadex G-25 columns (PD10) (GE Healthcare) equilibrated in 40mM Tris–HCl buffer with 0.1M NaCl, distributed into aliquots, snap-frozen

with liquid N_2 and stored at -80ºC until use.

Enzyme assays and kinetics

DXP synthase activity was determined by a modification of a previously described coupled assay (Jomaa, Wiesner et al. 1999). Reaction mixtures contained 100mM Tris-HCl (pH 7.4), 1.5mM TPP, 0.15mM NADPH, 1mM $MgCl_2$, 1mM $MnCl_2$, 1mM DTT, 1mM pyruvate, 1mM DL-glyceraldehyde 3-phosphate, 45µg/mL of *Synechocystis sp*. 1-deoxy-D-xylulose 5-phosphate reductoisomerase (produced in our laboratory 10.48U/mg) and 25,5 µg/mL TmDXS on 96-well microtiter plate and monitored at 340nm in a Benchmark Plus microplate spectrophotometer (BIORAD). A first step (named DXS-activity step) was done in 100mM Tris-HCl (pH 7.4), 1.5mM TPP, 1mM DTT, 1mM $MgCl_2$, 1mM pyruvate, 1mM DL-glyceraldehyde 3-phosphate and TmDXS in a final volume of 100µL. The reaction was allowed to proceed and stopped by immersion in liquid N_2. In the second step (named DXR-activity step), 100µL of DXR reaction mixture (100mM Tris-HCl (pH=7.4), 1mM DTT, 0.3mM NADPH, 2mM $MnCl_2$, and SsDXR) was added and the reaction was monitored at 340nm. This two-step activity method was designed based on the optimum activity of the enzymes and the reaction was monitored based on the time of reaction of the DXR-activity step (the amount of DXP formed). The DXS-activity step was done at different temperatures for 5 minutes and the DXR-activity step was monitored at room temperature for 1000s.

Optimum of activity. The DXS-activity step was done at 37ºC, 50ºC, 55ºC, 60ºC, 65ºC, 70ºC, 75ºC, 80ºC and 90ºC

Kinetic parameters. K_{pyr} and K_{cat} were determined by using a constant D,L- glyceraldehyde 3-phosphate concentration (1mM) and varying the pyruvate concentrations (1µM, 5µM, 10µM, 25µM, 50µM, 75µM, 100µM, 125µM, 150µM, 200µM, 300µM and 500µM). Kinetic parameters were obtained by fitting the reaction rates using the Grafit program.

Acknowledgements

The Spanish Ministry of Science and Technology supported this work through the projects BIO2002-04419-C02-02. The 'Generalitat de Catalunya' also supported this work through the grant 2005SGR00914

TMdxsF	5’ TC**GGATC**CATGCTTCTGGACGAGATCAAG 3’
TMdxsF2	5’ CGTA**CATATG**CTTCTGGACGAGATCA 3’
TMdxsR	5’ ATT**AAGCTT**TCATACTTTCCCCTCCCTAC 3’
TMdxsFNdeI(2)	5’ GGAC**CATATG**CTTCTTCTGGACGAGATCAA 3’
TMdxsRXhoI	5’ CC**GCTCGAG**TACTTTCCCCTCCCTACT 3’
MutdxstmH35AF	5’ AAGAACGGGGG**AGCC**CTCGCATCG 3’
MutdxstmH35AR	5’ CGATGCGAGG**GGCT**CCCCCGTTCTT 3’

Table 1. Sequence of primers used in PCR reactions. For amplification primers, enzymatic targets for cloning endonucleases are in boldface. Primers used in mutagenesis have the nucleotide changes in boldface and the mutated codons (amino acid changes) underlined.

Bibliography

Achenbach-Richter, L., R. Gupta, et al. (1987). "Were the original eubacteria thermophiles?" Syst Appl Microbiol **9**: 34-9.

Alos, E., M. Cercos, et al. (2006). "Regulation of color break in citrus fruits. Changes in pigment profiling and gene expression induced by gibberellins and nitrate, two ripening retardants." J Agric Food Chem **54**(13): 4888-95.

Altincicek, B., M. Hintz, et al. (2000). "Tools for discovery of inhibitors of the 1-deoxy-D-xylulose 5-phosphate (DXP) synthase and DXP reductoisomerase: an approach with enzymes from the pathogenic bacterium Pseudomonas aeruginosa." FEMS Microbiol Lett **190**(2): 329-33.

Bailey, A. M., S. Mahapatra, et al. (2002). "Identification, cloning, purification, and enzymatic characterization of Mycobacterium tuberculosis 1-deoxy-D-xylulose 5-phosphate synthase." Glycobiology **12**(12): 813-20.

Bean, L. E., W. H. Dvorachek, Jr., et al. (2001). "Analysis of the pdx-1 (snz-1/sno-1) region of the Neurospora crassa genome: correlation of pyridoxine-requiring phenotypes with mutations in two structural genes." Genetics **157**(3): 1067-75.

Bouvier, F., A. d'Harlingue, et al. (1998). "Dedicated roles of plastid transketolases during the early onset of isoprenoid biogenesis in pepper fruits1." Plant Physiol **117**(4): 1423-31.

Campos, N., M. Rodriguez-Concepcion, et al. (2001). "Escherichia coli engineered to synthesize isopentenyl diphosphate and dimethylallyl diphosphate from mevalonate: a novel system for the genetic analysis of the 2-C-methyl-d-erythritol 4-phosphate pathway for isoprenoid biosynthesis." Biochem J **353**(Pt 1): 59-67.

Cane, D. E., Y. Hsiung, et al. (1998). "Biosynthesis of vitamin B6: the oxidation of 4-(phosphohydroxy)-L-threonine by PdxA." J Am Chem Soc. **120**: 1936–1937.

Cassera, M. B., F. C. Gozzo, et al. (2004). "The methylerythritol phosphate pathway is functionally active in all intraerythrocytic stages of Plasmodium falciparum." J Biol Chem **279**(50): 51749-59.

Ehrenshaft, M., P. Bilski, et al. (1999). "A highly conserved sequence is a novel gene involved in de novo vitamin B6 biosynthesis." Proc Natl Acad Sci U S A **96**(16): 9374-8.

Ehrenshaft, M. and M. E. Daub (2001). "Isolation of PDX2, a second novel gene in the pyridoxine biosynthesis pathway of eukaryotes, archaebacteria, and a subset of eubacteria." J Bacteriol **183**(11): 3383-90.

Estevez, J. M., A. Cantero, et al. (2001). "1-Deoxy-D-xylulose-5-phosphate synthase, a limiting enzyme for plastidic isoprenoid biosynthesis in plants." J Biol Chem **276**(25): 22901-9.

Estevez, J. M., A. Cantero, et al. (2000). "Analysis of the expression of CLA1, a gene that encodes the 1-deoxyxylulose 5-phosphate synthase of the 2-C-methyl-D-erythritol-4-phosphate pathway in Arabidopsis." Plant Physiol **124**(1): 95-104.

Eubanks, L. M. and C. D. Poulter (2003). "Rhodobacter capsulatus 1-deoxy-D-xylulose 5-phosphate synthase: steady-state kinetics and substrate binding." Biochemistry **42**(4): 1140-9.

Fitzpatrick, T. B., N. Amrhein, et al. (2007). "Two independent routes of de novo vitamin B6 biosynthesis: not that different after all." Biochem J **407**(1): 1-13.

Frank, R. A., F. J. Leeper, et al. (2007). "Structure, mechanism and catalytic duality of thiamine-dependent enzymes." Cell Mol Life Sci **64**(7-8): 892-905.

Fukuchi, S. and K. Nishikawa (2001). "Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria." J Mol Biol **309**(4): 835-43.

Garrido-Franco, M. (2003). "Pyridoxine 5'-phosphate synthase: de novo synthesis of vitamin B6 and beyond." Biochim Biophys Acta **1647**(1-2): 92-7.

Gong, Y. F., Z. H. Liao, et al. (2006). "Molecular cloning and expression profile analysis of Ginkgo biloba DXS gene encoding 1-deoxy-D-xylulose 5-phosphate synthase, the first committed enzyme of the 2-C-methyl-D-erythritol 4-phosphate pathway." Planta Med **72**(4): 329-35.

Harker, M. and P. M. Bramley (1999). "Expression of prokaryotic 1-deoxy-D-xylulose-5-phosphatases in Escherichia coli increases carotenoid and ubiquinone biosynthesis." FEBS Lett **448**(1): 115-9.

Hawkins, C. F., A. Borges, et al. (1989). "A common structural motif in thiamin pyrophosphate-binding enzymes." FEBS Lett **255**(1): 77-82.

Huber, R., T. A. Langworthy, et al. (1986). "*Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90 ºC." *Arch. Microbiol.* **144**: 324-333.

John, R. A. (1995). "Pyridoxal phosphate-dependent enzymes." Biochim Biophys Acta **1248**(2): 81-96.

Kakugawa, S., S. Fushinobu, et al. (2007). "Characterization of a thermostable carboxylesterase from the hyperthermophilic bacterium Thermotoga maritima." Appl Microbiol Biotechnol **74**(3): 585-91.

Kim, B. R., S. U. Kim, et al. (2005). "Differential expression of three 1-deoxy-D: -xylulose-5-phosphate synthase genes in rice." Biotechnol Lett **27**(14): 997-1001.

Kuzuyama, T., M. Takagi, et al. (2000). "Cloning and characterization of 1-deoxy-D-xylulose 5-phosphate synthase from Streptomyces sp. Strain CL190, which uses both the mevalonate and nonmevalonate pathways for isopentenyl diphosphate biosynthesis." J Bacteriol **182**(4): 891-7.

Laber, B., W. Maurer, et al. (1999). "Vitamin B6 biosynthesis: formation of pyridoxine 5'-phosphate from 4-(phosphohydroxy)-L-threonine and 1-deoxy-D-xylulose-5-phosphate by PdxA and PdxJ protein." FEBS Lett **449**(1): 45-8.

Lange, B. M., M. R. Wildung, et al. (1998). "A family of transketolases that directs isoprenoid biosynthesis via a mevalonate-independent pathway." Proc Natl Acad Sci U S A **95**(5): 2100-4.

Le Fourn, C., M. L. Fardeau, et al. (2008). "The hyperthermophilic anaerobe Thermotoga Maritima is able to cope with limited amount of oxygen: insights into its defence strategies." Environ Microbiol **10**(7): 1877-87.

Lee, J. K., D. K. Oh, et al. (2007). "Cloning and characterization of the dxs gene, encoding 1-deoxy-d-xylulose 5-phosphate synthase from Agrobacterium tumefaciens, and its overexpression in Agrobacterium tumefaciens." J Biotechnol **128**(3): 555-66.

Lindqvist, Y., G. Schneider, et al. (1992). "Three-dimensional structure of transketolase, a thiamine diphosphate dependent enzyme, at 2.5 A resolution." Embo J **11**(7): 2373-9.

Lois, L. M., N. Campos, et al. (1998). "Cloning and characterization of a gene from Escherichia coli encoding a transketolase-like enzyme that catalyzes the synthesis of D-1-deoxyxylulose 5-phosphate, a common precursor for isoprenoid, thiamin, and pyridoxol biosynthesis." Proc Natl Acad Sci U S A **95**(5): 2105-10.

Lois, L. M., M. Rodriguez-Concepcion, et al. (2000). "Carotenoid biosynthesis during tomato fruit development: regulatory role of 1-deoxy-D-xylulose 5-phosphate synthase." Plant J **22**(6): 503-13.

Mandel, M. A., K. A. Feldmann, et al. (1996). "CLA1, a novel gene required for chloroplast development, is highly conserved in evolution." Plant J **9**(5): 649-58.

Medrano-Soto, A., G. Moreno-Hagelsieb, et al. (2004). "Successful lateral transfer requires codon usage compatibility between foreign genes and recipient genomes." Mol Biol Evol **21**(10): 1884-94.

Miller, B., T. Heuser, et al. (1999). "A Synechococcus leopoliensis SAUG 1402-1 operon harboring the 1-deoxyxylulose 5-phosphate synthase gene and two additional open reading frames is functionally involved in the dimethylallyl diphosphate synthesis." FEBS Lett **460**(3): 485-90.

Munoz-Bertomeu, J., I. Arrillaga, et al. (2006). "Up-regulation of 1-deoxy-D-xylulose-5-phosphate synthase enhances production of essential oils in transgenic spike lavender." Plant Physiol **142**(3): 890-900.

Nelson, K. E., R. A. Clayton, et al. (1999). "Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima." Nature **399**(6734): 323-9.

Osmani, A. H., G. S. May, et al. (1999). "The extremely conserved pyroA gene of Aspergillus nidulans is required for pyridoxine synthesis and is required indirectly for resistance to photosensitizers." J Biol Chem **274**(33): 23565-9.

Penuelas, J. and S. Munne-Bosch (2005). "Isoprenoids: an evolutionary pool for photoprotection." Trends Plant Sci **10**(4): 166-9.

Querol, J., M. Rodriguez-Concepcion, et al. (2001). "Essential role of residue H49 for activity of Escherichia coli 1-deoxy-D-xylulose 5-phosphate synthase, the enzyme catalyzing the first step of the 2-C-methyl-D-erythritol 4-phosphate pathway for isoprenoid Synthesis." Biochem Biophys Res Commun **289**(1): 155-60.

Reynen, M. and H. Sahm (1988). "Comparison of the structural genes for pyruvate decarboxylase in different Zymomonas mobilis strains." J Bacteriol **170**(7): 3310-3.

Schenk, G., R. G. Duggleby, et al. (1998). "Properties and functions of the thiamin diphosphate dependent enzyme transketolase." Int J Biochem Cell Biol **30**(12): 1297-318.

Settembre, E., T. P. Begley, et al. (2003). "Structural biology of enzymes of the thiamin biosynthesis pathway." Curr Opin Struct Biol **13**(6): 739-47.

Sprenger, G. A., U. Schorken, et al. (1997). "Identification of a thiamin-dependent synthase in Escherichia coli required for the formation of the 1-deoxy-D-xylulose 5-phosphate precursor to isoprenoids, thiamin, and pyridoxol." Proc Natl Acad Sci U S A **94**(24): 12857-62.

Spurgeon, S. and J. Porter (1981). in Biosynthesis of Isoprenoid Compounds **1**: 1-46.

Sudarsan, N., J. E. Barrick, et al. (2003). "Metabolite-binding RNA domains are present in the genes of eukaryotes." Rna **9**(6): 644-7.

Tambasco-Studart, M., O. Titiz, et al. (2005). "Vitamin B6 biosynthesis in higher plants." PNAS **102(38)**: 13687-92.

Thore, S., M. Leibundgut, et al. (2006). "Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand." Science **312**(5777): 1208-11.

Titiz, O., M. Tambasco-Studart, et al. (2006). "PDX1 is essential for vitamin B6 biosynthesis, development and stress tolerance in Arabidopsis." Plant J **48**(6): 933-46.

Vargas, M., K. Kashefi, et al. (1998). "Microbiological evidence for Fe(III) reduction on early Earth." Nature **395**(6697): 65-7.

Wagner, S., A. Bernhardt, et al. (2006). "Analysis of the Arabidopsis rsr4-1/pdx1-3 mutant reveals the critical function of the PDX1 protein family in metabolism, development, and vitamin B6 biosynthesis." Plant Cell **18**(7): 1722-35.

Webb, M. E., A. Marquet, et al. (2007). "Elucidating biosynthetic pathways for vitamins and cofactors." Nat Prod Rep **24**(5): 988-1008.

Xiang, S., G. Usunow, et al. (2007). "Crystal structure of 1-deoxy-D-xylulose 5-phosphate synthase, a crucial enzyme for isoprenoids biosynthesis." J Biol Chem **282**(4): 2676-82.

Yang, H., H. Ichinose, et al. (2006). "Characterization of a thermostable endo-beta-1,4-D-galactanase from the hyperthermophile Thermotoga maritima." Biosci Biotechnol Biochem **70**(2): 538-41.

Zein, F., Y. Zhang, et al. (2006). "Structural insights into the mechanism of the PLP synthase holoenzyme from Thermotoga maritima." Biochemistry **45**(49): 14609-20.

Zhang, M., K. Li, et al. (2008). "Identification and characterization of class 1 DXS gene encoding 1-deoxy-D: -xylulose-5-phosphate synthase, the first committed enzyme of the MEP pathway from soybean." Mol Biol Rep.

Procesamiento autoproteolítico de la 1-desoxi-D-xilulosa 5-fosfato sintasa

Victor Giménez-Oya, Meritxell Antolin-Llovera, Santiago Imperial, Albert Boronat

Dept. de Bioquímica i Biología Molecular, Facultat de Biologia, Universitat de Barcelona,, Avda Diagonal 645, E-08028 Barcelona, Spain.

Running title: Descubrimiento de reorganizaciones intramoleculares en la 1-desoxi-D-xilulosa 5-fosfato sintasa que podrían producir autoproteolisis

Palabras clave**:** 1-desoxi-D-xilulosa 5-fosfato sintasa, metileritritol fosfato, enlace ester, succinimida, proceso autoproteolítico, biosíntesis de isoprenoides, tiamina.

RESUMEN

La enzima 1-desoxi-D-xilulosa 5-fosfato sintasa (DXS) cataliza la condensación de (hidroxi)tiamina derivada de la decarboxilación de piruvato en el grupo carbonilo de D-gliceraldehido 3-fosfato (GAP) para producir 1-desoxi-D-xilulosa 5-fosfato (DXP). La DXP es el primer precursor de la ruta plastídica del metileritritol 4-fosfato de síntesis de isoprenoides y también es un intermediario de la síntesis de tiamina y piridoxol. Está comúnmente aceptado que la DXS juega un papel regulatorio en la biosíntesis de isoprenoides bacterianos y plastídicos, y que los niveles de DXP son controlados a nivel post-transcripcional. En este estudio mostramos evidencias que sugieren que la DXS de *Escherichia coli* se procesa por autoproteólisis. La segmentación del enlace entre la Ser^{206} y la Ser^{207} en la EcDXS nos permite pensar en un proceso relacionado con la formación de un enlace éster. Interesantemente, la hidroxilamina aumenta el proceso de proteólisis pero muestra la presencia de dos reordenaciones intramoleculares relacionas con residuos Asp en los residuos Asp^{152} y Asp^{294}, que están totalmente conservados en las DXS. La localización de los enlaces segmentados por la proteólisis y de los producidos por hidroxilaminosis parece mostrar la presencia de dos procesos diferentes. El residuo Asp^{152} es un residuo conservado presente en el motivo de unión a TPP ($GDGX_{25\text{-}30}N$) de las DXSs el cual juega un importante papel en la estabilización del TPP y el ión Mg^{2+}. La modificación post-traduccional de este Asp podría tener un importante rol en la regulación de la actividad DXS. Este residuo esta conservado en otras enzimas TPP-dependientes como la transcetolasa o la subunidad E1 de la piruvato deshidrogenasa.

Autoproteolytic processing of 1-deoxy-D-xylulose 5-phosphate synthase

Victor Giménez-Oya, Meritxell Antolin-Llovera, Santiago Imperial, Albert Boronat

Dept. de Bioquímica i Biología Molecular, Facultat de Biologia, Universitat de Barcelona,, Avda Diagonal 645, E-08028 Barcelona, Spain.

Running title: Discerning intramolecular rearrangements in 1-deoxy-D-xylulose 5-phosphate synthase that may induce autoproteolysis

Key words**:** 1-deoxy-D-xylulose 5-phosphate synthase, methylerythritol phosphate, ester-bond, succinimide residue, autoproteolysis, isoprenoid biosynthesis, thiamine

ABSTRACT

The enzyme 1-deoxy-D-xylulose 5-phosphate synthase (DXS) catalyzes the condensation of (hydroxyethyl)thiamin, derived from the decarboxylation of pyruvate, on the carbonyl group of D-glyceraldehyde 3-phosphate (GAP) to yield 1-deoxy-D-xylulose 5-phosphate (DXP). DXP is the first precursor of the plastidial methylerythrititol 4 phosphate pathway for isoprenoid biosynthesis and also an intermediate in the synthesis of thiamine and pyridixol. It is widely accepted that DXS plays a key regulatory role in the biosynthesis of bacterial and plastidial isoprenoids and that DXP levels are mainly controlled at post-transcriptional level. In this paper we show evidences that *Escherichia coli* DXS is processed most probably by autoproteolysis. Cleavage of DXS between Ser^{206} and Ser^{207} most likely involve an ester bond-related process but, interestingly, the increased proteolysis with hydroxylamine treatment showed the presence of at least two Asp-related cyclic intermediates at residues Asp^{152} and Asp^{294} which are totally conserved in DXSs. Both the cleavage sites at Ser residues and the sites obtained by hydroxylaminosis seem to involve different mechanisms. Asp^{152} is a conserved residue within the thiamine pyrophosphate (TPP) binding motif ($GDGX_{25\text{-}30}N$) of DXS and plays an essential role in the stabilization of the TPP and Mg^{2+} ion. The posttranslational modification of this Asp residue could have a relevant role on controlling DXS activity. Interestingly, these Asp residues are also conserved in other TPP-depending enzymes, like transketolase and pyruvate dehydrogenase.

INTRODUCTION

Isoprenoids are ubiquitous compounds that play essential functions in all life. Isoprenoids also represent one of the largest families of natural compounds, particularly in plants (Sacchettini and Poulter 1997). In spite of the high diversity of structures and functions, all isoprenoids are synthesized from the same 5-carbon precursors: isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP) (Chappell 1995). IPP and DMAPP are synthesized from two anabolic pathways: the mevalonate (MVA) pathway (Spurgeon and Porter 1981) and the methylerythritol 4-phosphate (MEP) pathway(Rodríguez-Concepción and Boronat 2002; Rohmer, Grosdemange-Billiard et al. 2004). Both pathways are differentially distributed in the kingdoms of life. While the MVA pathway is exclusively present in archaebacteria, fungus and animals, the MEP pathway is mainly present in eubacteria, including most of the pathogenic species, and in some protozoa, like *Plamodium falciparum*. Interestingly, both pathways are present in plants although in separate cell compartments. Thus, whereas the MVA pathway is present in the cytosol/endoplasmic reticulum, the MEP pathway is found in the plastids (Lichtenthaler, Rohmer et al. 1997; Eisenreich, Schwarz et al. 1998; Rohmer 1999; Lichtenthaler 2000; Boucher, Huber et al. 2001; Rodríguez-Concepción, Fores et al. 2004; Rohmer, Grosdemange-Billiard et al. 2004).

The first step of the MEP pathway consists of a transketolase-type condensation of pyruvate and glyceraldehyde 3-phosphate to form 1-deoxy-D-xylulose 5-phosphate (DXP) catalyzed by the enzyme DXP synthase (DXS) (Lange, Wildung et al. 1998). This reaction is a thiamine pyrophosphate (TPP)-dependent acycloin condensation (Sprenger, Schorken et al. 1997). This type of reaction is due to the secondary activity of TPP that has also been documented in pyruvate decarboxylating enzymes like the E1 component of pyruvate dehydrogenase (Sprenger, Schorken et al. 1997). Moreover, it is worth noting that DXP acts as a precursor for the synthesis of TPP and pyridoxal (Sprenger, Schorken et al. 1997), which play essential roles as enzyme cofactors (John 1995; Fitzpatrick, Amrhein et al. 2007; Frank, Leeper et al. 2007).

The reaction catalyzed by DXS has been described as one of the main rate-limiting steps of the MEP pathway in microorganisms and plants (Estevez, Cantero et al. 2001) and it has been proposed as one of the main regulatory enzymes of the MEP pathway. Several line of evidence has let to propose that DXS activity is mainly controlled at post-transcriptional level. Although very little is known on the mechanisms involved in the control of DXS, it has recently been proposed that protein degradation mediated the ClpP could represent a major process to adjust DXS levels in the plastids and thus controlling the flux of MEP pathway intermediates in the plastids (Flores-Perez, Sauret-Gueto et al. 2008). As a novel mechanism to control DXS levels and enzyme activity here we report that DXS is autoproteolytically cleaved at defined amino acid residues. The role of these post-translational modifications on the turnover and/or activity of DXS are discussed.

MATERIALS AND METHODS

Construction of E. coli expression vectors.

Escherichia coli DXS gene was cloned into the *NdeI/EcoRI* sites of pET28(a) and the *NdeI/XhoI* sites of pET23(b). These constructs encoded DXS variants containing N- and C-terminal extensions with a 6xHis tag, respectively. The obtained plasmids were named pET23-EcDXS and pET28-EcDXS respectively. A construct expressing an *E. coli* DXS derivative harboring a deletion extending from residues 191 to 239 was prepared as follows: i) the coding region from residues 1 to 190 of EcDXS was amplified with the primers T7promoter/190RDXSEco(SmaI) (which added three CCC nucleotides after the coding region of residue 190) (Table 1) and subcloned into pBluescript II (Stratagene) previously digested with *SmaI* to generate pBS-Nt190DXSEco, ii) the coding region extending from residues 240 to 620 of EcDXS was amplified with primers 240FDXSEco(SmaI)/T7terminator (which added three GGG nucleotides before the coding region for residue 240) (Table 1) and cloned into the *SmaI* site of pBluescript II to generate plasmid pBS-240CtDXSEco, iii) Plasmids pBS-Nt190DXSEco and pBS-240CtDXSEco were digested with *NdeI/SmaI* and *SmaI/XhoI*, respectively an cloned into a pET23(b) digested with *NdeI/XhoI* to generate construct pET23-EcDXSΔ191-239. An additional construct containing a Gly-flexible loop replacing the deleted region was constucted as follows: i) two complementary primers encoding the loop (named Flexlinker(P-G-S)D/ Flexlinker(P-G-S)Com) (Table 1) were heated for 5 min at 95ºC and let to hybridize for 3 hours at room temperature, ii) the duplex sequence encoding the Gly-flexible loop was cloned in-frame into pET23-EcDXS-Δ191-239 digested with *SmaI*. The generated construct was named pET23-EcDXS-Δ191-239[G-loop].

The coding region of *Arabidopsis thaliana DXS* without the transit peptide was amplified from a cDNA population and was cloned into a pET23(b) expression vector (Cairó 2009).

Thermotoga maritima DXS was cloned into the *NdeI/XhoI* sites of pET23(b). This construct encoded a DXS variant with a C-terminal extension containing a 6xHis tag.

Plamid pGEX derivatives were prepared as follows. The coding region of GFP from plasmids pGFP-PAU was excised with *NotI* and cloned in-frame into pGEX-5X-2 (GE healthcare) previously digested with *NotI* and dephosphorylated by SAP (USB corporation) and the obtained construct named pGST-GFP. The DNA region encoding amino acid residues 158 to 240 of EcDXS was amplified with primers 158FDXSEco/240RDXSEco(Table 1). This PCR product was inserted in-frame between the GST and the GFP encoding regions through *SmaI* digestion, obtaining the construct pGST-Ec158/240-GFP. The protein-encoding DNA from all the constructs were amplified with the primers GSTNdeIF(2)/GFPSacIR(pET23b) and subcloned into a pET23(b) vector digested with *NdeI/SacI* . The constructs were named pGST-GFPHis, pGST-Ec150,240-GFPHis and pGST-At226,318-GFPHis.

Complementation assays in *E. coli*

Complementation assays were performed by transforming the *E. coli dxs*-deficient strain EcAB4–2 with plasmids pET23-EcDXS-Δ191-239 and pET23-EcDXS-Δ191-239(G-loop). In strain EcAB4–2 the *dxs* gene is disrupted by the insertion of the *CAT* gene, encoding for cloramphenicol acetyltransferase and contains a copy of the MVA+ operon (Campos, Rodriguez-Concepcion et al. 2001) stably integrated in the chromosome, enabling cells to synthesize IPP from mevalonate supplied exogenously in the growth medium. The pPET23-EcDXS plasmid, expressing the full-length *Ec*DXS, was used as a positive control.

Expression and purification of recombinant proteins in *E. coli*

Protein expression from pET23 and PET28 plasmid derivatives were performed as follows. *Escherichia coli* BL21 [DE3] cells harboring the appropriate expression vectors were grown at 37ºC to an OD_{600} of 0.6. Then IPTG and thiamine were added to a final concentration of 1mM each. After 16h of incubation at 22ºC cells were harvested by centrifugation at 1,500xg for 15 min at 4ºC, resuspended in 60 mL of lysis buffer (40mM Tris-HCl pH 8.0, 100mM NaCl, 10mM Imidazole, 1mg/mL lysozime, 0.5mM EDTA and complete mini-EDTA-free protease-inhibitor cocktail (Roche)) and incubated for 20 minutes at 4ºC. The resuspended cells were sonicated and centrifuged at 11,000xg for 30 min at 4ºC. The supernatant was incubated with protamine sulphate (1,25mg/mL) for 15 min, and then centrifuged at 13,000x for 30min at 4ºC. Proteins were purified by immobilized metal affinity chromatography (IMAC) using Hi-trap Chelating columns (GE Healthcare) attached to an FPLC system (BioLogic Workstation, BIO-RAD). The resin was washed with 40mM Tris–HCl buffer, pH 8.0, containing 0.1M NaCl, 10 mM imidazole and 4 mM β-mercaptoethanol. The proteins retained on the resin were eluted using an imidazole gradient (10 to 500 mM) in 40 mM Tris–HCl buffer, pH 8.0, containing 0.1M NaCl buffer.

Protein expression from the pGEX plasmids were performed as follows. Cell cultures and lysates were prepared as indicated above for the pET23 and pET28 derivatives. Overexpressed proteins were purified by affinity chromatography using Glutathione (GSH)-Sepharose (GE Healthcare). The resin was washed with 40 mM Tris–HCl buffer, pH 8.0, containing 0.1M NaCl and the proteins eluted with buffer containing 25 mM reduced GSH.

SDS-PAGE and immunodetection

Proteins separated by SDS-PAGE were electrotransferred for 2 h into a PVDF membrane. Blots were blocked with 5% skimmed milk followed by incubation with primary antibody (polyclonal rabbit anti-GST or polyclonal mouse anti-His). Primary antibodies were recognized with a secondary anti-rabbit or anti-mouse secondary antibodies conjugated to horseradish peroxidase followed by enhanced chemiluminiscence using the ECL kit (GE healthcare) detection.

N-terminal protein sequencing

Proteins separated on SDS-PAGE were blotted into a special PVDF membrane (Sequi-Blot PVDF membrane Bio-Rad). Proteins were briefly stained in comassie and the bands of interests were excised. The bands were identified by N-terminal sequencing for 5 cycles using an Applied Biosystems Procise 494 N-terminal protein sequencer.

In vitro cell-free expression of recombinant DXS

DXS from *E. coli*, *T. maritima* and *A. thaliana* were expressed *in vitro* using the TNT Coupled Reticulocyte Lysate System (Promega) radiolabeled with L-[4,5-^{3}H] Leucine (Amersham Biosciences). The reaction mix contained 121 Ci/mmol of L-[4,5-^{3}H] Leucine (for 8 samples). The reaction mix containing the constructs pET23-EcDXS, pET23-AtDXS, pET23-TmDXS, respectively, was incubated for 90 min at 30ºC and the reaction stopped by heating at 95ºC for 5 min. Proteins were resolved by 10% SDS-PAGE. After electrophoresis the gel was incubated for 30 minutes in fixing solution (isopropanol : H_20 : acetic acid, 25:65:10). The radioactive signal was amplified with Amplify Fluorographic Reagent (Amersham Biosciences) during 30 minutes and the gel was then incubated with drying solution (7% acetic acid, 7% methanol, and 2% glycerol) for 5 minutes and let to dry. The dried gel was contacted with a Hyperfilm^{TM} MP (Amersham biosciences) during 1 month at -80ºC. The film was

fixed (AGFA G350 fixer) and developed (AGFA G150 developer).

Refolding of purified recombinant EcDXS

Purified recombinant EcDXS was denatured in a SDS-PAGE protein loading buffer and electrophoresed on a 10% polyacrylamide gel. The 70 kDa band was cut and was electroeluted (30mA, 4ºC, 16 h) in electroelution buffer (40 mM Tris-HCl pH 8.0, 100 mM NaCl and PMSF 0.1mg/mL). The electroeluted protein was incubated with an anion exchanger (SERDOLIT^{R}) 1h at 4ºC to remove residual SDS and some polyacrylamide impurities. Then, the protein sample was dialyzed during 6 h at 4º C in refolding buffer (40mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM DTT, 1 mM TPP, 1 mM $MgCl_2$ and PMSF 0.1mg/mL). EcDXS was also electroeluted and dialyzed with a denaturing buffer consisting of 40mM Tris-HCl pH 8.0, 100 mM NaCl and 8M urea).

Hydroxylamine treatment of recombinant EcDXS

The buffer of the purified EcDXS from pET23-EcDXS expression vector was exchanged for a non-nucleophile buffer (50mM Hepes pH 8.0, 100mM NaCl). The buffer for denatured samples was also exchanged (8M urea, 50mM Hepes pH 8.0, 100mM NaCl). Then, both the denatured and non-denatured samples were incubated at room temperature with hydroxylamine (250mM) under both neutral and acidic conditions. At 15 min and o/n of incubation times samples were subjected to SDS-PAGE and immunoblot analysis against the C-terminal His-tag.

RESULTS

Experimental evidences for autoproteolytic cleavage of purified DXS

Purified recombinant DXS expressed in *E. coli* always contains additional bands of lower molecular weight when analyzed by SDS-PAGE in all tested conditions(Fig. 1). As shown for *Escherichia coli* DXS, SDS-PAGE analysis of the recombinant C-terminal His-tagged enzyme (EcDXSc) purified by IMAC results in a typical pattern of bands with sizes around 69, 50, 40, 30 and 20 kDa (Fig. 1A). Although the size of the major band corresponds to the molecular weight predicted for the recombinant EcDXS (about 69 kDa, including the sequence containing the 6xHis affinity-tag), the aditional bands represent a significant proportion of the purified protein. Since all the aditional bands show a lower molecular weight than the observed for the full-lenght recombinant EcDXS it was tempting to speculate that they may actually correspond to proteolytic fragments generated during the preparation of cell extracts prior to purification. The presence of these contaminating bands could not be prevented by the inclusion of any protease inhibitors tested present in the protein extraction buffer.

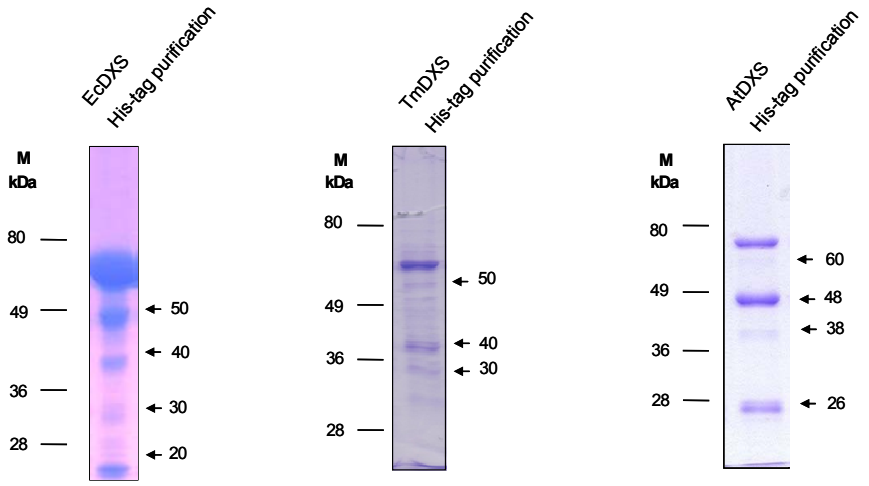


Fig.1 Bacterial expression of the EcDXS, TmDXS and AtDXS. Recombinant proteins were analyzed by10% SDS-PAGE. Molecular weight markers are marked on the left of each lane and the migration weights of the peptides are marked on the right with an arrow.

In any case, a Western blot analysis was performed using an antibody targeted against the C-terminal affinity tag. The results (shown in Fig. 2A) show that only the ~50 and ~40 kDa fragments reacted with the antibody. These results suggest that proteolytic fragments that do not have the 6xHis affinity tag could interact with the other ones that have it during the purification process or that DXS could be continuously processed even after affinity purification.

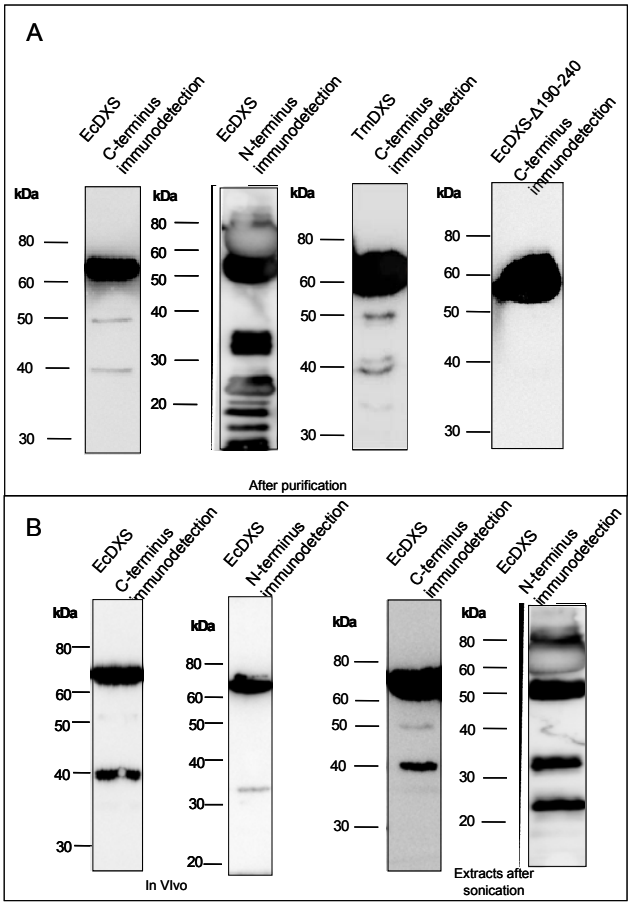


Fig.2 DXS derived products detected by immunoblot analysis. EcDXS with His-tag at the N-terminus (EcDXSn), EcDXS with His-tag at the C-terminus (EcDXSc), TmDXS with His-tag at the C-terminus (TmDXSc) or a EcDXS with a deleted loop and the His-tag at the C-terminus (EcDXS-Δ191-239c) were analysed with mouse anti-His polyclonal antibody. A Analysis of the recombinant proteins after protein purification. B Analysis of the recombinant EcDXSc and EcDXSn. On the left (*in vivo* samples), briefly, a culture of overexpressing cells was centrifuged and the pellet was directly resuspended in sample buffer. On the right, analysis of the recombinant EcDXS after sonication prior to affinity purification.

To check whether proteolysis was due to the presence of a contaminating protease(s), the region of the polyacrylamide gel containing the full-lenght 69 kDa band was cut and the intact EcDXS was recovered by electroelution, allowed to refold and, finally, analyzed by SDS-PAGE and Western blot using an anti-C-terminal 6xHis tag antibody. As shown in Fig. 3B the renatured protein fraction contained two main degradation products equivalent to those previously observed during enzyme preparations. It is worth noting however that the mobility of the ~50 kDa fragment present in the refolded protein sample was slightly lower than that found in the previously purified samples. Interestingly, no degradation products were observed when the electroeluted EcDXS protein was maintained under denaturing conditions (buffer with 8 M urea) (Fig. 3B). All together these results suggest that EcDXS is cleaved by means of an autoproteolytic process requiring the proper folding of the protein.
Purified recombinant DXS preparations from *Thermotoga maritima* (TmDXS) and *Arabidopsis thaliana* (AtDXS) also showed a pattern of degradation fragments with features similar to those reported for EcDXS (Fig.s 1B and

1C). After IMAC purification of recombinant TmDXS (predicted size of 68 kDa including the 6xHis affinity tag at the C-terminus) additional bands of 68, 50, 40 and 30 kDa were detected, along with a set of less intense bands, after SDS-PAGE and Coomassie staining (Fig 1B). Western blot analysis using an anti-6xHis antibody confirmed that only the fragments of 50 and 40 kDa contained the C-terminal tag (Fig. 2A). As to recombinant AtDXS (predicted molecular weigth of 78 kDa including the 6xHis affinity tag) bands of 78, 48, 38 and 26 kDa were also detected after SDS-PAGE and Coomassie blue staining (Fig. 1C).

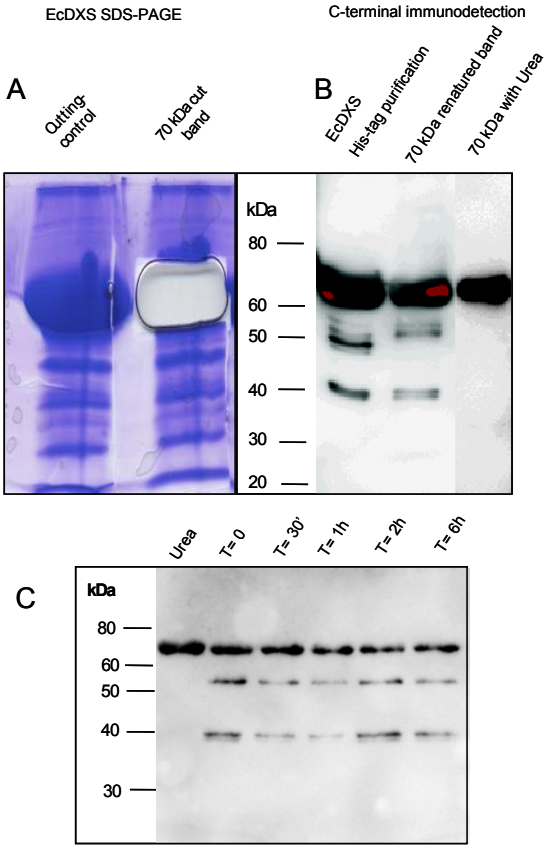


Fig.3 Excised full-length EcDXS undergoes proteolysis under renaturing conditions. A, SDS-PAGE of the electrophoresed EcDXS. After SDS-PAGE the cutting control (first lane of A) was stained and based on their migration the 70 kDa band was cut (second lane of A), after that, the lane was stained to check that no additional bands were additional bands were additional cut. B, His-tag at C-terminus immunoblotting of refolded EcDXS. Protein of the strip was electroeluted and refolded and the cleavage was assessed by western blot (second lane of B). The ~50 kDa cleavage fragment showed a slightly slow migration rate compared to the purified recombinant EcDXS (first lane of B). Protein electroelution and refold in the presence of a denaturing buffer (8M urea) strongly reduced the cleaved fragments (third lane of B). Molecular weights are indicated. C, refolding in urea buffer and time course of proteolysis after buffer exchange on refolding buffer at 0, 30 min, 1h, 2h and 6h.

To further confirm the autoproteolytic cleavage of EcDXS, TmDXS and AtDXS, the corresponding proteins were expressed in a cell-free eukaryotic system. Since genome of rabbits do not encode DXS, it is also very unlikely that the reticulocyte lysate system used in our experiments contains a specific protease system for DXS. Expression of EcDXS, TmDXS and AtDXS in the coupled rabbit reticulocyte lysate system yielded a pattern of radioactive bands similar to those previously observed in the preparation purified from *E. coli* (Fig. 4). In the case of *E. coli* and *T. maritima* the band of ~49 kDa was among the most prominent ones.

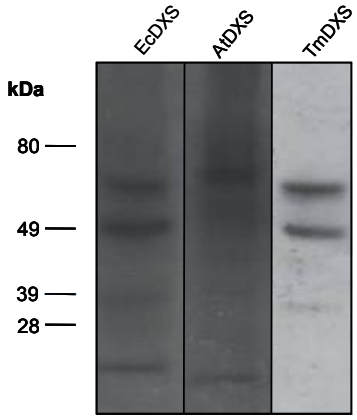


Fig.4 In vitro expression of DXS with TNT coupled retyculocyte lysate expression system. In vitro expression radiolabeled with L-[4,5-^{3}H] Leucine of EcDXS, AtDXS and TmDXS. Expression showed both the entire protein and a predominant cleaved fragment of 50kDa.

Analysis of DXS processing *in vivo*

To study whether the autoproteolysis of DXS detected *in vitro* also occurs *in vivo*, *E. coli* cells growing at mid exponential phase were collected by centrifugation, resuspended in electrophoresis loading buffer and heated at 100ºC for 5 min. Immunoblot analysis of the proteins separated by SDS-PAGE revealed a major band of ~69 kDa, corresponding to the intact DXS protein, and mainly, a band of about 40 kDa containing the 6xHis tag at the C-terminal. Furthermore, immunoblot analysis of the form containing the 6xHis tag at the N-terminal showed the intact DXS protein and a band of about 35 kDa that could be the corresponding part of the protein of the 40 kDa 6xHis C-terminal tag cleavage site. During the purification process, after the sonication step appearing two bands of about 50 kDa, all containg the 6xHis tag at the C-terminal (Fig. 2B). When similar experiments were done with *T. maritima* cell extracts, a band of 50 kDa and two bands of about 40 kDa were detected in addition to the intact DXS protein after the sonication step. In both cases, the obtained pattern of bands was equivalent to those shown in Fig. 2A. These results would suggest that the identified autoproteolytic processing of DXS could play a role in controlling DXS activity and/or stability *in vivo*.

Identification of cleavage sites

To characterize in more detail the autoprocessing of DXS, the N-terminal sequence of the degradation fragments was analyzed by Edman sequencing. For *E. coli* DXS, both the 40 and 50 kDa fragments were analyzed. The N-terminal end of 50 kDa fragment corresponds to Ser207 (SLREG) (Fig. 5) and the N-terminal end of 40 kDa fragment corresponds to Pro295 (PITFH) which has a peptide bond with an Asp residue (Fig. 5). For *T. maritima*, the N-terminal end of the 40 and 30 kDa fragments were established as residues Ser^{265} (SVVHV) and Met^{1} (MLLDE), respectively (Fig. 5). For *A. thaliana*, the N-terminal end of the 45 kDa fragment was identified as Ala^{302} (AAKV).

Hydroxylamine treatment of EcDXS

The enhancement of the rate of processing of the target protein by a robust nucleophile has been proposed as an additional proof of an autoproteolytic mechanism (Shao, Xu et al. 1996). Hydroxylamine is a nucleophilic compound very reactive with (thio)esters (Jencks, Cordes et al. 1960; Rosenblum and Blobel 1999) and succinimides (Kwong and Harris 1994), but too mild to break amide bonds. Hydroxylamine could increase the autoprocessing proteolysis through the hydrolysis of (thio)ester intermediates, which is considered the rate-limiting step in autoproteolysis (Shao, Xu et al. 1996).

Therefore, hydroxylamine was added to both non-denatured and urea-denatured EcDXS samples and aliquots taken after a short (15 min) and long (16 h) period treatments. Within 15 minutes, hydroxylamine treatment strongly enhanced the cleavage of the treated samples as compared to the control (hydroxylamine-untreated) samples (Fig. 6). After 16 h incubation, the treated samples only showed a slightly increased cleavage as compared to the 15 minutes treated samples, indicating that most of the cleavages occurred shortly after hydroxylamine treatment. Interestingly, the urea-denatured samples were cleaved at similar levels than the non-denatured samples. This fact strongly suggests the stable presence of succinimide residues in purified EcDXS. However, it should be stressed that whereas the ~40 kDa band produced after hydroxylamine treatment seems to results from a cleavage at the site identified previously, the band of ~50 kDa present in the control reaction was replaced by a new band of ~55 kDa in the treated samples (Fig. 5), indicating that another cleavage site is preferentially induced by hydroxylamine. Interestingly, the size of this band is similar to the band of ~55kDa already reported for the renatured enzyme (Fig. 3B). Edman sequencing of the 55 kDa hydroxylamine-cleaved fragment revealed Gly^{153} (GAITA) as the N-terminal residue (Fig. 5). This means that the cleavage is produced between Asp^{152}-Gly^{153} and corroborates the hypothesis of the presence of a succinimide or cyclic intermediates since this pair of residues have been described to undergo this intramolecular rearrangement with high probability. In this context, Edman sequencing of the ~40 kDa band produced by hydroxylamine treatment corroborated the cleavage between Asp^{294}-Pro^{295} (PITFH). In addition, the presence of other N-terminal fragments, like the Pro^{306}(PSSG) opened the possibility of the presence of other cyclic intermediates like between the residues Asp^{305}-Pro^{306}. The impossibility of the Asp-Pro peptide bond to form succinimide intermediates has been extensively described in the literature (Catak 2008, Li 2009, Oiyonen 1993). In spite of this, cyclic anhydride intermediates are formed through the attack of the carboxylic acid side chain of the Asp residue on the n+1 carbonyl center of the peptide bond (Li 2009). This intermediate has been described that undergoes hydrolysis to give the backbone cleavage more easily than succinimides or other Asp-aa peptide bonds (Li 2009). Taken

together, these results suggest that whereas the ~40 kDa fragment is induced by the hydrolysis of a cyclic anhydride intermediate at residue 294 both in control and in hydroxylamine treated samples, the hydroxylaminosis of succinimide of Asp^{152} could make an "epistatic" effect over the rearrangement site involved in the formation of the ~50 kDa fragment (cleavage between residues Ser^{206}-Ser^{207}) obtained from the native protein.

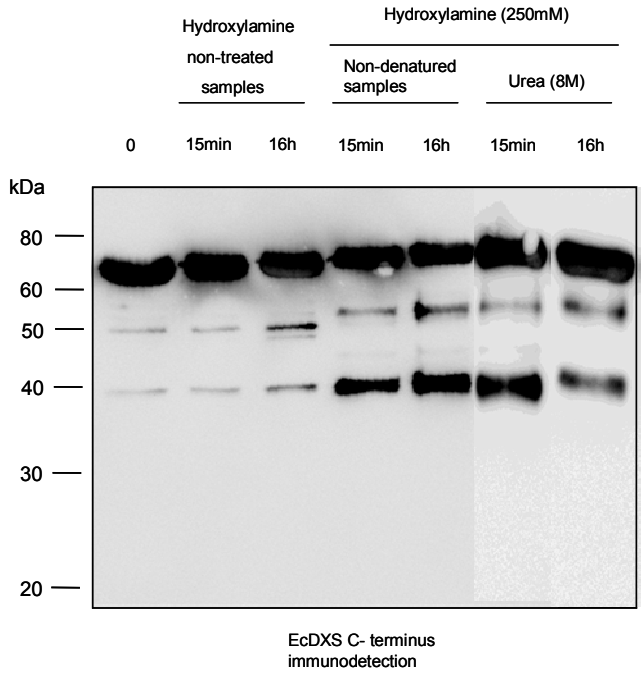


Fig.5 Hydroxylamine treatment of EcDXS (with His-tag in the C-terminus). EcDXS was incubated at room temperature and samples were removed at 15 min and 16h. Furthermore, EcDXS urea- and non urea-treated samples were incubated with 250mM hydroxylamine. Non denatured samples were removed at 15 min and 16h and denatured samples at 16h. Hydroxylamine treatment strongly enhances the cleavage of both denatured and non-denatured samples compared to non-treated samples. The hydroxylamine-cleaved 50 kDa peptides presented a slight slow migration in SDS-PAGE compared to the non hydroxylamine-treated 50 kDa peptides, in the same manner as observed in the novo fragment produced in the refold experiments. N-terminal sequencing of the 50 kDa peptides from non hydroxylamine-treated (SLREG) and hydroxylamine-treated samples (GAITA) were done. Molecular weights of the western blot analysis are marked on the left.

Activity of truncated forms of DXS lacking the major autoprocessing site in *E. coli*

As it has been shown above, cleavage between residues Ser^{206}-Ser^{207} represents one of the major autoprocessing sites of EcDXS. Interestingly, this cleavage site is located within a non-conserved region in bacterial and plant DXS that has been identified as an external and disordered region of DXS which extends from residues 183 to 238. Since this region could not be resolved in the crystallographic studies of bacterial DXS (Xiang, Usunow et al. 2007) its structural and/or functional role was not established. To check whether this region is relevant for enzyme activity a DXS derivative lacking the region extending from residues 191 to 239 was generated (pET23-EcDXS-Δ191-239). Since it can not be excluded that this region could have a structural role by acting as a hinge between functional domains a derivative in which the 191-239 region was replaced by a polyglycine loop was also prepared (pET23-EcDXS-Δ191-239[G-loop]). *EcAB4-2* cells transformed with either plasmid pET23-EcDXS-Δ191-239 or pET23-EcDXS-Δ191-239[G-loop] were able to complement the defective DXS strain (Fig. 7), thus

showing that the deleted region is not essential for enzyme activity. Interestingly, immunoblot analysis of purified recombinant EcDXS-Δ191-239 did not show degradation products (Fig. 2A), suggesting that this exposed domain of the protein may play a relevant role in targeting DXS autoprocessing, at least in *E. coli*.

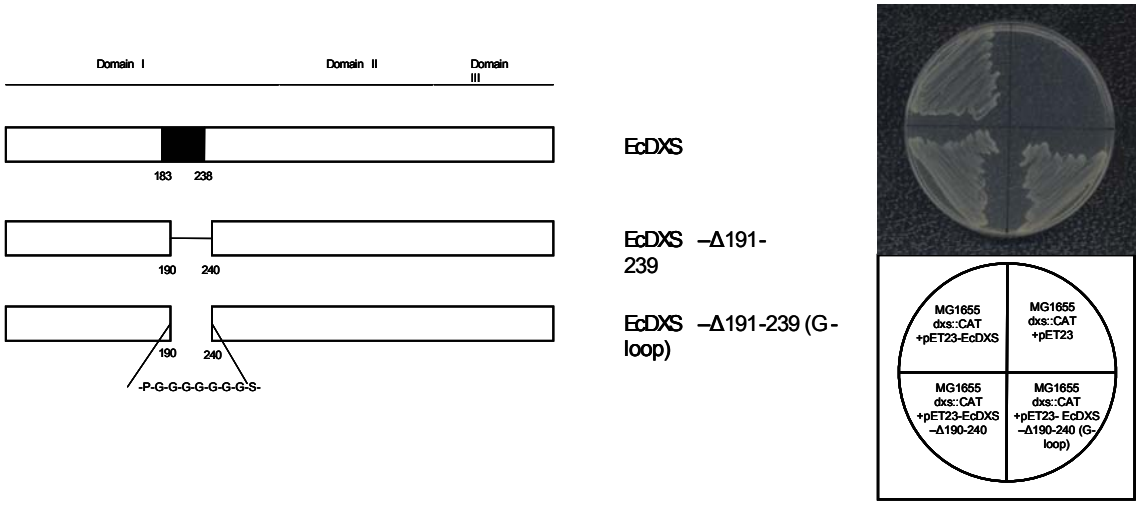


Fig.6 Activity of EcDXS constructs. (Left) Scheme of the constructs EcDXS-Δ190-240 and EcDXS-Δ190-240 (G-loop). Non-resolved region in the crystal structure of the EcDXS (Xiang 2007) is marked with black boxes. (Right) Complementation assays of the EcDXS constructs.

Heterologous protein autodegradation induced by *E. coli* DXS sequences

As described above, the cleavage site involving Ser^{207} seems to be a main target for the autoprecessing of EcDXS. However, cleavage at this site apparently required the proper folding of the protein (Fig. 3). To check whether cleavage at Ser^{207} requires the folding of the entire DXS protein or just a short domain around the cleavage site the region of EcDXS extending from residues 158 to 240 was cloned in the middle of a chimeric protein consisting of GST fused to GFP. A 6xHis-tag was also included in frame downstream of GFP (Fig. 8). The chimeric constructs (encoded in plasmids pET23-GST-GFP and pET23-GST-EcDXS158-240-GFP, respectively) were expressed in *E. coli* and the recombinant proteins purified by IMAC. Immunoblot analysis of the purified proteins revealed the presence of degradation fragments only in the case of the chimeric construct containing the internal EcDXS sequence. Assuming that the size of GFP is around 28 kDa, the results shown in Fig. 8 are in agreement with a major cleavage site within the EcDXS sequence. However, secondary cleavages within GST and GFP also seem to occur.

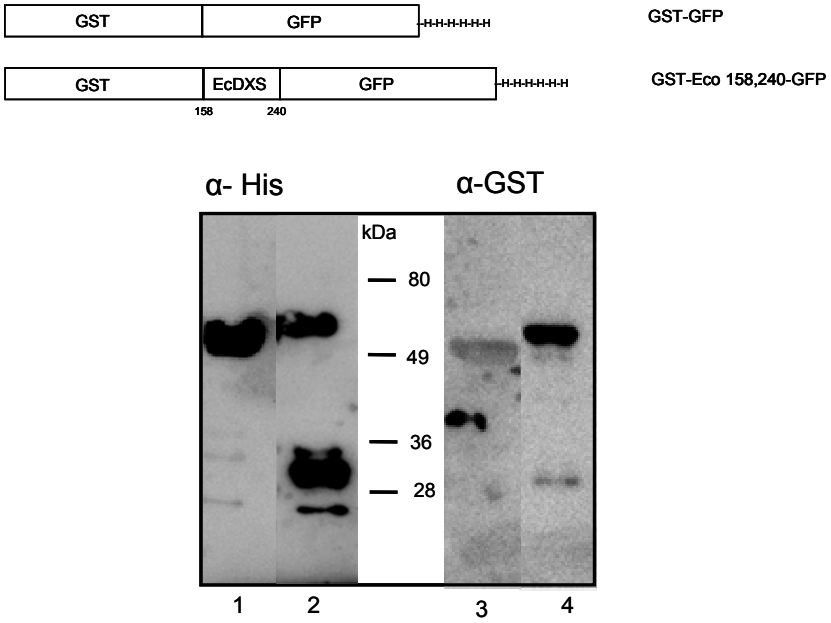


Fig.7 Chimerical constructs. (Up) Scheme of the constructs GST-GFP, GST-Eco158,240-GFP and GST-At226,318-GFP. (Down) Constructs were purified using the His-tag at C terminus and were analyzed by Western blot with anti-his and anti-GST antibodies. GST-Eco158,240-GFP (lanes 2 and 4) presented the cleavage of the protein compared to GST-GFP (lanes 1 and 3).

Discussion

Proteases play an important role in several processes in the life. However, the occurrence of autoproteolytically intramolecular processes has also been reported in a large number of protein families. Most of these intramolecular rearrangements involve the formation of an ester bond in exchange of the amide bond between two amino acid residues. These kinds of rearrangements have been described for instance in pyruvoil enzymes (Recsei and Snell 1984; van Poelje and Snell 1990), Ntn hydrolases (Brannigan, Dodson et al. 1995), nucleoporins (Rosenblum and Blobel 1999), proteins with SEA domain (Johansson, Macao et al. 2008) and HINT proteins (Hedgehog and inteins) (reviewed in (Paulus 2000)). Moreover the formation of succinimide residues, like in the cleavage of the interbranched peptide during protein splicing plays a relevant role in protein autoprocessing, and the formation of cyclic anhydride intermediates plays a relevant role in protein autoprocessing (Fig.8). The formation of succinimide residues also play a role in the deamination/formation of L-isoaspartate (isoAsp), residue that constitutes a major source of spontaneous protein damage under physiological conditions.(reviewed in (Reissner and Aswad 2003)).

Autoproteolysis is generally believed to proceed by a two-step reaction mechanism involving an ester intermediate (Perler 1998). This ester is generated via an intermediate structure (hydroxyoxazolidine or hydroxythiazolidine) derived from the nucleophilic attack of a serine/threonine hydroxyl (ROH) or a cysteine thiol (RSH) on the carbonyl carbon of the preceding amino acid. Such an N → O or N → S acyl shift is believed to be similar for the initial steps of autoproteolysis in all of the mentioned enzymes. What differs between some of the enzymes above is the subsequent resolution of the ester. In the pyruvoyl-dependent enzymes, the ester is eliminated into an N-terminal dehydroalanine that then hydrolyzes into a pyruvoyl group and ammonia (Recsei,

Huynh et al. 1983). In the other enzymes, the ester is instead broken by hydrolysis. In the Hedgehog proteins, the attacking nucleophile is a cholesterol hydroxyl, thus yielding covalently bound cholesterol (Porter, Young et al. 1996), and in the splicing reaction, it is a downstream serine hydroxyl (Xu and Perler 1996), whereas in the group of Ntn hydrolases, nucleoporins and SEA domain proteins the nucleophile is believed to be water, thus generating novel C and N termini (Brannigan, Dodson et al. 1995).

The fact that DXS undergoes proteolysis in *Escherichia coli* cells, in a cell-free system, in chimerical constructs and in a renatured form from a full-length non-proteolyzed EcDXS without any extraneous addition could suggests an autoproteolytic process as more an easy explanation. The data presented here has revealed that DXS is modified at post-translational level and these intramolecular rearrangements could produce an autoproteolytic process that cleaves the protein at specific amino acid residues. This autoproteolytic process, which was initially observed in the *E. coli* enzyme, has also been observed in the bacterium *T. maritima* and in the plant *A. thaliana*, thus suggesting that autoprocessing of DXS represents a novel and general feature of this enzyme. For *E. coli* DXS, the main cleavage sites has been identified between residues Ser^{206} and Ser^{207}. Similarly, one of the main cleavage sites in the *T. maritima* DXS has been found between residues Ser^{264} and Ser^{265}. The fact that these pair of Ser residues are located in different regions of the protein suggests that the position of this cleavage site is not a major trait in DXS autoprocessing. Mechanistically, the cleavage of DXS between the identified pairs of Ser residues in the *E. coli* and *T. maritima* enzymes can most likely involve an ester bond-related process equivalent to that described above for the autoprocessing of Nth hydrolases.

The second main cleavage site in the *E. coli* enzyme has been predicted between residues Asp^{294}-Pro^{295} which, in contrast to the pair of Ser residues indicated above, is highly conserved among DXSs(Fig.9C), thus suggesting that this cleavage site may be common to all known DXSs. The mechanism(s) involved in the cleavage between Asp^{294}-Pro^{295} in *E. coli* DXS could involve the hydrolysis of a cyclic anhydride intermediate in a process described for other protein models (Li 2009) and, in some ways, similar to that operating in protein splicing. It has been proposed that protein splicing begins with the attack of a downstream Asn residue at the upstream splice junction to form a succinimide. Although the C-terminal residue of inteins is typically an Asn, the presence of Gln and Asp has also been reported (Amitai, Dassa et al. 2004). The enhanced cleavage of the Asp^{294}-Pro^{295} peptide bond with hydroxylamine (Fig 5) is a clear indication that a cyclic anhydride intermediate is formed at the Asp^{294} residue. In fact, the hydroxylamine cleavage in this position is the same cleavage site observed in hydroxylamine-untreated samples and seems to be the main cleavage site observed in the *in vivo* experiments. It should be stressed that the Asp^{294} is located in the second non-resolved region of the crystal structure from the *E. coli* DXS.

The hydroxylamine treatment of E. coli DXS (Fig 5) has also shown that succinimide is formed at the Asp^{152} amino acid residue. The presence of the ester bond rearrangement between the Ser^{206} and Ser^{207} has not been corroborated in hydroxylamine assays most probably owed to the presence of this succinimide residue. Although this residue has not been directly involved in DXS cleavage, despite that the refolding of EcDXS showed a band of the corresponding weight, the possibility that this succinimide could be an intermediate in the conversion of Asp into isoAsp represents an attractive field of study considering that *E. coli* Asp^{152} is present in the conserved TPP binding motif ($G\underline{D}GX_{25\text{-}30}N$) found in all known DXSs (Fig. 9B). Furthermore, this residue plays an essential role in the stabilization of the TPP and Mg^{2+} ion (Xiang, Usunow et al. 2007) (Fig.10). The presence of stable succinimides observed in the hydroxylamine assays could be explained in the way of the intermediate of the IsoAsp formation. The modification of this Asp residue to isoAsp could certainly affect enzyme activity and thus

represent a novel regulatory mechanism to modulate DXS activity. Interestingly, this Asp residue is also conserved in other related TPP-dependent enzymes (like transketolase and pyruvate dehydrogenase) (Fig 9). In this proteins the TPP binding motif is present in a turn that separates a preceding β-strand from a following α-helical (Hawkins, Borges et al. 1989). This turn allows the orientation of the C2 of the TPP to the solvent (corroborating the mode of catalysis).

Although the autocatalytic processing of DXS represents a novel aspect on the characterization of this enzyme, more experiments are needed to really assess the biological significance of this post-translational modification in the regulation of isoprenoid biosynthesis.

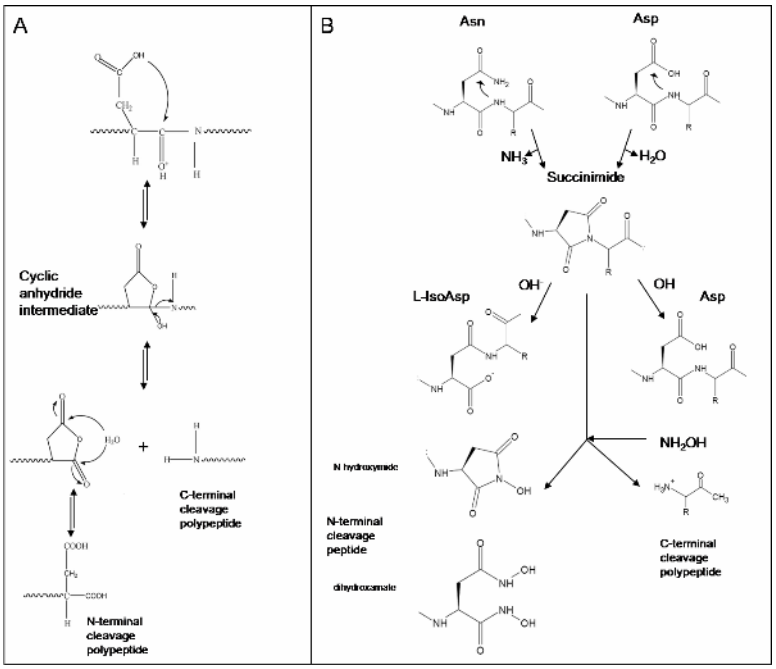


Fig.8. On the left, scheme of the cyclic anhydride intermediate formation. On the right, asparagine deamination/ L-isoaspartate formation through the succinimide via. The hydroxylaminosis are also shown.

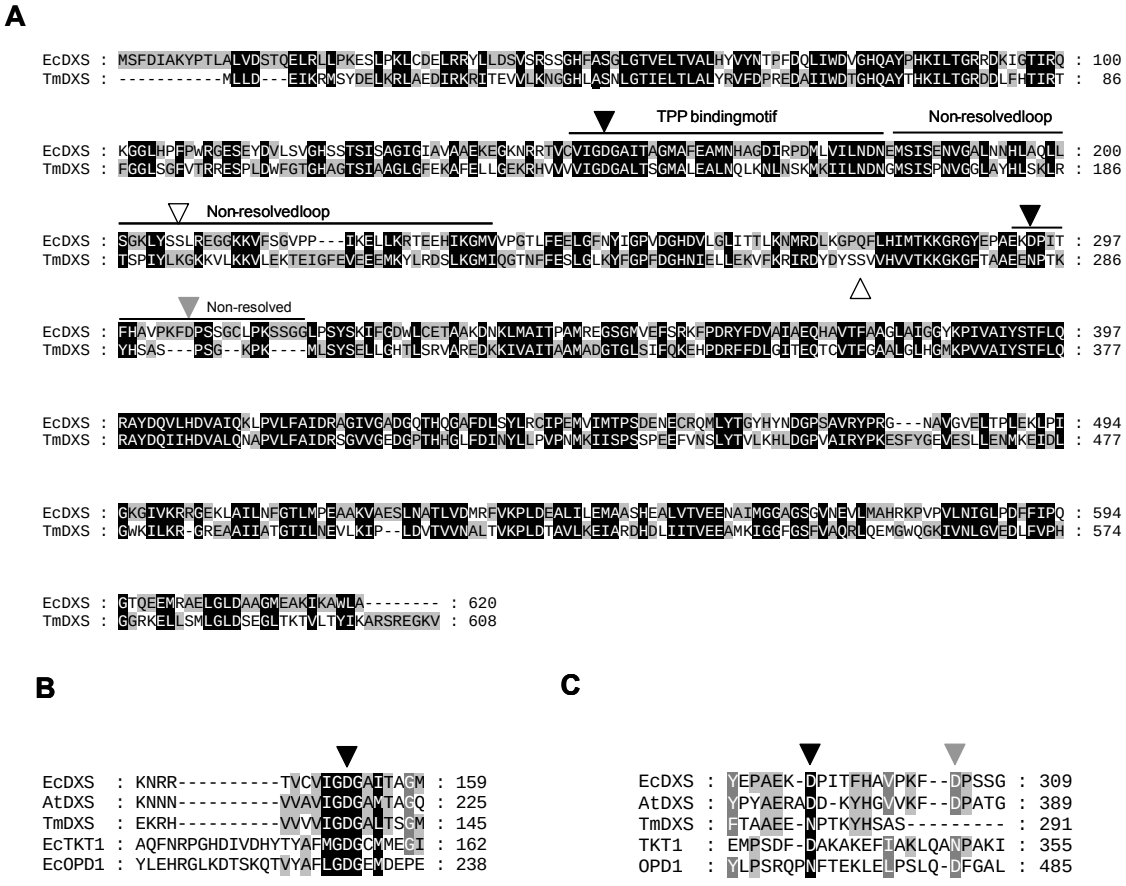


Fig.9. Alignment of DXS. A Alignment of EcDXS and TmDXS. The TPP-binding motif and the non-resolved region in crystal structure of EcDXS are marked. Succinimide/cyclic anhydride intermediates residues are marked with a black triangle, cleavage sites are marked with a white triangle and the second cleavage peptide obtained by hydroxylamine is marked with a grey triangle. B Alignment of the region surrounding the EcDXS Asp^{152}-succinimide rearrangement of EcDXS, TmDXS, AtDXS, *E.coli* transketolase 1 (TKT1) and E.coli E1 subunit of pyruvate dehydrogenase (OPD1). It is shown that this residue is totally conserved. C. Alignment of the region surrounding the EcDXS Asp^{294}-cyclic anhydride intermediate rearrangement of EcDXS, TmDXS, AtDXS, EcTKT1 and EcOPD1.

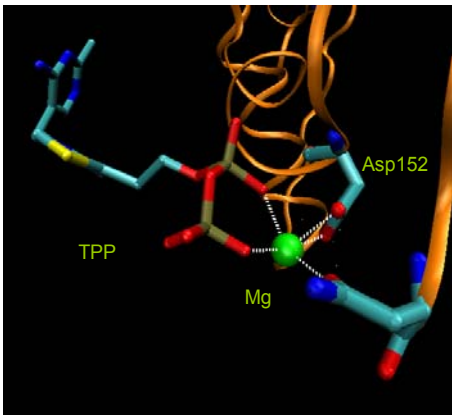


Fig.10. Scheme of the TPP-binding motif from EcDXS. The octahedral structure formed between the Mg^{2+} ion, the pyrophosphate moiety of the TPP and the residues of EcDXS are shown.

Conclusions

The 1-deoxy-D-xylulose 5-phosphate synthase (DXS), a crucial enzyme in some biosynthetic pathways, has been indirectly described as a target for proteases. In this paper, we show that the DXS proteolysis is observed in three homologues in "in vivo" *Escherichia coli* expression and after protein purification, and in a cell-free system. Furthermore, renaturated full length non-proteolyzed EcDXS presented the proteolytic event in comparison with non-renaturated one suggests a folding-dependent process. The presence of a Ser amino acid residue involving the cleaved peptide bond by N-terminal sequencing of the peptides allows to thinking in an ester-bond related process. Interestingly, EcDXS is cleaved by hydroxylamine treatment under proper folding and denatured conditions suggesting an Asx-related intramolecular rearrangement in two different sites, which was corroborated by N-terminal sequencing. Thus, hydroxylaminosis could not corroborate the presence of an ester-bond by the presence of the succinimides/cyclic anhydride intermediates suggesting the possibility of two processes. The presence of a succinimide in the Asp^{152} amino acid residue involved in the TPP binding domain could have implications in the enzymatic activity.

Name	(5' - 3')
158FDXSEco	GGCATGGCGTTTGAAGCGATGAATCAC
240RDXSEco	GCCAGGCACTACCATGCCTT
GSTNdeIF(2)	CGAGCATATGTCCCCTATACTAGGTTA
GFPSacIR(pET23b)	CCGGGAGCTCCACTTGTACAGCTCGTC
F214DXS	GATCGACATATGAAAGTTTTCTCTGGCGTGC
R214XhoI	ACTCGAGTTTCCCGCCTTCGCGCAG
T7promoter	TAATACGACTCACTATA
T7terminator	GCTAGTTATTATTGCTCAGCG
190RDXSEco(SmaI)	GGGGCCGACATTTTCGGAAATCGACATTTC
240FDXSEco(SmaI)	GGGACGTTGTTTGAAGAGCTGGGCTTTAAC
Flexlinker(P-G-S)D	CCCGGGGGCGGCGGCGGCGGCGGCTCT
Flexlinker(P-G-S)Com	AGAGCCGCCGCCGCCGCCGCCCCCGGG

Table 1 Primers used during the cloning procedures.

Bibliography

Amitai, G., B. Dassa, et al. (2004). "Protein splicing of inteins with atypical glutamine and aspartate C-terminal residues." J Biol Chem **279**(5): 3121-31.

Boucher, Y., H. Huber, et al. (2001). "Bacterial origin for the isoprenoid biosynthesis enzyme hmg-coa reductase of the archaeal orders thermoplasmatales and archaeoglobales." Mol Biol Evol **18**(7): 1378-88.

Brannigan, J. A., G. Dodson, et al. (1995). "A protein catalytic framework with an N-terminal nucleophile is capable of self-activation." Nature **378**(6555): 416-9.

Cairó, A. (2009). "Anàlisi molecular de la biosíntesis d´isoprenoides a Arabidopsis." phD Thesis.

Campos, N., M. Rodriguez-Concepcion, et al. (2001). "Escherichia coli engineered to synthesize isopentenyl diphosphate and dimethylallyl diphosphate from mevalonate: a novel system for the genetic analysis of the 2-C-methyl-d-erythritol 4-phosphate pathway for isoprenoid biosynthesis." Biochem J **353**(Pt 1): 59-67.

Chappell, J. (1995). "The Biochemistry and Molecular Biology of Isoprenoid Metabolism." Plant Physiol **107**(1): 1-6.

Eisenreich, W., M. Schwarz, et al. (1998). "The deoxyxylulose phosphate pathway of terpenoid biosynthesis in plants and microorganisms." Chem Biol **5**(9): R221-33.

Estevez, J. M., A. Cantero, et al. (2001). "1-Deoxy-D-xylulose-5-phosphate synthase, a limiting enzyme for plastidic isoprenoid biosynthesis in plants." J Biol Chem **276**(25): 22901-9.

Fitzpatrick, T. B., N. Amrhein, et al. (2007). "Two independent routes of de novo vitamin B6 biosynthesis: not that different after all." Biochem J **407**(1): 1-13.

Flores-Perez, U., S. Sauret-Gueto, et al. (2008). "A mutant impaired in the production of plastome-encoded proteins uncovers a mechanism for the homeostasis of isoprenoid biosynthetic enzymes in Arabidopsis plastids." Plant Cell **20**(5): 1303-15.

Frank, R. A., F. J. Leeper, et al. (2007). "Structure, mechanism and catalytic duality of thiamine-dependent enzymes." Cell Mol Life Sci **64**(7-8): 892-905.

Hawkins, C. F., A. Borges, et al. (1989). "A common structural motif in thiamin pyrophosphate-binding enzymes." FEBS Lett **255**(1): 77-82.

Jencks, W. P., S. Cordes, et al. (1960). "The free energy of thiol ester hydrolysis." J Biol Chem **235**: 3608-14.

Johansson, D. G., B. Macao, et al. (2008). "SEA domain autoproteolysis accelerated by conformational strain: mechanistic aspects." J Mol Biol **377**(4): 1130-43.

John, R. A. (1995). "Pyridoxal phosphate-dependent enzymes." Biochim Biophys Acta **1248**(2): 81-96.

Kwong, M. Y. and R. J. Harris (1994). "Identification of succinimide sites in proteins by N-terminal sequence analysis after alkaline hydroxylamine cleavage." Protein Sci **3**(1): 147-9.

Lange, B. M., M. R. Wildung, et al. (1998). "A family of transketolases that directs isoprenoid biosynthesis via a mevalonate-independent pathway." Proc Natl Acad Sci U S A **95**(5): 2100-4.

Lichtenthaler, H. K. (2000). "Non-mevalonate isoprenoid biosynthesis: enzymes, genes and inhibitors." Biochem Soc Trans **28**(6 Dec): 785-9.

Lichtenthaler, H. K., M. Rohmer, et al. (1997). "Two independent biochemical pathways for isopentenyl diphosphate and isoprenoid biosynthesis in higher plants." Physiologia plantarum **101**: 643-652.

Paulus, H. (2000). "Protein splicing and related forms of protein autoprocessing." Annu Rev Biochem **69**: 447-96.

Perler, F. B. (1998). "Breaking up is easy with esters." Nat Struct Biol **5**(4): 249-52.

Porter, J. A., K. E. Young, et al. (1996). "Cholesterol modification of hedgehog signaling proteins in animal development." Science **274**(5285): 255-9.

Recsei, P. A., Q. K. Huynh, et al. (1983). "Conversion of prohistidine decarboxylase to histidine decarboxylase: peptide chain cleavage by nonhydrolytic serinolysis." Proc Natl Acad Sci U S A **80**(4): 973-7.

Recsei, P. A. and E. E. Snell (1984). "Pyruvoyl enzymes." Annu Rev Biochem **53**: 357-87.

Reissner, K. J. and D. W. Aswad (2003). "Deamidation and isoaspartate formation in proteins: unwanted alterations or surreptitious signals?" Cell Mol Life Sci **60**(7): 1281-95.

Rodríguez-Concepción, M. and A. Boronat (2002). "Elucidation of the Methylerythritol Phosphate Pathway for Isoprenoid Biosynthesis in Bacteria and Plastids. A Metabolic Milestone Achieved through Genomics." Plant Physiol **130**(3): 1079-89.

Rodríguez-Concepción, M., O. Fores, et al. (2004). "Distinct Light-Mediated Pathways Regulate the Biosynthesis and Exchange of Isoprenoid Precursors during Arabidopsis Seedling Development." Plant Cell **16**(1): 144-56.

Rohmer, M. (1999). "The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in

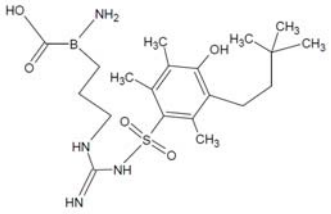
bacteria, algae and higher plants." Nat Prod Rep **16(5)**(Oct): 565-574.
Rohmer, M., C. Grosdemange-Billiard, et al. (2004). "Isoprenoid biosynthesis as a novel target for antibacterial and antiparasitic drugs." Curr Opin Investig Drugs **5**(2): 154-62.
Rosenblum, J. S. and G. Blobel (1999). "Autoproteolysis in nucleoporin biogenesis." Proc Natl Acad Sci U S A **96**(20): 11370-5.
Sacchettini, J. C. and C. D. Poulter (1997). "Creating isoprenoid diversity." Science **277**(5333): 1788-9.
Shao, Y., M. Q. Xu, et al. (1996). "Protein splicing: evidence for an N-O acyl rearrangement as the initial step in the splicing process." Biochemistry **35**(12): 3810-5.
Sprenger, G. A., U. Schorken, et al. (1997). "Identification of a thiamin-dependent synthase in Escherichia coli required for the formation of the 1-deoxy-D-xylulose 5-phosphate precursor to isoprenoids, thiamin, and pyridoxol." Proc Natl Acad Sci U S A **94**(24): 12857-62.
Sprenger, G. A., U. Schorken, et al. (1997). "Identification of a thiamin-dependent synthase in *Escherichia coli* required for the formation of the 1-deoxy-D-xylulose 5-phosphate precursor to isoprenoids, thiamin, and pyridoxol." Proc Natl Acad Sci USA **94(24)**(Nov 25): 12857-12862.
Spurgeon, S. L. and J. W. Porter (1981). Introduction. Biosynthesis of isoprenoid compounds. Wiley. Nueva York. **1:** 1-46.
van Poelje, P. D. and E. E. Snell (1990). "Pyruvoyl-dependent enzymes." Annu Rev Biochem **59**: 29-59.
Xiang, S., G. Usunow, et al. (2007). "Crystal structure of 1-deoxy-D-xylulose 5-phosphate synthase, a crucial enzyme for isoprenoids biosynthesis." J Biol Chem **282**(4): 2676-82.
Xu, M. Q. and F. B. Perler (1996). "The mechanism of protein splicing and its modulation by mutation." Embo J **15**(19): 5146-53.

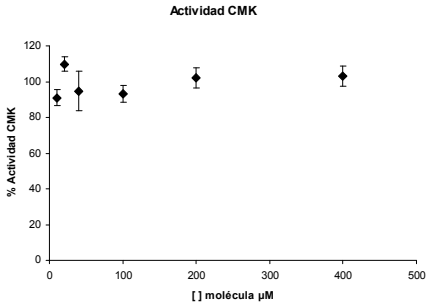
ANEXO 1

Moléculas y efecto sobre la actividad enzimática de CMK

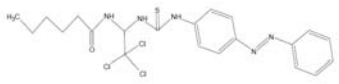
Hipótesis farmacofórica de 5 puntos

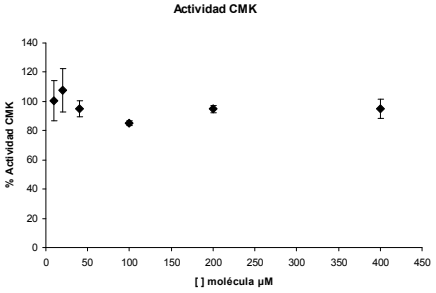
Base de datos/proveedor	MW	X-Score	RMSD	Referencia	Número CAS
ACD/Merck	440,557	5,09	1,1	MFCD00153419	112160-37-9



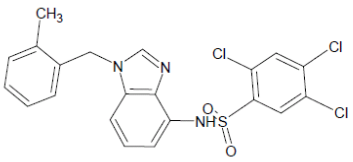


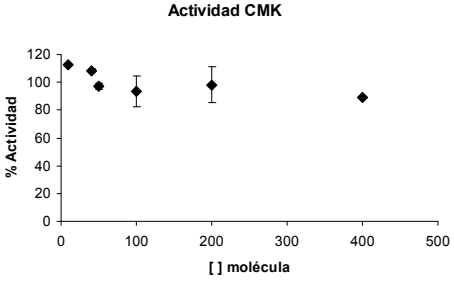
Base de datos/proveedor	MW	X-Score	RMSD	Referencia	Número CAS
ACD/Sigma Aldrich	500,88	5,2	1,64	MFCD00268048 (R286133 en sigma)	



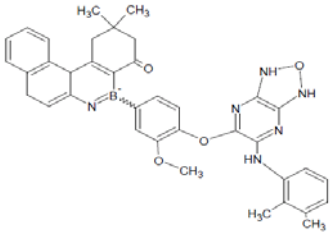


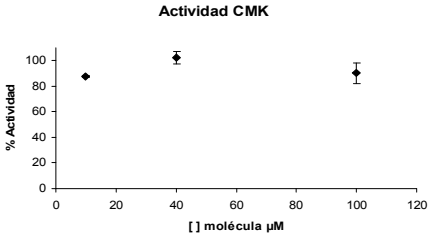
Base de datos/proveedor	MW	X-Score	RMSD	Referencia	Número CAS
ACD/BIONET	508,85	5,43	1,65	MFCD00231810 ('5J-338S en BIONET)	



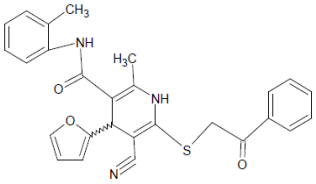


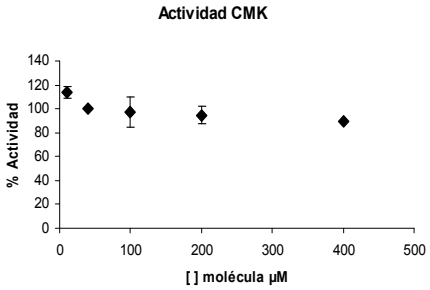
Base de datos/proveedor	MW	X-Score	RMSD	Referencia	Número CAS
SPECS	638,73	5,53	1,42	MFCD01359799 (AG-205/12084218 en SPECS)	



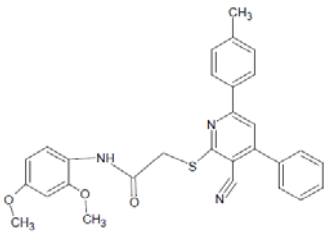


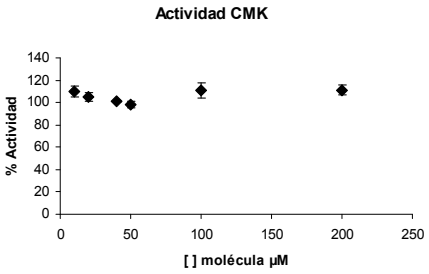
Base de datos/proveedor	MW	X-Score	RMSD	Referencia	Número CAS
SPECS	469,56	5,23	1,26	MFCD02234987 (AG-690/40721807)	



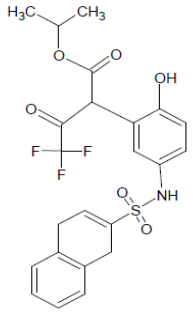


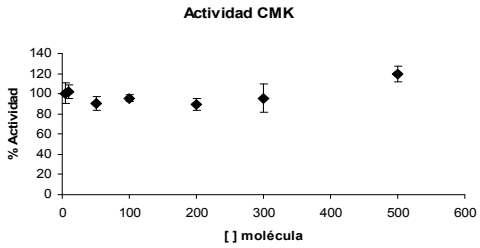
Base de datos/proveedor	MW	X-Score	RMSD	Referencia	Número CAS
SPECS	496	5,38	1,77	MFCD02739564(AG-690/40750925)	



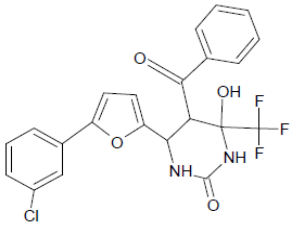


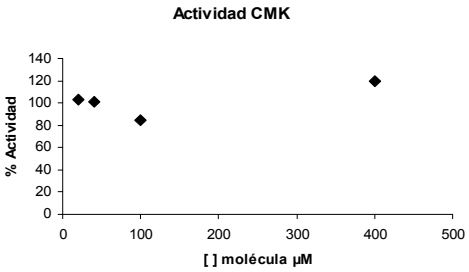
Base de datos/proveedor	MW	X-Score	RMSD	Referencia	Número CAS
ChemDiv	495,97	5,43	1,34	4456-0302	



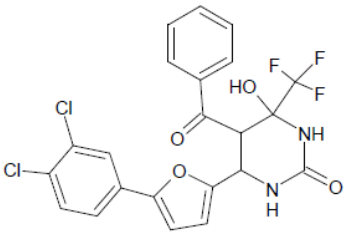


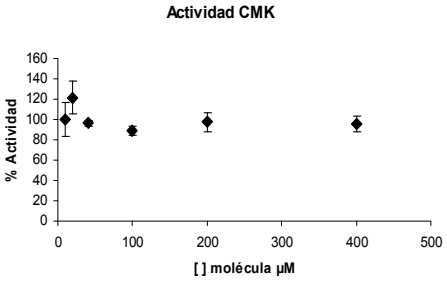
Base de datos/proveedor	MW	X-Score	RMSD	Referencia	Número CAS
ChemDiv	464,83	5,54	1,61	4618-0164	



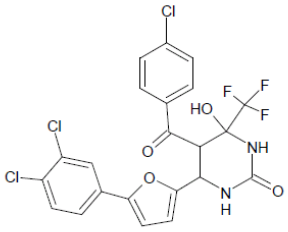


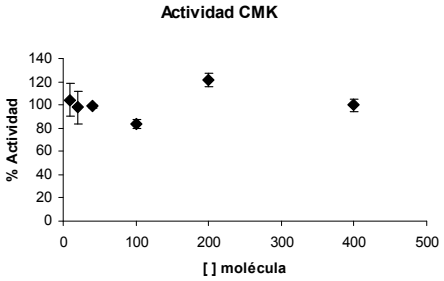
Base de datos/proveedor	MW	X-Score	RMSD	Referencia	Número CAS
ChemDiv	499,27	5,69	1,61	5310-4742	



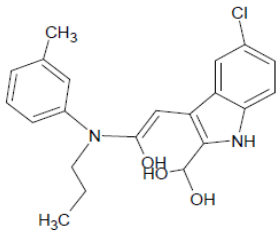


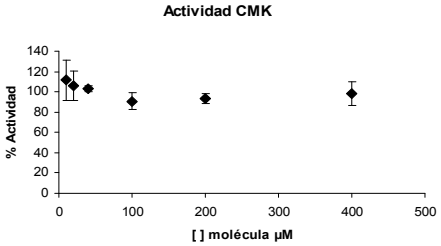
Base de datos/proveedor	MW	X-Score	RMSD	Referencia	Número CAS
ChemDiv	533,72	5,6	1,76	5310-4777	





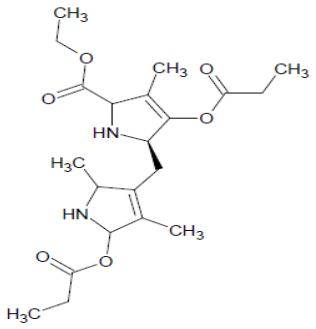
Base de datos/proveedor	MW	X-Score	RMSD	Referencia	Número CAS
ChemDiv	384,86	5,15	0,95	C289-0366	

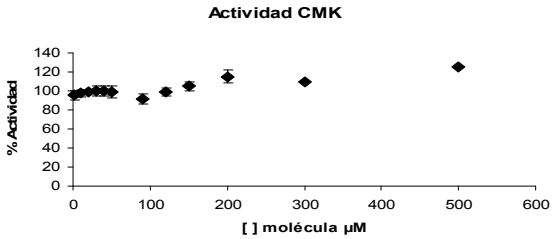




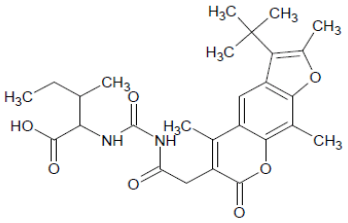
Hipótesis farmacofórica de 7 puntos

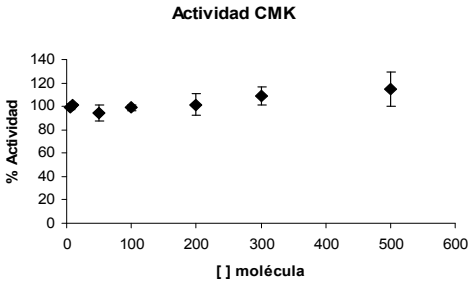
Base de datos/proveedor	MW	X-Score	RMSD	Referencia	Número CAS
NCI	404,3	4,64	1,33	NCI0031973	



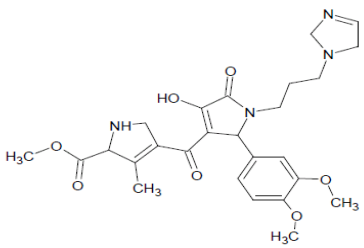


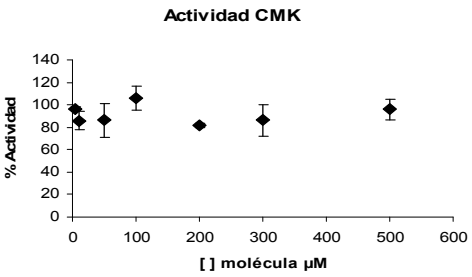
Base de datos/proveedor	MW	X-Score	RMSD	Referencia	Número CAS
Interbioscreen	498,3	5,14	1,68	COMPND STOCK1N-50533/ 859141-99-4	



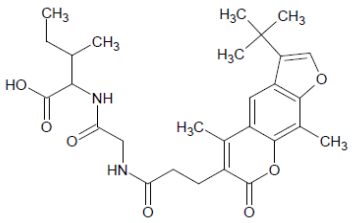


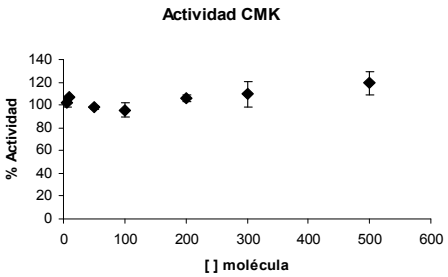
Base de datos/proveedor	MW	X-Score	RMSD	Referencia	Número CAS
Interbioscreen	522,3	4,96	1,47	COMPND STOCK4S-57225	





Base de datos/proveedor	MW	X-Score	RMSD	Referencia	Número CAS
Interbioscreen	512,3	5,24	1,68	COMPND STOCK1N-54193	





ANEXO 2

J Mol Model
DOI 10.1007/s00894-009-0458-5

ORIGINAL PAPER

Mimicking direct protein–protein and solvent-mediated interactions in the CDP-methylerythritol kinase homodimer: a pharmacophore-directed virtual screening approach

Victor Giménez-Oya · Óscar Villacañas · Xavier Fernàndez-Busquets · Jaime Rubio-Martinez · Santiago Imperial

Received: 15 September 2008 / Accepted: 2 December 2008

Abstract The 2C-methylerythritol 4-phosphate (MEP) pathway for the biosynthesis of isopentenyl pyrophosphate and its isomer dimethylallyl pyrophosphate, which are the precursors of isoprenoids, is present in plants, in the malaria parasite *Plasmodium falciparum* and in most eubacteria, including pathogenic agents. However, the MEP pathway is absent from fungi and animals, which have exclusively the mevalonic acid pathway. Given the characteristics of the MEP pathway, its enzymes represent potential targets for the generation of selective antibacterial, antimalarial and herbicidal molecules. We have focussed on the enzyme 4-(cytidine 5′-diphospho)-2-C-methyl-D-erythritol kinase (CMK), which catalyses the fourth reaction step of the MEP pathway. A molecular dynamics simulation was carried out on the CMK dimer complex, and protein–protein interactions analysed, considering also water-mediated interactions between monomers. In order to find small molecules that bind to CMK and disrupt dimer formation, interactions observed in the dynamics trajectory were used to model a pharmacophore used in database searches. Using an intensity-fading matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry approach, one compound was found to interact with CMK. The data presented here indicate that a virtual screening approach can be used to identify candidate molecules that disrupt the CMK–CMK complex. This strategy can contribute to speeding up the discovery of new antimalarial, antibacterial, and herbicidal compounds.

Keywords CMK-MEP · MVA · Solvent-mediated interactions · Protein–protein interactions · Molecular dynamics · Drug design · Intensity-fading MALDI-TOF mass spectrometry

V. Giménez-Oya and Ó. Villacañas contributed equally to this work.

V. Giménez-Oya · S. Imperial (✉)
Department de Bioquímica i Biología Molecular,
Universitat de Barcelona,
Avda Diagonal 645,
08028 Barcelona, Spain
e-mail: simperial@ub.edu

Ó. Villacañas · J. Rubio-Martinez
Department de Química Física, Institut de Recerca en Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona,
Martí i Franquès 1,
08028 Barcelona, Spain

X. Fernàndez-Busquets
Biomolecular Interactions Team, Nanobioengineering Group,
Institute for Bioengineering of Catalonia,
Barcelona Science Park, Baldiri Reixac 10–12,
08028 Barcelona, Spain

X. Fernàndez-Busquets · S. Imperial
Nanoscience and Nanotechnology Institute,
Universitat de Barcelona,
Barcelona, Spain

Present address:
Ó. Villacañas
Intelligent Pharma,
S.L. Barcelona Science Park, Baldiri Reixac 10–12,
08028 Barcelona, Spain

Introduction

Isopentenyl pyrophosphate (IPP) and its isomer dimethylallyl pyrophosphate (DMAPP) [1, 2] are the universal five-carbon precursors of one of the largest families of natural products—the isoprenoids. Isoprenoids include hopane triterpenes, ubiquinones and menaquinones in bacteria; carotenoids, plastoquinones, mono-, sesqui-, di and tri- terpenes and the

Published online: 07 February 2009

prenyl side chains of chlorophylls in plants; and quinones, dolichols, steroids and retinoids in mammals [3], and account for more than 30,000 naturally occurring molecules of both primary and secondary metabolism.

After the discovery of the mevalonic acid (MVA) pathway in yeast and animals, it was assumed that IPP was synthesised from acetyl-CoA via MVA and then isomerised to DMAPP in all organisms [2–5]. However, an alternative, MVA-independent, route for the biosynthesis of IPP was recently identified by labelling experiments in bacteria [6, 7] and plants [8]. It is named the MEP pathway, after what is currently considered its first committed precursor, 2C-methylerythritol 4-phosphate (for recent reviews see [9] and [10]).

Experimental evidence has shown that most organisms only use one of the two pathways for the biosynthesis of their isoprenoid precursors. Thus, the MEP pathway is essential in the malaria parasite *Plasmodium falciparum* and in most eubacteria, including the causal agents for diverse and serious human diseases like leprosy, bacterial meningitis, various gastrointestinal and sexually transmitted infections, tuberculosis, and certain types of pneumonia. The MEP pathway is absent in archaeobacteria, fungi and animals, which synthesise their isoprenoids exclusively through operation of the MVA pathway, whereas plants use both pathways for isoprenoid biosynthesis, although they are localised in different compartments [9–16].

Given the absence of the MEP pathway in mammals, its enzymes represent potential targets for the development of selective antibacterial, antimalarial and herbicidal compounds [10, 15–18]. An example of this new class of chemotherapeutic agents is fosmidomycin, a natural antibiotic originally isolated from *Streptomyces lavendulae*, which inhibits the second enzyme of the MEP pathway, deoxyxylulose 5-phosphate reductoisomerase [16]. Fosmidomycin and its derivatives have been shown to inhibit bacterial growth [19] and to be effective against *Plasmodium* infections [17, 20–22]. A combined therapy of fosmidomycin and clyndamicin, a chemical that inhibits the prokaryote-like translation machinery of the *P. falciparum* apicoplast, proved to be highly efficient against malaria, with a 3-day regime resulting in an overall cure rate of 95% [23]. These results are very promising and validate the usefulness of MEP pathway enzymes as drug targets.

In this work we have studied the enzyme 4-(cytidine 5′-diphospho)-2-C-methyl-D-erythritol kinase (CMK), which catalyses the fourth step in the MEP pathway, i.e. the ATP-dependent phosphorylation of the intermediate CDP-ME at the C-2 hydroxyl group to yield CDP-MEP. The enzyme from *Escherichia coli* has been cloned, overexpressed, purified, and crystallised. X-ray structure studies showed that CMK is organised as a homodimer. The asymmetric unit consists of two subunits, A and B, of approximate total mass 62 kDa, which assemble with C2 symmetry to form an extended homodimer, but only 4% of the total surface area of the two monomers is involved in dimer formation. Despite such a small area of interaction, a dimer is observed in gel filtration experiments that can be detected in matrix assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS) [24] and in analytical ultracentrifugation assays [25].

Disruption of protein–protein interactions through small molecules that target hotspots in the contact regions between protein surfaces [26] is an attractive therapeutic approach that can be very useful for drug design [27–30]. Thus, destabilising the CMK complex could represent a new strategy for the development of antibacterial and antimalarial agents. Although few direct protein–protein interactions are observed in the crystal structure of CMK, several water-mediated interactions seem be important for dimer stabilisation. Solvent-mediated protein–protein interactions play an important role in molecular recognition [31–35], and are being increasingly considered when modelling protein complexes [36]. Here, we have carried out molecular dynamics (MD) simulations on the CMK dimer in order to gain insight into such protein–protein interactions. With the objective of studying the intermolecular interactions of the complex in more detail, both direct and solvent-mediated protein–protein interactions were monitored throughout the dynamics trajectory. A pharmacophore was designed based on the position of atoms involved in the interactions, and used in a database search for small molecules capable of disrupting the CMK dimer. The association between candidate compounds and CMK was empirically checked by intensity-fading MALDI-TOF MS [37, 38]. This approach provides a quick method for the detection of complex formation between the protein and a ligand based on the reduction of ion intensities of the mixture, compared to the intensity of both protein and ligand alone. One hit provided by the pharmacophore-guided database search turned out to decrease dimer formation.

Materials and methods

System design

All minimisations and MD simulations were performed with the AMBER 7 suite of programs [39] and all-atom *parm99* and *gaff* force fields [40, 41]. For organic small molecules, atom types were assigned with the antechamber module of AMBER7, and atomic charges were calculated following the BCC-AM1 method [42]. Starting coordinates of the CMK complex were downloaded from the Protein Data Bank (PDB) [24], identification number PDBid=1oj4. The complex consists of two CMK proteins (each monomer comprising the full sequence, from residue 1 to residue 283), two

molecules of adenosine 5′-(β, γ-imido)triphosphate (AMP-PNP), two molecules of 4-(cytidine 5′-diphospho)-2-C-methyl-D-erythritol (CDP-ME), 192 water molecules and 1 chloride ion. The structure was resolved by X-ray diffraction at a resolution of 2.01 Å. Some side-chains presented two possible conformations. Because they were not involved in protein–protein close contacts, only the conformations marked in the PDB file as 'A' were considered further. Selenomethionine residues were replaced by methionines. Hydrogen atoms were placed where necessary with the LEaP module of AMBER7. The system was neutralised with Na^+ ions (31 ion atoms were required) placed at the most negative electrostatic potential points following a grid-based box-shaped procedure for mapping electrostatic potential surface (as implemented in LEaP).

Molecular dynamics simulations

A first minimisation step was carried out in vacuo with a distance-dependent dielectric constant of 4r. Then, a rectangular box of TIP3P water [43] was added, with a minimum distance between the protein and the edges of the box along the Cartesian axis of 10 Å. Water molecules closer than 1.8 Å to protein/CDP-ME/AMP-PNP were removed. The system was then relaxed by a two-step procedure. First, water molecule positions were optimised and, second, the whole system was optimised. MD simulations were performed using the particle mesh Ewald (PME) summation method [44] to compute long-range electrostatic energy. Bonds involving hydrogen atoms were constrained with the SHAKE algorithm [45]. The time step was set to 1 fs and the non-bonded pair list was updated every 25 ps, applying a cutoff value of 9 Å. Temperature and pressure regulation was simulated using Berendsen's algorithm [46], with time coupling constants of 1.0 and 0.2 ps, respectively, unless specified otherwise. The system was heated gradually to 300°K increasing 30°K each 10 ps, at constant volume. Then, a two-step pressure equilibration process followed. A temperature coupling constant of 1.0 ps was applied during the first 20 ps and increased to 2.0 ps for a further 40 ps. Up to this point, ions and protein atoms had been restrained under harmonic potentials with a force constant values of 1 and 5 kcal mol^1 $Å^{-1}$, respectively. Next, restraints were reduced gradually for 60 ps. Finally, simulations were carried out in the NVT ensemble with no restraints. The last 1,000 ps with stable total energy were considered as the production run for further analysis.

Energy analysis

Lennard-Jones (LJ) and Coulombic energies between every side chain of the first monomer protein and the entire second monomer protein, and vice versa, were computed for snapshots every picosecond throughout the 1,000 ps of production time with the ANAL module of AMBER 7. These calculations were also performed after adding water molecules to one of the interacting partners.

Solvent-mediated interaction analysis

Water molecules that established hydrogen bonds simultaneously with both proteins throughout the last 200 ps were analysed with a program developed by our group. Hydrogen bond distances were monitored and occupancy was defined as the percentage of time for which there existed at least one hydrogen bond to each monomer with a distance value lower than 2.5 Å.

Design of interaction models from MD trajectory

During the last 200 ps of production time, the positions of those atoms in one monomer that effectively interacted with its partner monomer were monitored. If several atoms contributed to the same effective interaction, the centre of masses of the mean positions was calculated and treated as a single interaction point. Maximum and minimum distances between every pair of points were used to construct the model for database searches, and mean positions of points as Cartesian coordinates were used for the docking procedure. Because all points must be referred to the same reference system, each snapshot was superimposed onto the backbone heavy atoms. An interaction feature was assigned to every point: hydrogen donor/acceptor, hydrophobic.

Database search and docking procedure

A selected interaction model was introduced as input for the compound search. Catalyst (Accelrys, San Diego, CA) software together with the Available Chemical Database (ACD), National Cancer Institute (NCI) database, ChemDiv database and Specs database were used. Matches were filtered and molecules with unusual bonds or with a molecular mass higher than 750 Da were not selected. Hits were saved as *mol2* format files containing all the conformations given by Catalyst. Each conformation was docked to the CMK monomer as follows. Pharmacophoric features were automatically assigned to the ligands, and distances between pharmacophoric points were calculated for all conformations. Next, for each conformation, we tested whether a combination that satisfied the interaction model existed and, in such cases, the ligand pharmacophoric points were superimposed on those of the interaction model, optimising the root mean square deviation (RMSD). At this point, the CMK monomer structure was introduced into the model to detect van der Waals (vdW) clashes between the ligand and the protein. In order to avoid bad

vdW contacts, a slight translational movement of the ligand was allowed. Atoms were not allowed to be closer than half of the sum of their vdW radii. If there were no forbidden vdW contacts, the dissociation constant was evaluated with X-Score [40].

A visual structure analysis was carried out on ligands with best docking, defined as those with best X-Score and RMSD consensus values. The compounds selected were finally classified according to their structure.

Preparation of recombinant CMK

Recombinant CMK from *E. coli* was prepared as previously described [47]. Briefly, M15[pREP4] *E. coli* cells were transformed with the pQE30-*ispE* construct containing the coding region of the *ispE* gene of *E. coli*. Transformed cells were grown at 37°C in LB medium supplemented with ampicillin (100 μg/mL) and kanamycin (50 μg/mL) until an OD_{600} of 0.3–0.4 was reached. Induction was then performed with 0.4 mM IPTG for 3 h at 30°C. Cells were harvested by centrifugation (3,000 *g*, 10 min) and resuspended in 40 mM Tris–HCl buffer, pH 8, containing 0.5 mM EDTA, lysozyme (1 mg/mL) and complete-mini EDTA-free (Roche) (one tablet/10 mL buffer). Cells were incubated at 4°C for 20 min and sonicated (5×30 s bursts with 1 min cooling at 0°C). The extract was centrifuged at 11,000 *g* for 30 min at 4°C; the supernatant was recovered and supplemented with protamine sulphate (1.25 mg/mL), incubated at room temperature for 15 min, and then centrifuged at 13,000 *g* for 20 min at 4°C.

Hi-trap chelating columns (GE Healthcare, Little Chalfont, UK) were used for the purification of recombinant CMK. The resin was washed with 40 mM Tris–HCl buffer, pH 8.0, containing 0.1M NaCl, and 10 mM imidazole. His6-CMK was eluted with an imidazole gradient (10–500 mM) in 40 mM Tris–HCl buffer, pH 8.0, containing 0.1M NaCl. The fractions containing His6-CMK were 95% pure as judged from SDS–polyacrylamide gel electrophoresis and Coomassie brilliant blue R-250 staining and appeared as a major band corresponding to an apparent molecular mass of 31 kDa. CMK preparations were desalted by gel filtration on disposable Sephadex G-25 columns (PD10) (GE Healthcare) equilibrated in 40 mM Tris–HCl buffer pH 8.0 containing 0.1M NaCl, distributed into aliquots, snap-frozen with liquid N_2 and stored at −80°C until use. Protein concentration was measured by the method of Bradford [48] using bovine serum albumin as a standard.

Intensity-fading MALDI-TOF MS

Intensity-fading MALDI-TOF MS was used to study the formation of a complex between one of the selected compounds in the database search, 2,4,6-trichloro-N-[5,6-dimethyl-1-(2-methylbenzyl)-1H-1,3-benzimidazol-4-yl] benzenesulfonamide (Bionet Research, Key Organics, Camelford, UK), Compound 1 (C1), and purified CMK from *E. coli*. In these experiments, a direct protocol could be used because both C1 and CMK are soluble in the concentration range used. This protocol also requires that the complex remains intact in solution at least in part during the process of MALDI-TOF MS determination [37]. Samples were prepared as follows: a 1 mM C1 solution prepared in water was mixed (1:1, v:v) with different dilutions in the purification buffer of recombinant CMK ranging from 0.002 mM to 0.2 mM. Then, the samples were mixed (1:1, v:v) with a 10 mg/mL solution of α-ciano-4-hydroxycinnamic acid (CHCA) (Aldrich; Sigma-Aldrich, St. Louis, MO) matrix solution in acetonitrile/deionised water (1:1, v:v). One microlitre of the mixture was overlaid onto the MALDI-TOF plate and dried using the dried droplet method [49]. MALDI-TOF mass spectra were recorded in a 4700 Proteomics Analyzer instrument (Applied Biosystems, Foster City, CA). Acquisition of mass spectra was performed in the MS reflector positive-ion mode. Typical parameters were set to source and grid voltages 20 and 14 kV, respectively, power laser from 5,200 to 5,800, signal/noise threshold 5, and a noise window width of 50.

Drawings

Figures depicting complex structures were constructed with Chimera [50] or VMD [51].

Results and discussion

Interactions deduced from X-ray crystal structure

The X-ray crystal structure of CMK [24] shows a homodimeric complex, where the two monomers are apparently positioned in an inverted position. The protein monomers interact with one another through two small interfaces: A and B (see Fig. 1), in which similar interactions are established. Residues of monomer 1 (chain A in PDB original file) found in interface A are the same as those of monomer 2 (chain B in PDB original file) in interface B, and vice versa. In the experimental structure of the complex, three direct protein–protein electrostatic interactions are detected. Two of them are observed in interface A, an intermolecular salt bridge between Arg^{21} and Asp^{80} (hereafter, monomer 1 and monomer 2, respectively) and a hydrogen bond between Ala^{22}(carbonyl) and Gly^{87}(amine). While the former is also observed in interface B (the third protein–protein electrostatic interaction), the latter interaction is detected in interface B not as a direct protein–protein interaction but as a solvent-mediated interaction (Table 1),

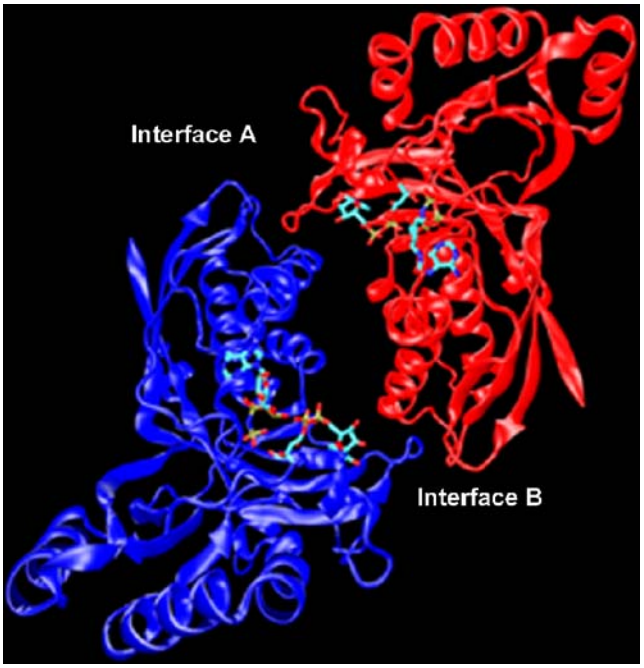


Fig. 1 X-ray structure of *Escherichia coli* 4-(cytidine 5′-diphospho)-2-C-methyl-D-erythritol kinase (CMK) dimer. Monomers are represented in cartoon mode; and adenosine 5′-(β, γ-imido) triphosphate (AMP-PNP) and 4-(cytidine 5′-diphospho)-2-C-methyl-D-erythritol (CDP-ME) molecules are in licorice mode as implemented in VMD visualisation software [51]. According to the reference used in the text, in the protein data bank (PDB) the first monomer is coloured blue and the second red

although with very distorted hydrogen bond angles between water molecule and protein atoms. A more relaxed geometry is found in the solvent-mediated interaction between Ser^{88}(carbonyl) and Ala^{22}(carbonyl) (Table 1). Table 1 lists all solvent-mediated hydrogen bonding networks between monomers.

Interactions from dynamics trajectory

Hydrogen bonds and salt bridges

The scenario obtained throughout the MD production time was slightly different from that suggested by the crystal structure. Our analysis revealed only one direct monomer–monomer electrostatic interaction, a salt bridge between Arg^{21} and Asp^{80} in interaction interface A [mean minimum hydrogen bond distance was 1.81 Å, with a root mean square (RMS) of 0.12 Å]. Although the interaction in the other contact region (interaction interface B) was not of the protein–protein type, a more detailed study revealed that a water bridge joined both side chains through hydrogen bonds (Fig. 2). The hydrogen bond distances were analysed at every snapshot extracted from the last 200 ps of the production time, and the respective percentages of time with values below 2.5 Å (occupancy) are noted in Table 2. The analysis showed that a water bridge (always the same water molecule) between Asp^{80} and Arg^{21} established hydrogen bonds during 96% and 99% of the time, respectively. As far as the interaction between Ala^{22} and Gly^{87} backbones is concerned (detected in the crystal structure), it was observed as water-mediated in both interfaces. In interface A, a water molecule acted as the centre of a hydrogen bond network and established favourable interactions with Ala^{22}(carbonyl) of monomer 1 and the backbones of Gly^{49}, Pro^{85}, Gly^{87}, and Ser^{88} of monomer 2. The occupancies of hydrogen bond distances were 83% and 97%, respectively (Table 2). In this latter case, the water molecule involved was labile and, therefore, different molecules were observed to play the role of mediating the interaction. However, the equivalent interaction in the other interface (B) was mediated by a water bridge, with respective hydrogen bond occupancy values of 65% and 99% (Table 2).

Table 1 Distances and angles of hydrogen bonds detected in the crystal structure of 4-(cytidine 5′-diphospho)-2-C-methyl-D-erythritol kinase (CMK). Distance 1 and angle 1 are calculated between water and residues of monomer 1, Distance 2 and angle 2 are calculated between water and residues of monomer 2. When a protein amino group is involved, both carbon-bonded atoms were used to calculate angles (carbon of the N-terminal side/carbon of C-terminal side)

Residues of monomer 1	Residues of monomer 2	Distance 1 (Å)	Angle 1 (degrees)	Distance 2 (Å)	Angle 2 (degrees)	Interface
Ala^{22} (amine)	Thr^{86} (side chain)	3.3	83/87	3.4	117	A
Ala^{22} (carbonyl)	Ser^{88} (carbonyl)	3.2	124	2.7	138	A
Gly^{49} (carbonyl)	Ala^{22} (carbonyl)	2.8	121	3.1	131	B
Lys^{76} (side chain)	Leu^{136} (carbonyl)	2.9	103	3.0	157	B
Asp^{80} (side chain)	Glu^{158} (side chain)	3.5	117	3.4	116	B
Ser^{88} (amine)	Ala^{22} (carbonyl)	3.2	93/102	2.1	141	B
Ser^{88} (carbonyl)	Ala^{22} (carbonyl)	2.8	124	2.1	141	B
Gly^{87} (amine)	Ala^{22} (carbonyl)	3.6	131/88	2.1	141	B
Gly^{87} (amine)	Ala^{22} (carbonyl)	2.8	118/119	3.7	88	B

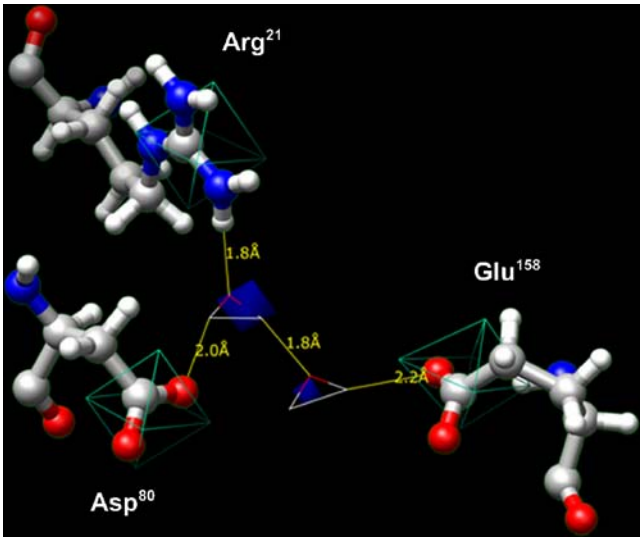


Fig. 2 Water-mediated interactions between Asp^{80} of monomer 1 and Arg^{21} and Glu^{158} of monomer 2 in interface B. Occupational analysis was carried out with Chimera software [50]. Grid volumes in which an oxygen atom of a water molecule is found at least 70% (approximately) of the time throughout the last 200 ps of the production time are coloured *solid blue*. Carbon atoms of carboxylic (Asp^{80} and Glu^{158}) and guanidinium (Arg^{21}) groups, and the oxygen atom of the water bridge between Asp^{80} and Arg^{21}, were used to superimpose snapshots before analysis. Cutoff and grid spacing values were set to 7 and 1.8 Å, respectively. For validation, volumes comprising the carbon atoms mentioned above 100% of the same time are outlined in *light blue*. As an example, the first snapshot is represented in the figure

Summarising, the intermolecular protein–protein hydrogen bonds elicited from the crystal structure were also observed in the dynamics trajectory but in a different structural pattern. Most direct protein–protein interactions in the crystal were observed in the dynamics trajectory as mediated by the solvent through hydrogen bond networks. The role of water molecules in mediating interactions between both monomers is also observed in the crystal structure. The CMK homodimer complex appears to overcome its low protein–protein interaction surface by water-mediated interactions. Further interactions of this kind were resolved from the trajectory and will be detailed in the following section.

Protein–protein solvent-mediated interactions

In addition to those described above, other stable water-mediated protein–protein interactions were found to be stable in the last 200 ps of production time (Table 2). In interaction interface A, three further water-mediated interactions were elucidated, two of which involving both the side chain and the carbonyl group of Asp^{23} in monomer 1. The side chain established a solvent-mediated interaction with the Lys^{76}(side chain) of monomer 2, whereas the carbonyl group did the same with Gly^{49}(carbonyl) and Arg^{72}(side chain) of monomer 2. Finally, the side chain of Glu^{158} in monomer 1 interacted with Lys^{76}(side chain) and Asp^{80}(side chain/carbonyl) of monomer 2 through hydrogen bonds established with water molecules. In interface B, one additional solvent-mediated interaction was observed between Ala^{79}(carbonyl) and Ala^{22}(amine). Similarities and differences between the crystal structure and the dynamics trajectory can be deduced by comparing values from Tables 1 and 2. The addition of temperature to the system changed the water-mediated monomer–monomer interaction pattern.

The interaction involving Glu^{158} and Asp^{80}(side chain)/Lys^{76}(side chain) in interface A was not detected in interface B. However, some evidence led us to carry out a spatial occupancy analysis on the oxygen atoms of the water molecules (Chimera software [50]). The results revealed

Table 2 Percentage of hydrogen bond (HB) distances under 2.5 Å in protein–protein solvent-mediated interactions throughout the last 200 ps of the production time

Residues of monomer 1	Residues of monomer 2	HB 1 (%)[a]	HB 2 (%)[b]	Interface
Ala^{22}(amine)	Ala^{79}(carbonyl)	19	94	A
Ala^{22}(carbonyl)	Pro^{85}(carbonyl), Gly^{87}(amine), Ser^{88}(carbonyl/amine) and Gly^{49}(carbonyl)	83	97	A
Asp^{23}(side chain)	Lys^{76}(side chain)	98	97	A
Asp^{23}(carbonyl)	Gly^{49}(carbonyl), Arg^{72}(side chain)	94	99	A
Glu^{158}(side chain)	Lys^{76}(side chain), Asp^{80}(carbonyl/side chain)	99	99	A
Pro^{85}(carbonyl), Thr^{86}(carbonyl), Ser^{88}(amine)	Ala^{22}(carbonyl)	65	99	B
Asp^{80}(side chain)	Arg^{21}(side chain)	96	99	B
Ala^{79}(carbonyl)	Ala^{22}(amine)	62	97	B

[a] Hydrogen bonds between water and monomer 1
[b] Hydrogen bonds between water and monomer 2

that this interaction was described as mediated by two water molecules (Fig. 2), which most of the time filled optimal sites for establishing a hydrogen bond network between the residues. There was also evidence of a further example of a double water-mediated interaction pattern between Ala^{22}(amine) and Ala^{79}(carbonyl) in interface A. However, occupational analysis did not show a clear picture of two oxygen atoms located in optimal positions (data not shown). In this last case, a single water-mediated interaction pattern was observed 19% of the time. It is likely that both patterns exchanged with one another several times throughout the trajectory.

On the whole, the results extracted from dynamics analysis show the importance of water-mediated intermolecular interactions in the CMK dimer complex. It has been argued that bound water should be considered as an integral part of the tertiary structure of proteins [52], a proof of which is the conservation of buried water molecules in homologous proteins [53]. Here, in the CMK complex, most water molecules that were involved in intermolecular solvent-mediated interactions were highly labile. Because of the small monomer interaction surfaces, these water molecules are highly exposed, that is, they are not buried in protein cavities, and have high exchange rates. The existence of the hydrogen bond network was detected for most of the 200 ps time, suggesting that the role of water molecules is crucial for binding and recognition, although they are not tightly bound to the proteins. The computation of the protein–protein interaction surface throughout the 20 snapshots extracted from the production time (carried out with MolSurf as implemented in AMBER7), gave a value of 355 $Å^2$. When considering water molecules involved in solvent-mediated interactions as an integral part of monomers, the surface increased to 422 $Å^2$, a 19% increase. Thus, taking into account water molecules significantly raised the interaction surface.

Van der Waals interactions

Van der Waals (vdW) interactions were analysed in terms of energy, as described in Materials and Methods. Interaction vdW energy between every side chain of one monomer and the entire counterpart protein is depicted in Fig. 3. The results revealed two main groups of amino acid residues with favourable interaction energies; the first group is composed of Arg^{21} [with a mean LJ energy value, considering both cases, of −4.7 kcal mol^1), Ala^{22} (−2.6 kcal mol^1), Asp^{23} (−4.7 kcal mol^1), and Tyr^{25} (−1.4 kcal mol^1). The second group is formed by Met^{75} (−1.9 kcal mol^1), Lys^{76} (−3.0 kcal mol^1), Ala^{79} (−2.3 kcal mol^1), Leu^{84} (−2.4 kcal mol^{-1}) and Thr^{86} (−1.2 kcal mol^{-1}). In fact, these two groups interacted with one another when considered in different monomers.

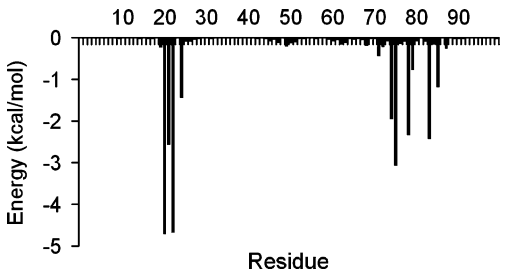


Fig. 3 Van der Waals (vdW) interaction energy between each individual side chain of one monomer and the entire partner protein. Mean values accounting for the same residue in both monomers are represented

Electrostatic interactions

Protein–protein electrostatic interaction energies throughout the production time were also analysed (Fig. 4a). Arginine and lysine residues had negative favourable values, while aspartic acid and glutamic acid residues had complex destabilising values. A single CMK monomer has 26 positively charged residues and 36 negatively charged residues. Therefore, one monomer has a high net negative charge that is translated into an unfavourable energy when analysing the electrostatic contribution of glutamic and aspartic acid residues. The net electrostatic interaction energy between monomers had a mean value of −27 kcal mol^{-1} (100 snapshots were analysed), but it was strongly favoured by the presence of intervening water molecules—the mean electrostatic interaction energy was computed to be −5,537 kcal mol^{-1} when water molecules were incorporated in the calculation. The inclusion of ions and/or CDP-ME/AMP-PNP did not change the effect of the water molecules (data not shown). In order to see the effect of the solvent on residue electrostatic contribution, water molecules were included in a further energy computation. As seen in Fig. 4b, the electrostatic energies of glutamic acid and aspartic acid residues decreased to negative values. Thus, water molecules are essential in explaining how protein–protein electrostatic interactions favour complex formation.

Database search for compounds disrupting protein–protein interactions

Atoms involved in protein–protein interactions were monitored throughout the last 200 ps of the dynamics trajectory to extract interaction models, as described in Materials and Methods. One of those models, depicted in Fig. 5, was introduced as a hypothesis in Catalyst (Accelrys). It is worth mentioning that one of the points of the model was deduced from a water molecule involved in a solvent-mediated interaction between monomers (between Asp^{23} and Lys^{76}

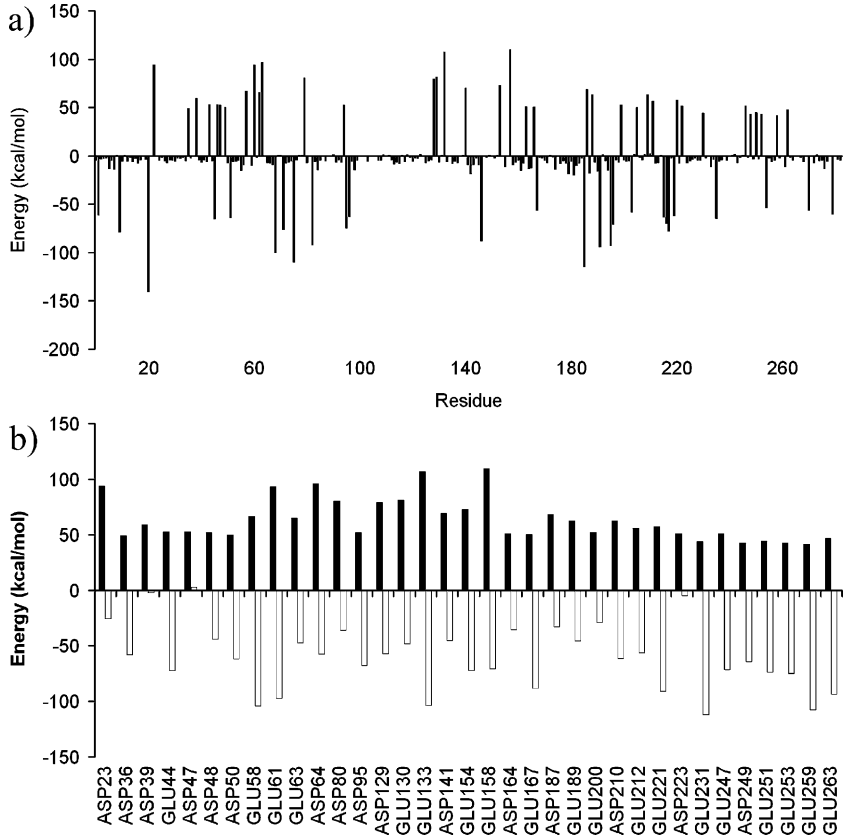


Fig. 4 **a** Coulombic interaction energy between each individual side chain of one monomer and the entire partner protein. **b** *White bars* Same as **a** for glutamic acid and aspartic acid residues but including water molecules in the partner monomer part, *black bars* same as **a** for comparison. Mean values accounting for same residue in both monomers are represented

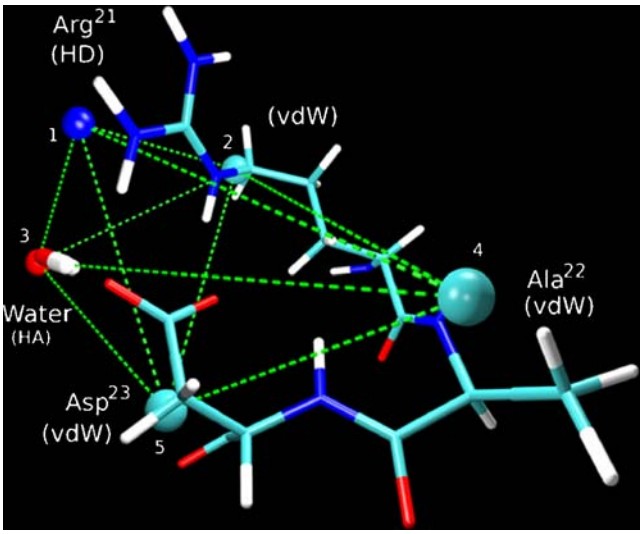


Fig. 5 Pharmacophoric points forming the interaction model used in the database search. Points are derived from residues 21–23 of monomer 2 and a water molecule. Arg^{21} provided a hydrogen donor (HD) and a vdW pharmacophoric point; Ala^{22} and Asp^{23} provided vdW pharmacophoric points; the water molecule provided a hydrogen acceptor point. Distances between pair of points are: 1–2: 2.644–3.099 Å; 1–3: 3.004–4.837 Å; 1–4: 6.965–8.907 Å; 1–5: 3.982–4.795 Å; 2–3: 3.737–5.753 Å; 2–4: 5.958–7.235 Å; 2–5: 3.755–4.644 Å; 3–4: 8.715–11.321 Å; 3–5: 3.111–5.839 Å; 4–5: 5.354–6.696 Å

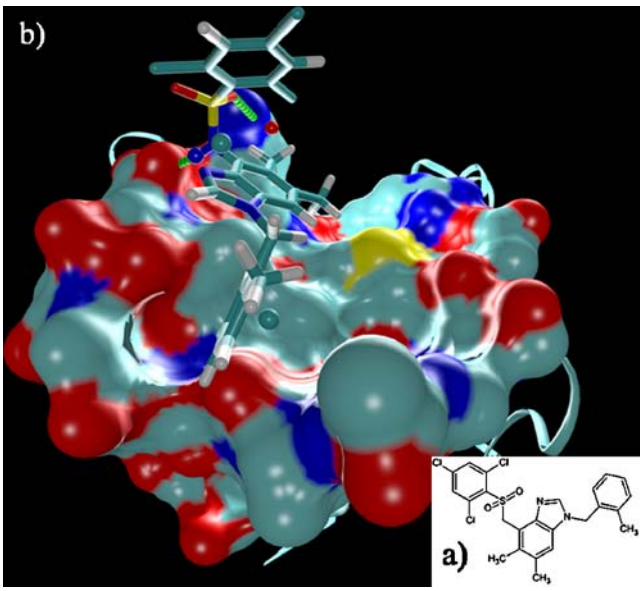


Fig. 6 **a** 2-D structure of compound 1. **b** Docking of C1 on CMK surface. CMK atoms closer than 5 Å were represented as a *molecular surface* using MSMS [51]. The remaining protein atoms were depicted following a *ribbon pattern*. Pharmacophoric points are represented as *spheres*

in interface A). Points marked *vdW* were modified to *hydrophobic* when defining the pharmacophoric hypothesis in Catalyst. The search procedure was carried out in the Available Chemical Database (ACD; 214 hits), Specs (512 hits) and ChemDiv databases (835 hits).

Hits were docked on CMK protein, as described in Materials and Methods, and poses were evaluated with the scoring function X-Score [40].

Final selection of compounds

A visual structure analysis was carried out on the 15 ligands with best docking from each database. Ranking was based on X-Score consensus values and the final selection took into account the compound structure.

Intensity-fading MALDI-TOF MS

Intensity-fading MALDI-TOF MS is a rapid and sensitive approach for detecting the formation of soluble complexes between proteins and small molecules [38]. This methodology has the advantages of low cost and rapid performance together with the capability of checking the affinity properties of the analysed compounds. In this approach, detection of the complex is based on a reduction in the molecular ion intensities of the partners in a MALDI mass spectrum compared to the mass spectra of the partners alone. In our study, we tested complex formation between one of the final selected compounds (C1) and CMK. Docking of C1 onto CMK is shown in Fig. 6. RMSD after docking between pharmacophoric features on compound

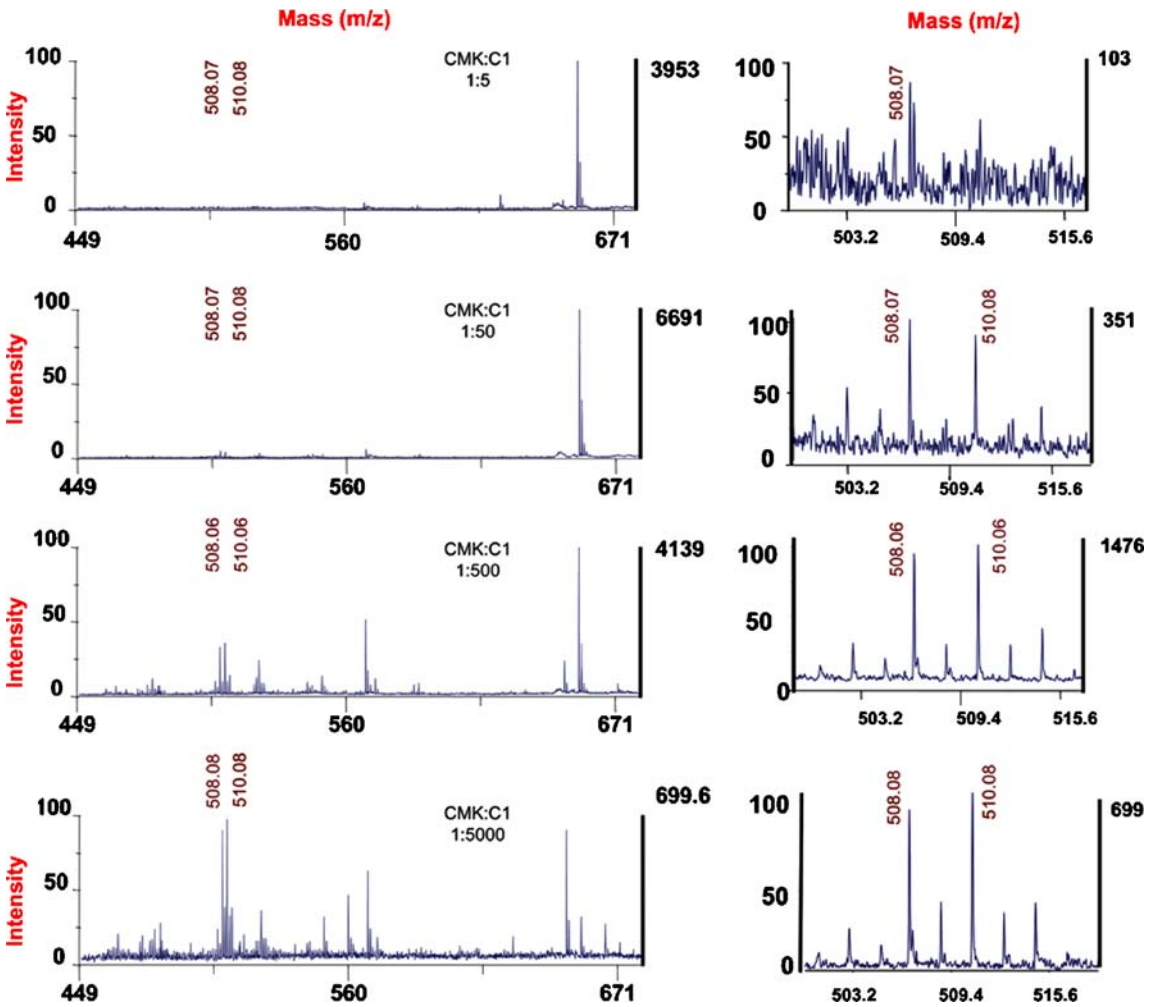


Fig. 7 Matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS) of the compound 1 (C1) molecule in the presence of increasing CMK concentrations. From *top* to *bottom*: 0.2, 0.02, 0.002, and 0.0002 mM CMK, respectively; *x*-axis mass(m/z), *y*-axis relative intensity (*left*) and ion intensity (*right*). Peaks at m/z 508 and 510 correspond to C1. A detailed view is shown on the *right panel*. Note that in order to visualize C1 m/z 508 and 510 peaks, spectra are presented at different intensity scales. The additional peaks correspond to adducts, corresponding to the α-cyano-4-hydroxycinnamic acid used as a matrix

and pharmacophore from Fig. 5 was 1.65 Å, with a pK_d estimated value (X-Score) of 5.43. Because the complex is formed between a small molecule and an enzyme, the reduction of the small-molecule ion intensity can be associated directly to formation of the complex.

The mass spectrum for C1 was acquired by adding 1 μL 1 mM C1 to 1 μL CHCA matrix and by depositing 1 μL of the mixture onto a MALDI-TOF plate. The MALDI-TOF mass spectrum of the C1+CHCA matrix showed two major peaks at m/z 508 and 510. The m/z 508 peak correlates with C1 molecular mass. In the subsequent series of experiments, 1 μL 1 mM C1 was mixed with different concentrations of CMK and the corresponding MALDI-TOF mass spectrum was acquired. As shown in Fig. 7, all C1+CMK+CHCA mass spectra obtained showed C1 peaks at m/z 508 and 510. The intensities of these peaks showed a reduction (fading) in the presence of CMK that was CMK concentration-dependent. At the highest concentration assayed (0.2 mM, Fig. 7, top), the intensity-fading of the m/z 508 and 510 peaks was almost complete, making the peak intensities similar to the baseline noise of the spectrum. These results demonstrate the formation of a complex between CMK and C1.

Conclusions

This paper describes a MD study carried out on a solvated dimer complex of CMK. The intermolecular hydrogen bonds elicited from the crystal structure were also observed in the dynamics trajectory but in a different structural pattern. Most direct protein–protein interactions in the crystal structure were observed in the dynamics trajectory as being mediated by solvent through hydrogen bonds, a pattern already seen in the crystal structure of the complex. A deeper analysis of the solvent-mediated interactions indicates that four water-mediated protein–protein interactions remain along the MD trajectory, with one of them, in interface B, following a double-water motif. These "wet spots" [31], which likely are interfacial contact residues that interact through a water molecule, and double-water bonds, play an important role in protein–protein interaction between contact surfaces, as has been described in other protein models [32–34].

The study of intermolecular electrostatic interaction energies highlights the importance of water structure. Thus, the contribution of negatively charged residues was unfavourable when considering only protein atoms, but became instead stabilising when water molecules were included in the computation. These results indicate that water molecules are essential in explaining how protein–protein electrostatic interactions favour complex formation and, together with the observations of solvent-mediated intermolecular interactions, reveal the importance of water for protein–protein interactions in the CMK dimer.

According to intermolecular interactions deduced from MD trajectory, some CMK residues could be hotspots for complex stabilisation, and thus useful targets for protein–protein interaction disruption. By monitoring the positions of Arg^{21}, Ala^{22}, Asp^{23} and a water molecule during the MD trajectory, a pharmacophore model was constructed where the first three residues of monomer 1 interacted favourably with monomer 2. The water molecules established a hydrogen bond network between Asp^{23} of monomer 1 and Lys^{76} of monomer 2. Compounds that matched the pharmacophore model were searched for in the ACD, Specs and ChemDiv databases. After the docking procedure and visual analysis, four compounds were empirically tested and one of them exhibited binding to recombinant CMK in a concentration-dependent manner according to intensity-fading MALDI-TOF MS.

The data presented here indicate that a virtual screening approach can be used to identify candidate molecules that disrupt the CMK–CMK complex. This strategy can contribute to speeding up the discovery of new antimalarial, antibacterial and herbicidal compounds.

Acknowledgements The excellent technical assistance of Drs. Eliandre de Oliveira and Maria Antonia Odena Caballol, from the Proteomics Platform, Barcelona Science Park, in the intensity-fading MALDI-TOF mass spectrometry experiments is gratefully acknowledged. This work was financed in part with grants from the University of Barcelona (ACES-UB), the Spanish Ministerio de Ciencia y Tecnología (CTQ2006-06588/BQU, BIO2002-04419-C02-02 and BIO2008-01184) and the Generalitat de Catalunya (2005SGR00914).

References

1. Sacchettini JC, Poulter CD (1997) Science 277(5333):1788–1789
2. Chappell J (1995) Annu Rev Plant Physiol Plant Mol Biol 46:521–547
3. McGarvey DJ, Croteau R (1995) Plant Cell 7(7):1015–1026
4. Croteau R, Kutchan T, Lewis N (2000) Natural products (secondary metabolites). In: Buchanan B, Gruissem W, Jones R (eds) Biochemistry and molecular biology of plants. American Society of Plant Biologists, Rockville, MD, pp 1250–1268
5. Chappell J (2002) Curr Opin Plant Biol 5(2):151–157
6. Flesch G, Rohmer M (1988) Eur J Biochem 175(2):405–411
7. Rohmer M, Knani M, Simonin P, Sutter B, Sahm H (1993) Biochem J 295(Pt 2):517–524
8. Schwarz MK (1994) Terpenbiosynthese in *Ginkgo biloba*: Eine überraschende Geschichte. PhD Thesis. Eidgenoessische Technische Hochschule, Zurich
9. Rodríguez-Concepción M, Boronat A (2002) Plant Physiol 130 (3):1079–1089
10. Rohmer M, Grosdemange-Billiard C, Seemann M, Tritsch D (2004) Curr Opin Investig Drugs 5(2):154–162
11. Boucher Y, Doolittle WF (2000) Mol Microbiol 37(4):703–716
12. Lichtenthaler HK, Schwender J, Disch A, Rohmer M (1997) FEBS Lett 400:271–274

13. Eisenreich W, Schwarz M, Cartayrade A, Arigoni D, Zenk MH, Bacher A (1998) Chem Biol 5(9):R221–R233
14. Lichtenthaler HK (1999) Annu Rev Plant Physiol Plant Mol Biol 50:47–65
15. Rohmer M (1999) Nat Prod Rep 16(5):565–574 (Oct)
16. Kuzuyama T, Shimizu T, Takahashi S, Seto H (1998) Tetrahedron Lett 39(43):7913–7916
17. Jomaa H, Wiesner J, Sanderbrand S, Altincicek B, Weidemeyer C, Hintz M, Turbachova I, Eberl M, Zeidler J, Lichtenthaler HK, Soldati D, Beck E (1999) Science 285(5433):1573–1576
18. Lichtenthaler HK (2000) Biochem Soc Trans 28:785–789
19. Testa CA, Brown MJ (2003) Curr Pharm Biotechnol 4(4):248–259
20. Wiesner J, Borrmann S, Jomaa H (2003) Parasitol Res 90(2):S71–S76
21. Missinou MA, Borrmann S, Schindler A, Issifou S, Adegnika AA, Matsiegui PB, Binder R, Lell B, Wiesner J, Baranek T, Jomaa H, Kremsner PG (2002) Lancet 360(9349):1941–1942
22. Zeidler JG, Schwender J, Müller C, Wiesner J, Weidemeyer C, Beck E, Jomaa H, Lichtenthaler HK (1998) Z Naturforsch 53 (c):980–986
23. Borrmann S, Issifou S, Esser G, Adegnika AA, Ramharter M, Matsiegui PB, Oyakhirome S, Mawili-Mboumba DP, Missinou MA, Kun JF, Jomaa H, Kremsner PG (2004) J Infect Dis 190 (9):1534–1540
24. Miallau L, Alphey MS, Kemp LE, Leonard GA, McSweeney SM, Hecht S, Bacher A, Eisenreich W, Rohdich F, Hunter WN (2003) Proc Natl Acad Sci USA 100(16):9173–9178
25. Gabrielsen M, Bond CS, Hallyburton I, Hecht S, Bacher A, Eisenreich W, Rohdich F, Hunter WN (2004) J Biol Chem 279 (50):52753–52761
26. Wells JA, McClendon CL (2007) Nature 450(7172):1001–1009
27. Sperandio O, Miteva MA, Segers K, Nicolaes GA, Villoutreix BO (2008) Open Biochem J2:29–37
28. Smrcka AV, Lehmann DM, Dessal AL (2008) Comb Chem High Throughput Screen 11(5):382–395
29. Fletcher S, Hamilton AD (2007) Curr Top Med Chem (10):922–7
30. Whitty A, Kumaravel G (2006) Nat Chem Biol 2(3):112–118
31. Samsonov S, Teyra J, Pisabarro MT (2008) Proteins 73(2):515–525
32. Jiang L, Kuhlman B, Kortemme T, Baker D (2005) Proteins 58 (4):893–904
33. Furukawa Y, Morishima I (2001) J Biol Chem 20 276(16):12983–12990
34. Langhorst U, Backmann J, Loris R, Steyaert J (2000) Biochemistry 39(22):6586–6593
35. Janin J (1999) Structure 7(12):R277–R279
36. van Dijk AD, Bonvin AM (2006) Bioinformatics 22(19):2340–2347
37. Yanes O, Villanueva J, Querol E, Aviles FX (2005) Mol Cell Proteomics 4(10):1602–1613
38. Villanueva J, Yanes O, Querol E, Serrano L, Aviles FX (2003) Anal Chem 75(14):3385–3395
39. Case DA, Pearlman DA, Caldwell JW, Cheatham TEI, Wang J, Ross WS, Simmerling C, Darden T, Merz KM, Stanton RV, Cheng A, Vincent JJ, Crowley M, Tsui V, Gohlke H, Radmer R, Duan Y, Pitera J, Massova I, Seibel GL, Singh UC, Weiner P, Kollman PA (2002) AMBER7. University of California, San Francisco, CA
40. Wang R, Lai L, Wang S (2002) J Comput Aided Mol Des 16:11
41. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) J Comput Chem 25:1157
42. Jakalian A, Bush BL, Jack DB, Bayly CL (2000) J Comp Chem 21(2):132–146
43. Jorgensen WL, Chandrasekhar J, Madura JD (1983) J Chem Phys 79:926
44. Darden T, York D, Pedersen L (1993) J Chem Phys 98:10089
45. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) J Comp Phys 23:327
46. Berendsen HJC, Postman JPM, Van Gunsteren WF, DiNola A, Haak JR (1984) J Chem Phys 81:3684
47. Bernal C, Mendez E, Terencio J, Boronat A, Imperial S (2005) Anal Biochem 340(2):245–251
48. Bradford MM (1976) Anal Biochem 72:248–254
49. Karas M, Hillenkamp F (1988) Anal Chem 60(20):2299–2301
50. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) J Comput Chem 25(13):1605–1612
51. Humphrey W, Dalke A, Schulten K (1996) J Mol Graph 14(1):33–38,27–38
52. Levy Y, Onuchic JN (2006) Annu Rev Biophys Biomol Struct 35:389–415
53. Sreenivasan U, Axelsen PH (1992) Biochemistry 31(51):12785–12791