

107  
29



UNIVERSIDAD NACIONAL AUTONOMA  
DE MÉXICO

FACULTAD DE CIENCIAS

"ESTUDIOS ESTRUCTURALES SOBRE LA  
GLUCOSAMINA 6-FOSFATO DESAMINASA DE  
*ESCHERICHIA COLI* Y OTRAS ESPECIES"

**T E S I S**

QUE PARA OBTENER EL TÍTULO DE  
**B I O L O G O**

**P R E S E N T A :**  
**RODRIGO ARREOLA ALEMÓN**

Director de tesis: **Dr. Eduardo Horjales Reboredo**



MÉXICO, D.F.

FACULTAD DE CIENCIAS  
SECCIÓN ESCOLAR

1996

**TESIS CON  
FALLA DE ORIGEN**

**TESIS CON  
FALLA DE ORIGEN**



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL  
AUTÓNOMA DE  
MÉXICO

M. en C. Virginia Abrín Batule  
Jefe de la División de Estudios Profesionales de la  
Facultad de Ciencias  
P r e s e n t e

Comunicamos a usted que hemos revisado el trabajo de Tesis:

"Estudios estructurales sobre la Glucosamina-6-fosfato desaminasa de  
Escherichia coli y otras especies."

realizado por Arreola Alemón Rodrigo

con número de cuenta 8836505-9 , pasante de la carrera de Biología

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis	Dr. Eduardo	Horales Reboredo
Propietario	<i>Raul Carrasco</i>	
Propietario	Biól. Saul	Cano Colín
Propietario	M.enC. Víctor Manuel	Valdes Lopez
Suplente	Dr. Rene de Jesús	Cárdenas Vázquez
Suplente	Dr. Juan Enrique	Morett Sánchez

Consejo Departamental de Biología

COORDINACION GENERAL  
DE BIOLOGIA

**A mi Familia:**

*A Sara Eugenia, Arturo(aunque lo dudes), Dante, Alejandro, Roberto, Rogelio, Sandra, Eric y Carlos.*

*A Mi Madre: María Eugenia Alemón Tovar.*

*A Mi Padre: Roberto Arreola Miranda.*

*A Mi Tía Rosa y Aarón.*

A todos ustedes por apoyarme **GRACIAS.**

**GRACIAS EDUARDO POR TU CONFIANZA Y PACIENCIA Y A TU FAMILIA POR AYUDARME AL PRINCIPIO CUANDO LLEGUE.**

**Agradesco a Gabriel Moreno Hagelsieb por su apoyo técnico.**

*A los chavos y chavos del Laboratorio de Reconocimiento molecular y Bioestructura. Por hacerme la vida más "relax".*

**EL PRESENTE TRABAJO FUE REALIZADO EN EL DEPARTAMENTO DE  
RECONOCIMIENTO MOLECULAR Y BIOESTRUCTURA DEL INSTITUTO  
DE BIOTECNOLOGÍA DE LA UNAM BAJO LA DIRECCIÓN DEL DR.  
EDUARDO HORJALES REBOREDO.**

# ÍNDICE

## PÁGINAS

ÍNDICE	i
RESUMEN	iii
INTRODUCCIÓN	1
OBJETIVOS	11
MATERIALES Y MÉTODOS	12
a) Búsqueda de secuencias en el Protein Data Bank.	12
b) Construcción del modelo computacional de la glucosamina 6-fosfato desaminasa humana mutante pro268stop a partir del modelo cristalográfico de la enzima homóloga de <i>E. coli</i> en su conformero R.	13
c) Construcción del modelo computacional de la probable glucosamina 6-fosfato desaminasa de <i>Haemophilus influenzae</i> a partir del modelo cristalográfico de la enzima homóloga de <i>E. coli</i> en su conformero R.	15
d) Construcción del modelo computacional de la probable glucosamina 6-fosfato desaminasa de <i>Mycoplasma capricolum</i> a partir del modelo cristalográfico de la enzima homóloga de <i>E. coli</i> en su conformero R.	15
e) Alineamiento estructural de los modelos monoméricos con el monómero de <i>E. coli</i> .	17
f) Obtención del hexámero del modelo humano por aplicación de la simetría cristalográfica de <i>E. coli</i> .	17
RESULTADOS Y DISCUSIÓN	21
Posiciones conservadas en los alineamientos de secuencias y su función estructural.	21
Secuencias homólogas a la GlcN6PD de <i>E. coli</i> : U18997 y L10328.	22
Estabilidad y calidad de los modelos.	24

Alineamiento estructural de los monómeros.	26
Hexámeros de la enzima humana.	28
Hexamerización de la proteína de <i>H. influenzae</i> .	29
La GlcN6PD de <i>Candida albicans</i> no es hexámerica.	31
Monómero de la enzima de <i>M. Capricolum</i> .	33
Entrada del amonio.	36
Sitio alostérico de la enzima humana.	37
Sitio activo de la enzima humana.	40
Sitio catalítico de la enzima de <i>Candida albicans</i> .	40
Sitio catalítico de la enzima de <i>M. capricolum</i> .	41
Caracterización del plegamiento del dominio de unión a NAD.	42
<b>CONCLUSIONES</b>	<b>45</b>
<b>APÉNDICES</b>	<b>46</b>
Apéndice 1: Homology	46
Apéndice 2 : Discover	48
Apéndice 3 : Profiles-3D	53
Apéndice 4 : Predicción de estructura secundaria.	56
<b>REFERENCIAS</b>	<b>67</b>

## RESUMEN

Usando métodos computacionales de modelaje de proteínas y de dinámica molecular se hizo un estudio sobre la relación de la estructura y la función de la catálisis química de las proteínas homólogas de la glucosamina-6-fosfato isomerasa desaminasa (GlcN6PD) de *Escherichia coli*. Además de un estudio de la homología estructural de este grupo de enzimas con el dominio de unión a NAD presente en diversas enzimas cuya estructura terciaria es conocida. De la GlcN6PD de *E. coli* se conoce la estructura cristalográfica que fue empleada como base para realizar los modelos de las proteínas homólogas. Las enzimas modeladas son las correspondientes ortólogas de *Haemophilus influenzae*, *Homo sapiens*, *Candida albicans* y *Micoplasma capricolum*. Tienen una identidad con *E. coli* de 72%, 58%, 47% y 30% respectivamente. El estudio se apoyó en datos cinéticos y de estructura de *E. coli*; y cinéticos de *C. albicans* y *H. sapiens*; de éstas se sabe que son las enzimas homólogas que realizan la misma función de desaminación de glucosamina-6-fosfato a fructuosa-6-fosfato y amonio. En el alineamiento de secuencias se muestra que los aminoácidos catalíticos (Asp 72 y His 143) están conservados en todas las secuencias y de dos posiciones variables (141 y 148) pero que siempre se mantiene un residuo ácido en alguna de las dos o ambos lo cual es esencial para la actividad de isomerasas. Las enzimas modeladas mostraron ser estables en la dinámica molecular lo cual sugiere que su plegamiento es similar al de *E. coli*. Los modelos proponen para *H. influenzae*, y confirman estructuralmente para *H. sapiens* y *C. albicans* la función catalítica de isomerización-desaminación de éstas enzimas con un mecanismo catalítico similar al de *E. coli*. Así mismo también confirman para *H. influenzae* y *H. sapiens* la estructura cuaternaria de un hexámero, formado por un trímero de dímeros, como está formado el hexámero de *E. coli*. El modelo de *C. albicans* muestra un motivo estructural por el cual esta enzima no puede ser un hexámero. La interpretación del modelo de *M. capricolum* sugiere la capacidad de isomerasa de esta enzima al conservar los aminoácidos catalíticos ya mencionados, además de conservar los ligantes principales el asa de ligadura del fosfato de la glucosamina-6-fosfato. Estos datos nos permiten suponer que esta enzima realiza una reacción similar a la de *E. coli* pero debido a los cambios en residuos del sitio activo y su periferia probablemente el sustrato sea otro azúcar. A las enzimas modeladas de *H. influenzae*, *H. sapiens*, *C. albicans* y *M. capricolum* se les propone una topología que es similar a la GlcN6PD de *E. coli*. Finalmente la comparación de estas secuencias con el dominio de unión a NAD propone que estas enzimas son derivados de este motivo, al conservar una topología similar, agregándose dos inserciones (la asa catalítico entre los residuos 137-157 y el motivo hebra beta B'/hélice 5/hebra beta C' que aísla el sustrato en el sitio activo, del solvente). También comparamos la secuencia del perfil para dominios de unión a NAD con la resultante del conjunto de las secuencias analizadas. Las regiones claves se mantienen en ambos perfiles. En conclusión estas enzimas tuvieron un ancestro común que fue derivado de un dominio de unión a NAD.



## INTRODUCCIÓN

Las proteínas, como uno de los componentes más importantes de los seres vivos, son ampliamente estudiadas ya que participan en la mayoría de las funciones celulares. Este conocimiento nos permite hacer uso de las proteínas como herramientas en la ciencia, en aplicaciones industriales y médicas. Las técnicas empleadas para su estudio son diversas (bioquímicas, de biología molecular) y de éstas, la difracción de rayos x a través de un cristal proteico es hasta el día de hoy el método mas empleado para dilucidar la estructura tridimensional de una molécula proteica. En la actualidad el número de estructuras conocidas (por difracción de rayos x y NMR) es mucho menor que la cantidad de secuencias de proteínas que se conocen. En el Protein Data Bank (Bernstein, et.al.1977) se tiene 4113 estructuras tridimensionales reportadas y en el GenBank (Rudd,K.E.1993) se tiene aproximadamente 744,000 secuencias hasta abril de 1996, de las cuales 329,000 corresponden a ESTs (expressed sequence tags). Si se sabe cómo es la estructura tridimensional de una enzima, se puede inferir cómo se llevan acabo las reacciones de catálisis, las interacciones internas de una proteína, así como las interacciones que pudiera tener con otras moléculas. Sin embargo, la obtención de un cristal proteico requiere grandes cantidades de proteína para realizar una prueba de cristalización (20mg para pruebas completas), situación que no siempre es posible, además sin considerar que una proteína puede tardar meses en cristalizar bajo condiciones especiales. Debido a estos inconvenientes desde hace unos 20 años se han desarrollado técnicas basadas en computación gráfica para poder proponer modelos tridimensionales a partir de homología de secuencias y de modelos obtenidos por cristalografía de proteínas (BrooksIII,C.L.,1988).

### **Modelaje y simulaciones.**

El modelaje es el inicio de los métodos computacionales para realizar un modelo teórico de la estructura terciaria (tridimensional) de una proteína y comprende las siguientes etapas: a) Un alineamiento de secuencias entre la proteína cuya estructura tridimensional es conocida y la proteína que se pretende modelar. b) Cuando se cuenta con más de una estructura tridimensional conocida se realiza un alineamiento estructural para inferir un posible arreglo estructural de los aminoácidos en regiones de baja similitud. Las regiones variables se reconstruyen por búsqueda de una estructura conocida adecuada. c) Finalmente, el modelo se refina con una simulación. Esta última fase es una

técnica que simula los movimientos de una molécula, ya sea en el vacío o en solución (con una esfera de solvente), a cierta temperatura y presión (la técnica a utilizar puede variar con la finalidad del modelo).

Los métodos de simulaciones se han implementado ampliamente en los últimos años debido al desarrollo de computadoras más poderosas en el manejo de gráficos y la rapidez en el procesamiento de datos. Asimismo, se han desarrollado "software" cada vez más complejos para el manejo de gráficos y simulaciones (Brooks III, C.L., 1988). De los métodos de simulación, la minimización de energía y la dinámica molecular son los empleados en este trabajo. Una simulación sirve para proponer un modelo aproximado de la estructura terciaria de una proteína que sea químicamente y estructuralmente competente, para poder interpretar la posible función biológica (Almagro, J.C. et al. 1995. Hoffrén, A.-M., et al. 1995. Stigler, R.-D., et al. 1995. Fuxreiter, M. et al. 1995). Estos modelos nos pueden ayudar a resolver, a nivel atómico, los procesos y mecanismos moleculares de la catálisis química fundamentados en estructura terciaria. Actualmente, es posible obtener modelos confiables basados en estructuras cristalográficas. Por lo tanto, la interpretación de un modelo bien refinado puede proponer mecanismos catalíticos e interacciones con algunas moléculas que pueden ser ciertos con un alto grado de confianza, al compararse con la estructura cristalográfica que originó el modelo o con otra estructura cristalográfica homóloga.

Los modelos computacionales son propuestas aproximadas y presentan ciertos límites en la exactitud de los mismos (como en las asas que suelen ser variables en secuencia y tamaño) por lo cual deben apoyarse por los datos del modelo cristalográfico en el que se basó para su construcción y datos de resultados experimentales de la proteína modelada.

Sin embargo, un modelo computacional puede someterse a verificación de calidad por medio de programas, los cuales buscan los errores y la posibilidad de la existencia de ciertas posiciones de los residuos con base en las comparaciones con estructuras conocidas de los bancos de datos de estructuras y también verifican que los aminoácidos del modelo estén en un ambiente adecuado (Profile3-D) (Lüthy, et al. 1992). El modelo se trabaja como si se tratara de un modelo cristalográfico verificando la calidad del refinamiento.

### **Tipos de simulación.**

Las simulaciones pueden ser del tipo dinámico o de tipo termodinámico. Las simulaciones del

tipo dinámico emplean fórmulas de movimiento (la ecuación de movimiento de Newton) y parámetros de energía potencial para describir interacciones de los átomos. Además, en las simulaciones de tipo termodinámico pueden incluirse conceptos termodinámicos, efectuando dinámicas a temperatura y presión constante (apéndice 1). El método utilizado en este trabajo es de tipo termodinámico. Un aspecto importante que hay que dejar claro es que las minimizaciones de energía (también llamadas mecánica molecular) nos dan como resultado modelos geoméricamente correctos muy similares a los modelos iniciales. La dinámica molecular se emplea en modelización, para verificar la estabilidad del plegamiento propuesto del modelo. Un modelo con baja estabilidad tenderá a ser destruido en el proceso de dinámica y a deformarse.

El actual trabajo propone algunos modelos tridimensionales usando computación gráfica y pretende realizar una comparación estructural entre los modelos computacionales y un modelo cristalográfico, a partir del cual se obtuvieron los primeros. Se pretende deducir, a partir de la estructura cristalográfica de la glucosamina-6-fosfato desaminasa isomerasa y de la estructura de los modelos teóricos de sus enzimas homólogas, la forma en que estas enzimas funcionan de manera general y marcar sus diferencias.

#### **La Glucosamina-6-fosfato desaminasa isomerasa (GlcN6PD).**

Los modelos teóricos se basan en la estructura del modelo obtenido por cristalografía de rayos x de la GlcN6PD de *Escherichia coli* (Oliva, G., et al. 1995) cuyas coordenadas están registradas en el Protein Data Bank (entrada 1DEA). De este grupo de enzimas, es la única proteína cuya estructura se conoce. También ha sido caracterizada bioquímicamente (Calcagno, M. et al. 1984). La GlcN6PD de *E. coli* cataliza la reacción de isomerización y desaminación de la D-glucosamina-6-fosfato (GlcN6P) en D-fructuosa-6-fosfato y amonio (Comb, D.G. 1958; Midelfort, C.F., 1977). La secuencia de aminoácidos se dedujo a partir de su DNA (Rogers, M.J., 1988). La GlcN6PD de *E. coli* es homóloga a la GlcN6PD de humano (Lara-Lemus, R., 1996; Weidanz, J.A. et al., 1995a) y a la GlcN6PD de *Candida albican* (Datta, A., 1993) las cuales realizan la misma función. Así también se ha detectado su actividad en varias especies, en mamíferos (Lara-Lemus, R., 1992), en protozoarios (Weidanz, J.A. et al., 1995b), y bacterias (Vogler, A.P., 1989; Homer, K.A., 1993) pero no se conoce la secuencia del gen. Sin embargo, se han secuenciado genes que tienen homología con el de la GlcN6PD de *E. coli*, como es el caso de *Haemophilus influenzae* (Fleischmann, R.D., et al., 1995) y

*Mycoplasma capricolum* (Gillevet P.1994) y muy recientemente de *Mesocricetus auratus* (Parrington,P.,et.al.,1996).

#### **Bioquímica y genética de la GlcN6PD de *E. coli*.**

La GlcN6PD de *E. coli* es una enzima alostérica típica de la clase K que está bien caracterizada (Comb,D.G.1958;Midelfort,C.F.,1977;Calcagno,M.,1984); activada por N-Acetil-D-glucosamina-6-fosfato (GlcNAc6P). Las enzimas alostericas, por lo general, muestran curvas sigmoideas cuando se mide la velocidad de reacción frente a la concentración de sustrato, a diferencia de las curvas hiperbólicas predichas por la ecuación del modelo de Mikaelis-Menten. En una enzima alostérica el centro activo de una subunidad enzimática afecta a otro sitio activo localizado en otra subunidad enzimática de la misma proteína. La unión de la proteína al sustrato se vuelve mas afin (la unión del sustrato se vuelve cooperativa) como resultado de esta interacción entre subunidades (Palmer,T.,1985). La interacción entre las subunidades esta regulada por una molécula (o varias moléculas) diferentes del sustrato que se ligan en sitios diferentes al sitio activo. Una proteína que se cataloga como clase K (un sistema K puro), es aquella en que el efector y el sustrato tienen diferentes afinidades hacia las conformaciones T y R. El activador cambia la afinidad de la enzima por el sustrato, pero no afecta la  $V_{max}$  de la reacción. La presencia del efector es suficiente para modificar la afinidad de la enzima por su sustrato y esto puede ocurrir en sentido opuesto, en donde el sustrato modifica la afinidad por el efector (Monod,J.,1965). Todo esto pasa sin que se modifiquen las propiedades catalíticas (donde la  $K_{cat}$  se mantiene igual) al pasar del estado R al T.

Esta enzima presenta una  $K_m$  de  $2.01 \pm 0.05$  mM en el tipo Wild-type, el coeficiente de Hill,  $h_{max}$ , es  $3.02 \pm 0.09$  (Altamirano,M.M.,1995). La GlcN6PD forma parte de una vía para la degradación de azúcares aminados (especialmente glucosa) y junto con sus actividades antagónicas de la ruta biosintética, tienen una regulación coordinada entre ambas vías (Plumbridge,J.A.,1983). Los aminoazúcares GlcN y GlcNAc forman parte esencial de la composición de los peptidoglucanos de la pared celular bacteriana y de los lipopolisacáridos de la membrana exterior en bacterias Gram negativo .xterior en bacterias Gram negativo .

Cuando la bacteria está en un medio rico en GlcN y/o GlcNAc se activan los genes para su degradación. Éstos están en los operones *nagE* y *nagBACD* a 15 minutos sobre el mapa de *E.coli*.

El gen *nagE* codifica una proteína de membrana para la asimilación de GlcNac y *nagBACD* codifica tres proteínas, *nagA* para N-Acetilglucosamina-6-fosfato desacetilasa (EC.3.5.1.25), *nagB* para Glucosamina-6-fosfato desaminasa (EC.5.3.1.10) involucradas en la degradación de la GlcN6P y GlcNac6P y su transformación a Fru6P y amonio (9, 10, 11, 12, 13), y *nagC* que codifica para un represor que controla la expresión de los genes de la desaminación y desacetilación (*nagB* y *nagA*), el cual está regulado por la presencia de GlcNac6P, que impide que el represor se una a su operador. *NagC* controla la expresión divergente de los operones *nagB-nagBACD*. *NagC* presenta una doble función como activador y como represor, e implica el control de las vías biosintéticas y catabólicas para aminoazúcares en *E. coli* (Plumbridge, J.A., 1995). Cuando la bacteria se encuentra en un medio donde la GlcN es escasa o falta, ésta se ve obligada a sintetizarla vía glucosamina-6-fosfato sintetasa (EC. 2.6.1.16) codificada por el gen *glmS* localizado en el minuto 84 del mapa cromosómico de *E. coli*, a partir de Fru6P y glutamina. Cultivos con glucosa, GlcN y GlcNac muestran que estos azúcares producen altos niveles de glucosamina-6-fosfato desacetilasa y desaminasa (se induce unas 5 veces con GlcN y unas 20 veces con GlcNac). La GlcN y GlcNac también reprimen de dos a tres veces la síntesis de la enzima glucosamina sintetasa. A nivel de control genético se ha visto que en cultivos con GlcN y GlcNac se presentan incrementos considerables de los transcritos de *nagB* y decrementos en los del gen *glmS* comparados con cultivos en glucosa. La regulación de la GlcN6PD de *E. coli* también se lleva a cabo de manera alostérica cuando los niveles de GlcNac bajan y la enzima se comporta con menor afinidad por el sustrato.

#### **Descripción de la estructura tridimensional de la GlcN6PD de *E. coli*.**

La enzima de *Escherichia coli* es un oligómero cuya estructura cuaternaria está compuesta por seis cadenas polipeptídicas idénticas de una masa molecular de 29.7 KDa cada una, formando un hexámero con estructura de trímero de dímeros, pues presenta un eje de simetría de orden tres y tres ejes de orden dos de simetría local (Altamirano, M.M., 1992), (Hojales, E., 1992). El plegamiento del monómero de la GlcN6PD puede ser descrito como un dominio de unión a NAD modificado por medio de la adición de algunas partes a la estructura; es una típica estructura  $\alpha/\beta$  abierta conformado por 7 hebras  $\beta$  paralelas que forman una hoja  $\beta$  central rodeada de 8 hélices  $\alpha$  y una hélice 3-10. Una segunda hoja  $\beta$  antiparalela de tres hebras aparece de la continuación de la hebra  $\beta$  C por su C-terminal. Estas estructuras son nombradas por números para las  $\alpha$  hélices (de 1 a 8), por letras para

las hebras  $\beta$  paralelas (de A a G) y por letras también para las hoja antiparalela (A',B' y C'). Las adiciones al dominio de unión a NAD son la hebra  $\beta$  A con la hélice 1 (del N-terminal al residuo número 35), la región que va del residuo 243 al 266 (el C terminal asa/hélice 8/asa ) y la sección que va del residuo 141 al 188 (zona asa/beta B'/hélice 5/asa/beta C'). La hélice 7, una hélice  $\alpha$  que esta seguida por una asa de siete aminoácidos y por una hélice 3-10, es equivalente a una hélice  $\alpha$  en un dominio de unión a NAD. Esta zona está implicada en los contactos intermonómero próximos al eje de simetría de orden tres. La hebra  $\beta$  A está implicada en el proceso alostérico y forma parte de este sitio con el N-terminal junto con los residuos 158 y 160 de la hebra  $\beta$  B' que es una de las adiciones. Esta adición esta formada por las hebras  $\beta$  C', B' y por la hélice  $\alpha$  5, localizada entre ambas hebras. Esta sección adicional contiene también la asa catalítica (141-148); las hebras  $\beta$  de esta estructura forman parte de la hoja  $\beta$  antiparalela junto con la hebra  $\beta$  A'. Este motivo estructural ( $\beta$  B', hélice 5,  $\beta$  C') es de gran importancia pues está también implicado en el proceso catalítico funcionando como una tapa que aísla al substrato para que se lleve a cabo la reacción. La última región adicionada es la región del C-terminal la cual esta compuesta por dos zonas: una asa y la hélice 8. Esta asa forma parte de las interacciones intermonoméricas cercanas al eje de simetría de orden dos del cual se conforma el dímero.

#### **Descripción como dominio de unión a NAD.**

Se ha dicho anteriormente que el plegamiento de la GlcN6PD de *E. coli* se puede describir como un dominio de unión a NAD. Los dominios de unión a NAD tienen una estructura tridimensional similar y presentan homología de secuencia (Branden,C.&Tooze,J.,1991;Wierenga,R.K.et.al,1986). Un dominio de unión a NAD es una hoja  $\beta$  abierta de 6 hebras paralelas con hélices a ambos lados de la hoja  $\beta$ . Esta estructura es simétrica en su conformación y esta formada por dos mitades iguales formadas por dos motivos estructurales  $\beta$ - $\alpha$ - $\beta$  cada uno. A cada una de las mitades se les denomina motivo de unión a mononucleótido y esto se debe a que cada mitad enlaza a uno de los dos nucleótidos del NAD, también se les conoce como "Rossmann folds" (Rossmann,M.G.,1974). Cada mononucleótido del NAD esta formado por una ribosa, un fosfato (en posición 5) y una base nitrogenada. Ambos "Rossmann folds" forman un dominio de unión a NAD por medio de las hebras  $\beta$  1 y 4 para formar una hoja  $\beta$  de 6 hebras (figura X). Esto sucede para un típico dominio de unión a NAD. En la glucosamina-6-fosfato desaminasa

se presentan modificaciones. Sin embargo, el plegamiento sigue las mismas características. Los dominios de unión a NAD se presentan en diferentes enzimas como la alcohol dehidrogenasa, lactato dehidrogenasa, glutatión reductasa y algunas otras. De algunas de éstas se conoce su estructura cristalográfica. El plegamiento de la glucosamina-6-fosfato desaminasa se modifica para ligar solo un azúcar fosfatado. El dominio de unión a mononucleótido tiene una región a la cual se enlaza parte del pirofosfato del NAD, esta región está muy conservada en varias enzimas que ligan NAD y es el motivo estructural  $\beta A-\alpha 1-\beta B$  del dominio de unión a NAD. Dicho motivo estructural es equivalente al motivo estructural  $\beta B-\alpha 2-\beta C-\beta A'$ . Recuérdese que las hebras  $\beta C$  y  $A'$  están separadas por dos residuos únicamente y forman parte de las modificaciones de la glucosamina-6-fosfato desaminasa. El otro dominio de unión a mononucleótido de la glucosamina-6-fosfato desaminasa de *E. coli* se localiza en la estructura  $\beta E-\alpha 6-\beta F$ , pero ésta no es funcional como ligante de azúcares fosfatados (se pierde la función de ligar pirofosfato de un NAD) y está modificada por la inserción de la asa catalítica (137-157). Después de este loop se continúa con la inserción del motivo  $\beta B'-\alpha 5-\beta C'$  que termina en la hélice 6.

#### **Mecanismo catalítico.**

El mecanismo catalítico (Oliva, G., 1995) de la GlcN6PD se conoce casi en su totalidad y se deriva del modelo cristalográfico de la enzima en complejo con el inhibidor 2-desoxi-2-amino D-glucitol-6-fosfato, que es un análogo del estado de transición de la reacción de isomerización-desaminación. A partir de éste se afinó un modelo con el sustrato con el anillo cerrado. El residuo de la His 143 se encuentra en una posición junto con el Asp 141 y el Glu 148 para formar un sistema de intercambio de protones observado en varias enzimas (Blow, D.M., 1992), (Fink, A.L., 1987) donde el aminoácido que cataliza la reacción es polarizado por un grupo ácido cargado cercano. Este sistema es el responsable de la abertura del anillo, el O5 contacta con el Ne2 de la His 143 (que es un contacto observado en el inhibidor y en el modelo con el sustrato cerrado). Esta His 143 junto con los aminoácidos Asp 141 y Glu 148 llevan a cabo la abertura del anillo. Una vez que se logra la abertura del sustrato, la siguiente etapa está relacionada con el aminoácido Asp 72, el cual reacciona con el H del C2 del sustrato para llevar a cabo una enolización del sustrato para formar un intermediario cis-enolato-amonio. Posteriormente, el protón extraído se adiciona al O1 para formar el enol-amonio. De la siguiente etapa pueden existir dos formas distintas de llevarse a cabo. En la

primera forma, nuevamente, un protón del amonio es trasferido al Asp 72, la cis-enol-amina resultante atrae el protón del Asp 72 para el C1 del intermediario con lo que se reestructura la molécula formando una fructosimina 6-fosfato. El grupo imina de la fructosimina 6-fosfato reacciona con agua para formar un carbinolamónio que es inestable. La segunda forma involucra una molécula de agua directamente para atacar al cis-enol-amónio en el enlace  $\pi$  C1-C2. La fase final incluye al intermediario final el cual se espera se descomponga en fructosa 6-fosfato y el ion amonio (carbinolamónio). Estudios no publicados de cinética realizados en el laboratorio de M.L.Calcagno muestran que el substrato de la reacción reversa parece ser el ion amonio y no el amoniaco.

La GlcN6PD pertenece a la clase de proteínas aldosa-cetosa isomerasas. Dicha clase lleva a cabo una catálisis de manera muy similar la cual remueve un hidrógeno del carbono dos de la aldosa y consecuentemente esta reacción lleva a un intermediario cis-endiol y posteriormente una nueva abstracción de un protón y la trasferencia a lo largo del plano del endiol, catalizado por un grupo básico. Adicionalmente esta enzima también es una aminotransferasa.

#### **Proteínas conocidas homólogas a la GlcN6PD de *E. coli*.**

La enzima homóloga humana de la cual se conoce la secuencia nucleotídica de su gen (Namura,N.,1994) cuyo producto peptídico es de 289 aminoácidos, contra 266 de la enzima de *E. coli*, muestran una identidad del 58.27%. La enzima humana, al igual que la de *E. coli*, presenta un hexámero con los mismos tipos de simetría. Se sabe que presenta mayor afinidad al substrato y su comportamiento alostérico es heterotrópico exclusivamente y no presenta activación homotrópica. Es un sistema "v" puro, aunque presenta trazas de cooperatividad. Un sistema "v" puro es aquel donde un activador modifica la  $V_{max}$  pero no el valor aparente de la  $K_m$  para el substrato (Palmer,T.,1985). De esta enzima se ha obtenido un mutante al cual le faltan 22 aminoácidos terminales (posee únicamente 267) y que puede ser muy similar en estructura a la de *Escherichia coli*. A la mutante se le ha nombrado pro268stop (Lara-Lemus,R. et.al. comunicación personal) y de datos experimentales se sabe que no es alotérica. Sin embargo el activador (GlcNAc6P) conserva la afinidad por su sitio de unión. En ausencia de activador el hexámero se disocia lentamente en monómeros inactivos y cuando esta presente parece estabilizar el hexámero.

Actualmente, se conocen también la secuencia primaria de las probables glucosamina-6-fosfato desaminasa isomerasa de dos bacterias: *Haemophilus influenza* y *Mycoplasma capricolum*,



localizadas por comparación de las secuencias de *E. coli* con los distintos bancos de genes. Estas dos secuencias presentan alto grado de similitud; en especial la de *Haemophilus influenza* con una identidad del 72.6% con *E. coli*. En *M. capricolum* se encontró una identidad del 32.54% con la enzima humana y de 30.66% con *E. coli*. Se han relacionado los aminoácidos esenciales para la catálisis y el plegamiento y en ambos casos están conservados. También se conoce la secuencia de la proteína de una levadura: *Candida albicans* con una identidad del 47% con *E. coli*. *C. albicans* con la enzima humana tiene una identidad del 51%. Con base en la similitud con *E. coli* se ha realizado un modelo computacional (Richard Garrat, comunicación personal). La enzima desnaturalizada de *C. albicans* tiene un peso de 28 a 27.5 kDa (Datta.,A.,1993) y la enzima nativa de 43 kDa. Presenta una Km para el sustrato de 0.1 mM en un intervalo de pH de 6.5 a 8 (Das,M.,1982). El activador alostérico, GLcNAc6P, de *E. coli* no presenta efecto alostérico sobre la actividad enzimática de *Candida albicans*; sin embargo, la GlcNAc induce la ruta catabólica para la utilización de este compuesto como fuente de carbono. La molécula además forma parte de la pared celular de la levadura y es un componente de algunas glicoproteínas.

Topología propuesta para las enzimas conocidas de la familia de la GlcN6P-D

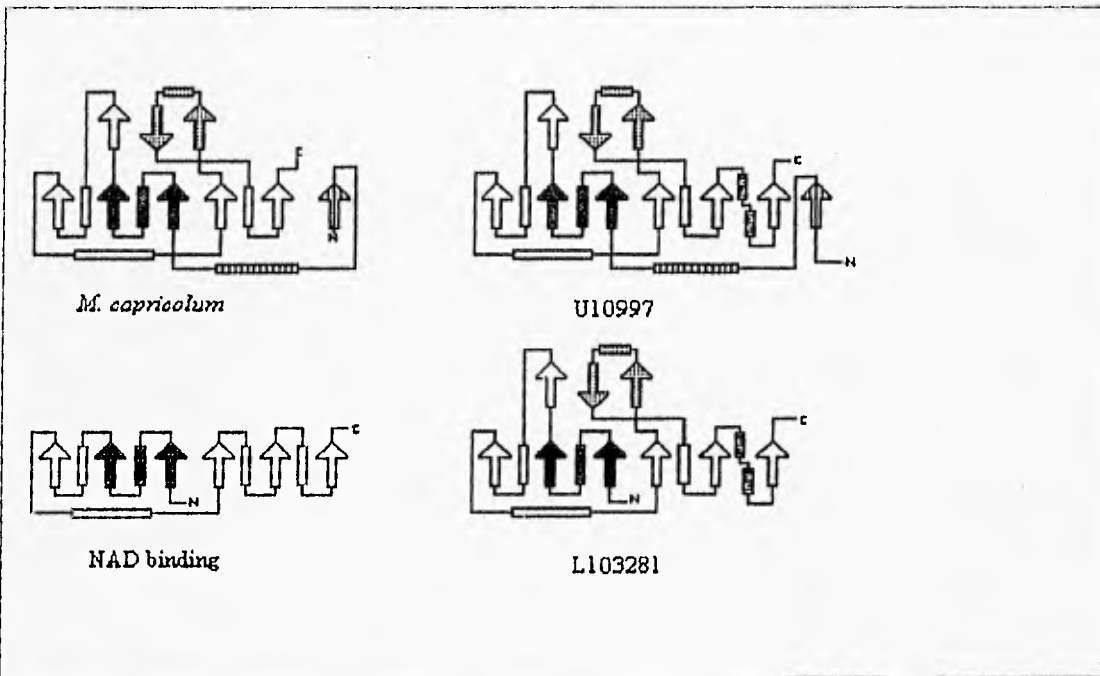
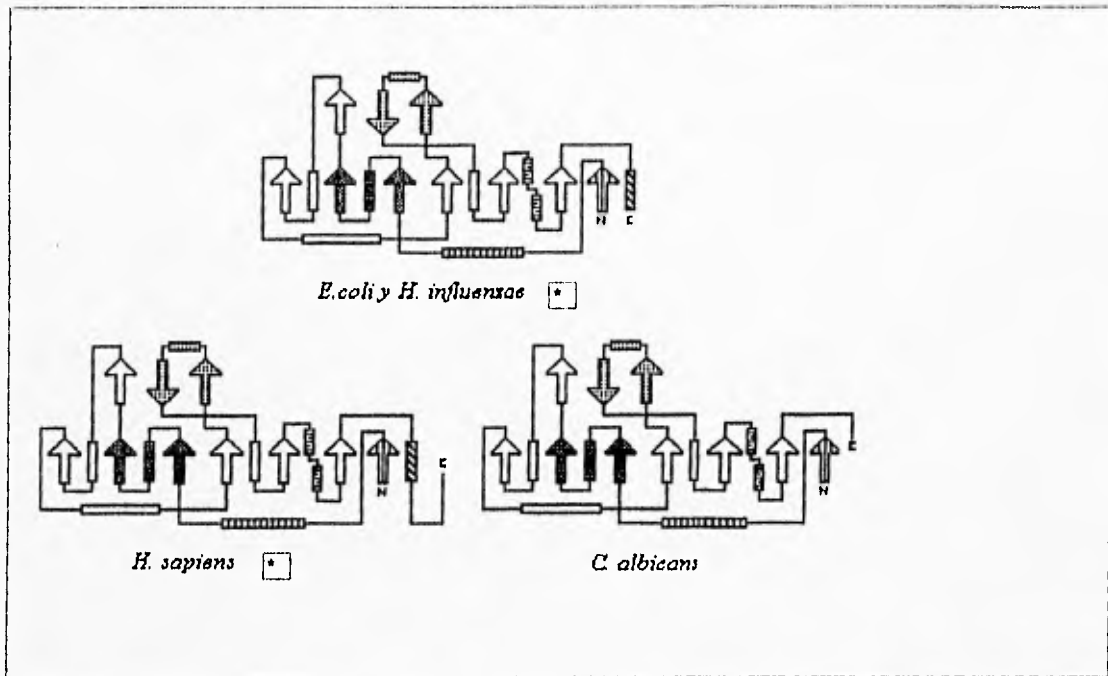


Figura 1.- Topologías. El "•" indica las enzimas hexaméricas.

- |  |                         |
|--|-------------------------|
| Inserción del motivo estructural beta B'/hélice S/beta C'. | Dominio de unión a NAD. |
| Modificación de la hélice 7.                               | Inserción N-terminal.   |
| Motivo estructural de unión a NAD.                         | Inserción C-terminal.   |

## Objetivos

- Localizar las secuencias homólogas a la glucosamina-6-fosfato desaminasa de *E. coli* en los bancos de datos de nucleótidos y aminoácidos; y realizar un alineamiento múltiple basado en la estructura cristalográfica de esta proteína.
- Obtener un modelo tridimensional de la estructura de la enzima mutante de la glucosamina-6-fosfato desaminasa humana pro268stop mediante el uso de la modelización computacional, a partir de la similitud de secuencias y de la estructura de la proteína de *Escherichia coli*.
- Obtener un modelo tridimensional de la estructura de las enzimas probables de la glucosamina 6-fosfato desaminasa de *Haemofilus influenza* y *Micoplasma capricorum* mediante el uso de la modelización computacional a partir de la similitud de secuencias y de la estructura cristalográfica de *Escherichia coli*.
- Relacionar las diferentes propiedades catalíticas y alostéricas del modelo cristalográfico de *E. coli* con las enzimas modeladas mediante la comparación de estructura tridimensional y datos bioquímicos de las enzimas humana y de *Candida albicans*.

## MATERIALES Y MÉTODOS

### a) Búsqueda de secuencias en el Protein Data Bank.

La búsqueda de secuencias homólogas a la secuencia proteica de *E.coli* se realizó utilizando el programa BLAST 1.4.6MP (Altschus,S.F.,et.al.,1990), en la versión para proteínas "blastp" y con los parámetros modelo. De esta búsqueda se encontró la secuencia de *Candida albicans*, *Haemophilus influenzae* y *Mycoplasma capricolum*, además se localizaron dos secuencias de *E.coli* que presentan conservados los aminoácidos catalíticos (TablaX). Estas dos secuencias se les denominó L10328 (Burland,V.D.,et.al.,1993) y U18997 (Plunkett,G.,1994). Los bancos de datos consultados fueron: PDB (Bernstein, et.al.1977), SWISSPROT(Bairoch,A.&Apweiler,R1996), PIR(George,D.G.et. al.1996), SPUPDATE, GENPEPT y GPUPDATE. De L10328 y U18997 no se realizaron modelos y únicamente se trabajó sobre la secuencia para explicar la homología. La secuencia L10328 se reportó como una proteína hipotética de 23.3 kDa de 213 aminoácidos perteneciente a la familia glucosamina/galactosamina-6-fosfato isomerasa (Plunkett,G.,1994). Se consideraron los alineamientos del programas BLAST, pero también se realizó un alineamiento (programa Pileup, Devereux,J.,et.al.1984) de las secuencias encontradas que fueron siete en total y no incluye la secuencia de hámster dorado: *Mesocricetus auratus* (Parrington,P.,et.al.,1996) que es 96.88% idéntica a la secuencia humana. Posteriormente con un criterio estructural se refinó el alineamiento de secuencias: El criterio estructural se basa en el hecho de que estructuras secundarias con secuencias de aminoácidos distintos pueden formar estructuras secundarias similares manteniendo un perfil característico. Un perfil de la estructura caracteriza el plegamiento de una proteína por conservar aminoácidos con características similares en las mismas posiciones de la secuencia. La secuencia de aminoácidos es continua y un indel suelen provocar cambios en las longitudes de las asas. Sin embargo, el patrón de tipos de aminoácidos que mantiene el plegamiento, principalmente los grupos hidrófobicos, se suelen mantener en las mismas posiciones en proteínas relacionadas. Muchas veces, una delección o una inserción colocada a la mitad de una hélice, por un programa que realiza alineamientos, podrá recorrerse a una de las dos asas que la flanquean. La secuencia de la hélice se puede modificar, pero conserva el patrón hidrofóbico. En otras ocasiones los cambios son reales, provocando modificaciones de un tipo de hélice a otro (cambio de una hélice  $\alpha$  a una hélice

3-10) o la pérdida o aumento de una vuelta. Este problema surge cuando la similitud es baja entre las secuencias.

**b) Construcción del modelo computacional de la glucosamina 6-fosfato desaminasa humana mutante pro268stop a partir del modelo cristalográfico de la enzima homóloga de *E.coli* en su conformero R.**

1-. Obtención de coordenadas tridimensionales a partir del modelo cristalográfico de *E.coli* y rearreglo estructural de aminoácidos.

La construcción del modelo tridimensional del mutante pro268stop de la enzima Glucosamina-6-fosfato desaminasa humana requirió un alineamiento previo de secuencias con *E.coli* para buscar las similitudes entre ambas. El alineamiento no mostró ningún tipo de inserción o deleción y fue realizado con el programa BLAST 1.4.6MP. La obtención de coordenadas iniciales del modelo humano se obtienen sustituyendo la secuencia de aminoácidos (InsightII 2.3.Homology 2.3) de la estructura cristalográfica de la glucosamina-6-fosfato desaminasa de *E. coli* por la secuencia homóloga de la enzima humana hasta el aminoácido 266. Esto permite asignar coordenadas a la secuencia de aminoácidos de una proteína. Esto se realizó así debido a que la secuencia de aminoácidos de una proteína que presenta más del 25% de identidad y en particular tiene todos los aminoácidos importantes para la función catalítica, conserva la estructura del plegamiento. Una vez teniendo coordenadas tridimensionales, los aminoácidos que presentaron diferencias fueron acomodados, tanto en el interior de la molécula como en el exterior, según los rotámeros teóricos posibles de las cadenas laterales del Programa "O" v.5.9.1(Jones,J.A.,et.al.,1991). Los rotámeros teóricos de los aminoácidos se acomodaron según el espacio disponible dentro de la molécula, dando preferencia a la posición del rotámero cristalográfico. Los aminoácidos diferentes se acomodan de acuerdo con el entorno que los rodea dentro de la proteína. Se elige un rotámero teórico lo más adecuado posible respecto a las distancias de los aminoácidos que le rodean, evitando contactos que superponga átomos. Los cambios no conservativos fueron analizados para tratar de ubicar posibles interacciones con otros aminoácidos de la proteína. Esto también permitió ubicar los cambios de mayor importancia dentro de la molécula. Así se obtuvo un modelo primario para ser refinado con los métodos de simulación computacional de modelaje molecular.

2.-Refinamiento del modelo computacional de la glucosamina 6-fosfato desaminasa humana mutante pro268stop, tratado como monómero, por minimización de energía y dinámica molecular computacional.

El modelo computacional basado sólo en la alineación de secuencia sobre estructura cristalográfica se refinó mediante el programa InsightII y el módulo Discover, con el cual se realizó una minimización de energía y una dinámica. Este modelo sólo toma en cuenta la cantidad de aminoácidos igual al modelo cristalográfico de *E. coli* (266 aminoácidos) ya que no se sabe cómo pueda continuar la estructura. La mutante pro268stop es de 267 aminoácidos por lo tanto el modelo presenta dos aminoácidos menos. El modelo además se trató como una proteína monomérica (en realidad la proteína nativa forma un hexámero).

Se realizó una dinámica molecular global de toda la estructura de 20,000 interacciones, equivale a 20 picosegundos de simulación, con un resumen de la dinámica de 500 pasos, de los cuales se obtuvo un promedio de las últimas conformaciones para obtener el modelo final de la dinámica molecular. El refinamiento incluye una previa minimización de 1000 interacciones para obtener los datos iniciales de los que parte la dinámica molecular. El proceso de la dinámica se realizó a 300° K con una constante dieléctrica de 1 dependiente de la distancia. Los cálculos se realizaron con base en un corte (cut-off) de 10 Å para las interacciones entre átomos que no presentan enlaces covalentes. Para la simulación se adicionó una esfera de solvente (agua) de 3 Å al rededor de la proteína a pH 7. (Insight II 2.3). Durante la minimización de energía, así como durante la dinámica molecular, no hubo restricciones de los movimientos de los átomos y el modelo se refinó como un monómero.

### 3.-Obtención del modelo computacional del sitio alostérico.

Para obtener este modelo fue necesario sobreponer el modelo humano con el modelo cristalográfico de *E.coli*. El acoplamiento de una estructura con otra (InsightII, homology) consiste en tomar intervalos de la secuencia de aminoácidos homólogos de ambas moléculas y sobreponerlas en un alineamiento estructural. Posteriormente, por el método de mínimos cuadrados se busca la orientación que hace mínima la suma de los cuadrados de las distancias entre los carbonos  $\alpha$  correspondientes entre los modelos. El alineamiento de cajas se realizó con el criterio de estructuras secundarias más conservadas: del aminoácido 35 al 71 que es una zona muy conservada entre ambas

proteínas (hebra beta B, asa, hélice 2, asa, hebra beta C, asa). También del 73 a 78: continua muy conservado (hebra beta A', asa); 133 a 136: (hebra beta E); 199 a 203: (hebra beta F); 235 a 239: (hebra beta G). Se incluyen las hebras  $\beta$  porque conservan mejor los patrones geométricos y de secuencia del plegamiento de Rossmann. Las hebras  $\beta$  E, F y G forman el núcleo de uno de los motivos de unión a mononucleótido (y que está modificado por una inserción) y las hebras  $\beta$  B, C y D el segundo núcleo y principal el cual liga al sustrato. La hebra  $\beta$  D no se incluyó porque está muy expuesta al solvente y varía más en su secuencia y puede presentar desviaciones mayores. En estructuras  $\alpha/\beta$  de hoja abierta entre proteínas emparentadas, el patrón de hebras  $\beta$  es muy similar y las hélices suelen variar más en su geometría (Branden, C. & Tooze, J., 1991). Dos monómeros del modelo teórico humano se sobrepusieron con el hexámero cristalográfico de *E. coli* para obtener el sitio alostérico de la enzima humana. El sitio alostérico, el cual se localiza entre la interfase de dos monómeros pertenecientes a dímeros distintos cada uno, también se realizó con la misma metodología descrita anteriormente.

**c) Construcción del modelo computacional de la probable glucosamina 6 fosfato desaminasa de *Haemophilus influenzae* a partir del modelo cristalográfico de la enzima homóloga de *E. coli* en su conformero R.**

La técnica utilizada es la misma que se utilizó con el modelaje de la enzima humana y los mismo criterios se aplicaron en los pasos 1 y 2 (ver obtención del modelo computacional humano) para la construcción del modelo. Al realizar el alineamiento de la secuencia de aminoácidos sobre la estructura cristalográfica de *E. coli* se recortó la enzima de *Haemophilus influenzae* en 4 aminoácidos (pues es de 270 comparado con *E. coli* de 266). Se obtuvo únicamente el monómero.

**d) Construcción del modelo computacional de la probable GN6PD de *Mycoplasma capricolum* a partir del modelo cristalográfico de la enzima homóloga de *E. coli* en su conformero R.**

Obtención de coordenadas tridimensionales a partir del modelo cristalográfico de *E. coli* y arreglo estructural de aminoácidos.

Aunque la técnica empleada es la misma, esta enzima tuvo un proceso un poco diferente en su elaboración debido a que el alineamiento de secuencias mostró inserciones y deleciones en regiones

de las asas. El alineamiento nos muestra que esta enzima no presenta las regiones de la enzima relacionadas con la asociación con otros monómeros (dimerización y hexamerización) eliminando el motivo estructural del que forma parte la hélice 7, la asa siguiente y la hélice 3-10, así como también la hebra  $\beta$  G y la hélice 8. Esto provoca también un movimiento de la estructura para reemplazar la posición de la hebra beta G perdida, por la hebra  $\beta$  A, la cual toma su lugar espacial en esta posición. Para obtener las coordenadas iniciales de la hebra  $\beta$  A de *M. capricolum* que ocupa una posición diferente a la hebra  $\beta$  A de *E. coli*, se tomaron coordenadas de la hebra beta G del modelo cristalográfico para colocarla en posición adecuada. Para la colocación de esta hebra  $\beta$ , sustituyendo a la que se pierde, se quitaron los tres aminoácidos iniciales de la secuencia de *M. capricolum* (la metionina inicial, una fenilalanina y una tirosina) debido a que su ubicación resultó ser difícil de determinar, pues comprende aminoácidos hidrofóbicos que no es común que estén expuestos al solvente. Pese a esta modificación, la enzima parece conservar la estructura y el plegamiento del dominio de unión a NAD, excepto por la región equivalente a la hélice 7 de la cual carece. Para las asas modificadas se empleó una búsqueda de estructuras (InsightII.Biopolymer) conocidas en el banco de datos del programa para las asas y se eligió aquella conformación que ofreciera una menor desviación del r.m.s. (root-mean-squared, raíz promedio del cuadrado de las distancias entre los carbonos  $\alpha$  de dos estructuras) de las regiones inmediatamente antes (preflexibles) y después (posflexibles) de cada asa tomando 5 aminoácidos en cada región (ver apéndices). La región flexible (la asa) no interviene en la búsqueda de las estructuras posibles pues a esta región se le permite variar. Estas estructuras, las asas, se unen al modelo con distancias de ligadura y ángulos entre los átomos inadecuadas, pero esto se soluciona con la minimización de energía que obliga a las estructuras a tomar distancias adecuadas. El alineamiento de secuencias entre *M. capricolum* y *E. coli* nos muestra una inserción importante entre los aminoácidos 169 y 170 de *E. coli* (la inserción son los aminoácidos 165 y 166 de *M. capricolum*) en la hélice  $\alpha$  5 que forma parte de una importante zona cercana a aminoácidos que están en contacto con el sustrato. Dicha hélice en *E. coli* presenta hacia su extremo C-terminal una modificación del aminoácido 168 al 173, de una hélice  $\alpha$  a una hélice 3-10. Con la inserción de 2 aminoácidos la sección de la hélice 3-10 fue modificada a hélice  $\alpha$  en *M. capricolum*, por búsqueda de una estructura adecuada en el banco de datos del programa. También muestra una deleción de un aminoácido en la asa siguiente a la hélice 5 al cual se le realizó una búsqueda de la asa.



Cabe mencionar que las formas alternativas de esta hélice de *M. capricolum* para acomodar esta inserción fueron inestables en la dinámica de 20 picosegundos. El modelo minimizado también se le realizó una dinámica con las mismas características que las anteriores y este modelo fue tratado también como proteína monomérica.

**e) Alineamiento estructural de los modelos monoméricos con el monomero de *E. coli*.**

Los modelos fueron alineados estructuralmente con el programa "O" el cual realiza un sobreposición de las estructuras. Con el programa "O" se elige una orienta entre ambos modelos de manera inicial con el comando "lsq\_ex" y posteriormente con el comando "lsq\_im" el programa va eligiendo fragmentos de la secuencia en los cuales los carbonos  $\alpha$ , que están a menos de 0.8 Å, son incluidos. De esta forma el ciclo se repite hasta encontrar la orientación del modelo con la suma mínima de los cuadrados de las distancias entre los carbonos  $\alpha$  correspondientes entre los dos modelos (r.m.s.). Para las comparaciones se emplearon los resultados de las minimizaciones de energía, ya que en una dinámica se deforman más las moléculas.

**f) Obtención del hexámero del modelo humano y de *C. albicans* por aplicación de la simetría cristalográfica de *E. coli*.**

Para obtener el hexámero se tiene que obtener el dímero del modelo humano. Este modelo se obtuvo con el mismo método de alineamiento de estructura de insithII (ver paso anterior no.3 sobre el dímero cristalográfico de *E. coli*). Cada monómero del modelo humano se alineó con el mismo número de aminoácidos para lograr la mejor simetría del dímero. El dímero del modelo obtenido se alineó sobre el dímero cristalográfico en el programa "O" para obtener el alineamiento bajo el método automático del programa (ver seccion e para explicación del programa "O"). Con el nuevo alineamiento del dímero del modelo humano obtenido se le aplicó la celda de simetría (a125.9 Å, b125.9 Å, c223.2 Å, 90 90 120 con el grupo espacial R32) del modelo cristalográfico de *E. coli* en el programa "O". Con esto se pudo contruir todo el hexámero del modelo. Para el hexámero de *C. albicans* se realizó un alineamiento similar y se aplicó la misma celda.

### Alineamiento de las secuencias homólogas de la GlcN6P-D de *E. coli*

	1			1	
E. coli					S S S S L L L H H H
H. influenzae					R R L I P L T T T A E Q Q
Humana					R R F I P L L Q T T E Q Q
C. albicans					K R L I I I F S H N P S D
M. capricolum					R Q A I I F S N P N D
u18997				M F Y L V	K I I I F D K I E D
I103281					Q T L Q Q V E N Y T
CONSENSO				h	+ . h h lh . p . p p
	12			22	
E. coli	H H H H H H H H H H			H H H H H L L L L	L L L S S S S S L L
H. influenzae	V S K W A A R Q H I V			N R I N N A F K K P T A	D R P P F V V L G L P T
Humana	V S C W A A R Q H I R			N R I N N A F K K P T A	D R P P F V V L G L P T
C. albicans	A S E W A A K Y I R			N R I N N A F K K P T A	E K Y F F T L G L P T
M. capricolum	A A E Y L A N Y I I			A K I N S T . . . .	P R T F V L G L P T
u18997	L Q E Y C A Q L F I			D Q V K Q N . . . .	. P F A V I G F A T
I103281	A L S E R A S E Y L			L A V I R S K P . . .	. . R N A V I C L A T
CONSENSO	h . . a h A p . lh h			. . . h . . p p P . h	. + . h . lh sh lh sh T
	42			52	
E. coli	L L L L H H H H H			H H H H H L L L L	L L L L S S S S L
H. influenzae	G G T P M T T Y K A			L V E M H K A G Q V	S F K H V V T F N M
Humana	G G T P L K T Y Q E			L I R L Y Y Q A G K V	S F K H V V T F N M
C. albicans	G S T P L G C Y K K			L I E A N K Q G D L	S F K Y V K T F N M
M. capricolum	G S S P E G I Y A K L			L I E A N K Q G R V	S F K N V V T F N M
u18997	G V S P V D C Y K L			I I K H S K Q N N I	S W K N I T F N L
I103281	G A T P L L T Y H Y			L V E K I H Q Q Q V	D V S Q L T F V K L
CONSENSO	G s o P . . . Y .			lh h p . . p . . p h	p a p p h . . h N lh
	72			82	
E. coli	S S S L L L L L L L			L L L H H H H H H	H H H H H L L L L
H. influenzae	D E Y V A L L P K E E H			P E S Y Y H S F M H R	N F F D H V D I P A
Humana	D E Y V A L L P K E E H			P E S Y Y H S F M H Y	N F F D H V D I P A
C. albicans	D E Y L G F A P S D			P E S Y Y H S F M W N	N F F K H I D I H P
M. capricolum	D E F V N I S K D H			L Q S Y H Y F M Y D	K F F N H I D I P R
u18997	D E W V D L S K D H			K Q A F I K Q M K D	N L F D H L D I D P
I103281	D E I P F R G K E G			P G T C E T F L Q Q	H I V Q P L G L R E
CONSENSO	D E a h . h . . p p			E G V T I T N L R N	L F F T P A G I K E
				. p s h . . . lh . p	p lh F p . h . h .

Alineamiento de las secuencias homólogas de la GlcN6P-D de *E. coli*

	102		112		118		122	
<i>E. coli</i>	S S S S L L L L L	L L H H H H H H H	H H H H H H H H	H H H H H H H H	H H H H H H H L L L			
<i>H. influenzae</i>	E N I N L L N G N A	P D I D A E C R R Q Y	E E K I R S Y G K I	E E K I K S Y G K I	E E K I K A A G G I			
Humana	E N I N I L N G N A	D D H N A E C R R R Y	E E K I K A A G G I	E E K I K A A G G I	E E K I K A A G G I			
<i>C. albicans</i>	E N I H I L N G L A	A N I D E E C A N Y	E K K I K Q Y G R I	E K K I K Q Y G R I	E K K I K Q Y G R I			
<i>M. capricolum</i>	K N I N F L D C Q T	S D P K K E C Q R Y	E N L I R S V D G I	E N L I R S V D G I	E N L I R S V D G I			
u18997	D Q L I S F R S E E	I N . E T E C E R V	T N L I A R K G G L	T N L I A R K G G L	T N L I A R K G G L			
I103281	E N I Q K L T I D .	. . . N Y R E H	D Q K L A R E G G L	D Q K L A R E G G L	D Q K L A R E G G L			
CONSENSO	- N lh p . h p s p .	. p . p . E C p p a	p p . lh . p . G . lh	p p . lh . p . G . lh	p p . lh . p . G . lh			
	132		139	141	143	148	152	
<i>E. coli</i>	L S S S S L L L L	L L L L L L L L L	L L L L L L L L L	L L L L L L L L L	L L L L L L L L L	L L L L L L L L L	L L L L L L L L L	L L L L L L L L L
<i>H. influenzae</i>	H L F M G G V G D	G H I A F N E P A S	G H I A F N E P A S	G H I A F N E P A S	G H I A F N E P A S	G H I A F N E P A S	G H I A F N E P A S	G H I A F N E P A S
Humana	H L F M G G V G D	G H I A F N E P A S	G H I A F N E P A S	G H I A F N E P A S	G H I A F N E P A S	G H I A F N E P A S	G H I A F N E P A S	G H I A F N E P A S
Humana	E L F V G G I G P D	G H I A F N E P A S	G H I A F N E P A S	G H I A F N E P A S	G H I A F N E P A S	G H I A F N E P A S	G H I A F N E P A S	G H I A F N E P A S
<i>C. albicans</i>	D L F L G G L G P E	G H L A F N E A G S	G H L A F N E A G S	G H L A F N E A G S	G H L A F N E A G S	G H L A F N E A G S	G H L A F N E A G S	G H L A F N E A G S
<i>M. capricolum</i>	D F Q Y I S L G I N	G H M A Y N E P N T	G H M A Y N E P N T	G H M A Y N E P N T	G H M A Y N E P N T	G H M A Y N E P N T	G H M A Y N E P N T	G H M A Y N E P N T
u18997	D L C V I L G L G K N	G H L G L N E P G E	G H L G L N E P G E	G H L G L N E P G E	G H L G L N E P G E	G H L G L N E P G E	G H L G L N E P G E	G H L G L N E P G E
I103281	D L V V L G L G A D	G H F C G N L P N T	G H F C G N L P N T	G H F C G N L P N T	G H F C G N L P N T	G H F C G N L P N T	G H F C G N L P N T	G H F C G N L P N T
CONSENSO	p lh h h . G h G . p	G H lh sh lh N E P . o	G H lh sh lh N E P . o	G H lh sh lh N E P . o	G H lh sh lh N E P . o	G H lh sh lh N E P . o	G H lh sh lh N E P . o	G H lh sh lh N E P . o
	162		172		182		189	
<i>E. coli</i>	L L H H H H H H H	H H H H L L L L L	H H H H L L L L L	H H H H L L L L L	L L L L L S S S S L L	L L L L L S S S S L L	L L L L L S S S S L L	L L L L L S S S S L L
<i>H. influenzae</i>	L T H D T R V A . .	N S R F F D N D V N T	N S R F F D N D V N T	N S R F F D N D V N T	Q V P K Y A L T V G	Q V P K Y A L T V G	Q V P K Y A L T V G	Q V P K Y A L T V G
Humana	L T A Q D T L I A . .	N S R F F D N D V N T	N S R F F D N D V N T	N S R F F D N D V N T	Q V P K Y A L T V G	Q V P K Y A L T V G	Q V P K Y A L T V G	Q V P K Y A L T V G
Humana	L A M D T I L A . .	N S R F F D N D V N T	N S R F F D N D V N T	N S R F F D N D V N T	Q V P K Y A L T V G	Q V P K Y A L T V G	Q V P K Y A L T V G	Q V P K Y A L T V G
<i>C. albicans</i>	L V E S T I K A . .	N C R F F G N D E S	N C R F F G N D E S	N C R F F G N D E S	K V P K Y A L S V G	K V P K Y A L S V G	K V P K Y A L S V G	K V P K Y A L S V G
<i>M. capricolum</i>	L T K E T I L D M V	N K K K F . N S L D	N K K K F . N S L D	N K K K F . N S L D	D C P T H A M T M G	D C P T H A M T M G	D C P T H A M T M G	D C P T H A M T M G
u18997	L D A R T . . . .	Q Q H E M L K T A G	Q Q H E M L K T A G	Q Q H E M L K T A G	R P V T R G I T L G	R P V T R G I T L G	R P V T R G I T L G	R P V T R G I T L G
I103281	P I Q G E M V D I V	A H G E L G G D F S	A H G E L G G D F S	A H G E L G G D F S	L V P D S Y V T M G	L V P D S Y V T M G	L V P D S Y V T M G	L V P D S Y V T M G
CONSENSO	L . . p T . h . . .	p . + . lh . p p h p	p . + . lh . p p h p	p . + . lh . p p h p	p sh P p . h lh T h G	p sh P p . h lh T h G	p sh P p . h lh T h G	p sh P p . h lh T h G
	190		205	208				
<i>E. coli</i>	H H H H L L L L L	S S S S S L L H H H	S S S S S L L H H H	S S S S S L L H H H	H H H H H H H H L	H H H H H H H H L	H H H H H H H H L	H H H H H H H H L
<i>H. influenzae</i>	V G T L L D . A E E	V M I L V L L G S H Q K	V M I L V L L G S H Q K	V M I L V L L G S H Q K	A L A L V Q A A V E	A L A L V Q A A V E	A L A L V Q A A V E	A L A L V Q A A V E
Humana	V G T L L D . A E E	V M I L V L L G S H Q K	V M I L V L L G S H Q K	V M I L V L L G S H Q K	A L A L V Q A A V E	A L A L V Q A A V E	A L A L V Q A A V E	A L A L V Q A A V E
Humana	V G T V M D . A R E	V M I L V L L G S H Q K	V M I L V L L G S H Q K	V M I L V L L G S H Q K	A L A L V Q A A V E	A L A L V Q A A V E	A L A L V Q A A V E	A L A L V Q A A V E
<i>C. albicans</i>	I S T I L D N S D E	I A I I V L G K S K	I A I I V L G K S K	I A I I V L G K S K	Q F A L D K T V N G	Q F A L D K T V N G	Q F A L D K T V N G	Q F A L D K T V N G
<i>M. capricolum</i>	I Q T L L N W T K K	A I T V S Y G I H K	A I T V S Y G I H K	A I T V S Y G I H K	D F S Y . . . .	D F S Y . . . .	D F S Y . . . .	D F S Y . . . .
u18997	L K D I L N . A R E	V L L V T G E G K K	V L L V T G E G K K	V L L V T G E G K K	Q D A T D R F L T A	Q D A T D R F L T A	Q D A T D R F L T A	Q D A T D R F L T A
I103281	P K S I M A . A K N	L L I I V S G A G K	L L I I V S G A G K	L L I I V S G A G K	A Q A L K N V L Q G	A Q A L K N V L Q G	A Q A L K N V L Q G	A Q A L K N V L Q G
CONSENSO	h p o lh lh p . s . .	h lh h lh s . G . p K	h lh h lh s . G . p K	h lh h lh s . G . p K	. . A h p . h h p .	. . A h p . h h p .	. . A h p . h h p .	. . A h p . h h p .

### Alineamiento de las secuencias homólogas de la GlcN6P-D de *E. coli*

	219																228													238					
<i>E. coli</i>	<b>L</b>	<b>L</b>	<b>L</b>	<b>L</b>	<b>L</b>	<b>L</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>L</b>	<b>L</b>	<b>L</b>	<b>S</b>	<b>S</b>	<b>S</b>	<b>S</b>	<b>S</b>	<b>S</b>	<b>S</b>	<b>L</b>	<b>L</b>	<b>L</b>	<b>L</b>	<b>L</b>	<b>L</b>	<b>L</b>	<b>L</b>	<b>L</b>	<b>L</b>	<b>L</b>		
<i>H. influenzae</i>	C	.	V	N	H	M	W	T	I	S	C	L	Q	L	H	P	K	A	I	M	V	C	D	E	P	S	T	M	E	L	.	.	.		
Humana	S	.	I	N	H	L	W	T	V	S	A	L	Q	M	H	R	H	F	L	L	V	C	D	E	A	A	Q	Q	E	L	.	.	.		
<i>C. albicans</i>	G	.	N	N	H	M	W	T	V	S	A	F	Q	Q	H	P	R	T	V	F	V	C	D	E	D	A	T	L	E	L	.	.	.		
<i>M. capricolum</i>	K	P	N	D	P	K	Y	P	S	S	Y	L	Q	D	H	A	N	V	L	I	V	C	D	N	A	A	A	.	.	.	.	.	.		
<i>M. capricolum</i> u18997	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
I103281	K	.	V	S	T	A	I	P	A	S	F	L	W	L	H	S	N	F	I	C	L	I	N	T	.	.	.	.	.	.	.	.	.	.	
CONSENSO	P	.	V	T	E	D	V	P	A	S	V	L	Q	L	H	P	S	L	M	V	I	A	D	K	A	V	P	Q	N	.	.	.	.	.	
	h	.	.	.	.	h	a	s	h	S	h	lh	p	.	H	.	p	h	h	h	sh	sh	D	p	.	s	.	.	.	.	.	.	E	L	

	248	250									254									266															
<i>E. coli</i>	<b>L</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>		
<i>H. influenzae</i>	K	V	K	T	L	R	Y	F	N	E	L	E	A	E	N	I	K	G	L	L	E	G	A	V	A	G	T	D	.	.	.	.	.		
Humana	K	V	K	T	V	K	Y	F	K	G	L	M	L	V	H	N	K	L	V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
<i>C. albicans</i>	K	S	K	L	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
<i>M. capricolum</i>	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
<i>M. capricolum</i> u18997	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
I103281	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CONSENSO	K	V	K	T	h	p	Y	F	p	.	L	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	

Tabla 1.- Alineamiento estructural de las secuencias homólogas a la GlcN6P-D de *E. coli*. La estructura secundaria de la proteína de *E. coli* se representa por las letras en cursivas *S* (strand), *H* (helix) y *L* (loop). Arriba de esta línea el número de aminoácido de *E. coli*. En la línea del consenso se representa con letras mayúsculas en negritas los aminoácidos conservados en su totalidad. Con mayúsculas normales los aminoácidos con alta conservación. Con una "h" se representa una posición con aminoácidos hidrofóbicos de cualquier tipo (A,V,L,I,F,W,M,C,P); con una "a" los aminoácidos aromáticos (W,F,Y); con "sh" a los aminoácidos hidrofóbicos pequeños (A,V,G); una "s" representa aminoácidos pequeños en general (A,V,G,S,T,C,P); "lh" representa aminoácidos hidrofóbicos grandes (I,L,F,M,W); una "p" significa aminoácidos polares en general (N,Q,H,S,T,Y,D,E,R,K); con un "+" aminoácidos cargados positivamente (R,K,H); con un "-" aminoácidos cargados negativamente (D,E); con una "o" se indican los grupos "OH" (S,T). El alineamiento termina a 266 aminoácidos que es el no. de aminoácidos conocidos de estructura terciaria de *E. coli*.

## Resultados y discusión

### Posiciones conservadas en los alineamientos de secuencia y su función estructural.

El alineamiento múltiple realizado utilizando el programa Blast y su posterior refinamiento, junto con los estudios de estructura cristalográfica (Oliva,G.,et.al.1995) y cinéticos (Miller,G.,et.al.1987) nos permiten obtener los aminoácidos fundamentales de las enzimas homólogas a la GlcN6PD de *E. coli* para manter el plegamiento y función. Además de los aminoácidos implicados en la catálisis, si se observan las posiciones de los aminoácidos conservados y su papel estructural se les puede asignar funciones estructurales concretas en algunos casos que se detallan a continuación. El Glu 73 es fundamental para formar un enlace entre la hélice 3 y la hebra beta A' en todas las enzimas para mantener en su posición a la hebra  $\beta$  A' que contiene al Asp 72, éste también se conserva en todos por ser el aminoácido catalítico (Oliva,G.,et.al.1995). El oxígeno E1 del Glu 73 se relaciona por medio de un puente de hidrógeno con el nitrógeno del residuo 85 y el nitrógeno del residuo 74. El oxígeno E2 del Glu 73 mantiene un puente de hidrógeno con el nitrógeno del residuo 86. Los aminoácidos de las posiciones 66, 69 y 71 mantienen el carácter hidrofóbico y el plegamiento local sustentando la posición de la hebra  $\beta$  A'. También se encontró la Gly de la posición 42 como altamente conservada ya que es el ligante principal del fosfato de la glucosamina 6-fosfato. La asa de la que forma parte el residuo 42 se conserva en sus regiones claves como la Thr 44 (que en *C. albicans* y en *M. capricolum* son una Ser que conserva el hidróxilo) y la prolina 45. La posición 40 y 41 que en *E. coli* es Pro y Thr se conserva en *H. influenzae*, humana y *C. albicans*. *M. capricolum* y U18997 mantiene un par de aminoácidos que conserva el patrón estructural: Ala y Thr. En la asa catalítica se mantienen la posición 142 con Gly y 143 con una His. Además el 141 mantiene un aminoácido polar que puede estar cargado negativamente dependiendo de que aminoácido se presente en la posición 148 (que es por lo general uno cargado negativamente). Es importante mencionar que estas dos últimas posiciones (141 y 148) forman junto con la His 143 el sistema de intercambio de protones durante la abertura del anillo del sustrato. Pese a que existen cambios en el aminoácido 141 y 148 siempre se conserva en alguna de las dos posiciones, un aminoácido con carga negativa (Asp o Glu) o ambos. En el sitio activo existen otras posiciones altamente conservadas como la Gly 205, que se requiere para mantener el plegamiento local, y la Lys

208, que es uno de los ligante del fosfato. A lo largo de los alineamientos aparecen otras posiciones muy conservadas, así como posiciones que conservan el carácter hidrofóbico y que estabilizan el plegamiento (ver alineamiento tabla 1). Las comparaciones nos permiten decir que estas enzimas pertenecen a una familia de proteínas que realizan una función muy similar de desaminación e isomerización, algunas con el mismo sustrato y otras probablemente con uno similar. El plegamiento por lo tanto debe ser similar siguiendo los mismos patrones de hidrofobicidad en la mayoría de los casos y para reafirmar esto, tanto para los modelos como para las secuencias no modeladas, se hicieron predicciones de estructura secundaria con los programas SSP (Solovyev, V.V. & Salamov, A.A. 1994) y PHD (Rost & Sander. 1993, 1994. Levin, et al. 1993) para U18997, L10328 y *M. capricolum*. Los métodos empleados se complementan para predecir la topología de las secuencias, las cuales coinciden con la topología de *E. coli* en casi todas las estructuras (ver apéndices para las enzimas de *E. coli* y *M. capricolum*). Algunas no son muy exactas, pero dejan ver la presencia de una estructura homóloga. Para una comparación más efectiva se realizó el mismo método de predicción para *E. coli*, el cual demostró que hay regiones que no se pueden predecir con estos métodos como la hebra  $\beta$  del Asp catalítico por su pequeño tamaño y la hebra  $\beta$  D la cual se encuentra muy expuesta al solvente. Una predicción de estructura secundaria por alineamiento de todas las secuencias nos muestra con mas claridad la topología y con menos error (excepto los dos mencionados anteriormente, ver apéndice). Las predicciones se complementan para disminuir el error y por lo general las predicciones débiles como en las hélices presentan un perfil hidrofóbico similar al del modelo cristalográfico. En el caso de U18997 en sus primeros 30 aminoácidos mostro muy baja identidad con *E. coli* (región equivalente a la hebra  $\beta$  A y la hélice 1) y los resultados de la predicción de estructura secundaria mostraron una estructura diferente para la hebra  $\beta$  A. El perfil hidrofóbico de esta zona es poco similar.

#### **Secuencias homólogas a la GlcN6PD de *E. coli*: U18997 y L10328.**

Las secuencias que son homólogas a la glucosamina-6-fosfato desaminasa de *E. coli*, U18997 y L10328, de las cuales la primera es una secuencia de la cepa K-12 de *E. coli* localizada en la región cromosómica de 67.4 a 76.0 minutos y la segunda es una proteína hipotética de 23.3 KDa de la región intergénica TNAB-BGLB también de *E. coli* y de 213 aminoácidos. Ambas secuencias tienen una identidad baja con *E. coli*. L10328 tiene una identidad de 26.8% y U18997 de 25%. Aunque

estas dos secuencias presentaron baja identidad con *E. coli* conservan los aminoácidos principales de la catálisis (Tabla 1). Ambas secuencias mostraron cambios en alguno de los aminoácidos relacionados con el sistema de intercambio de protones, al igual que *M. capricolum*, en los tres casos se conserva la posición equivalente de la His 143. En *M. capricolum* y en U18997 la posición 148 se conserva con un Glu, pero en L10328 hay un cambio por Leu que es un cambio de aminoácido con carga a uno hidrofóbico. Sin embargo, L10328 en la posición 141 mantiene el Asp que *E. coli* presenta, y en *M. capricolum* y U18997 en la posición equivalente del aminoácido 141 presentaron un cambio idéntico para ambos de un Asp por un Asn el cual es polar pero no cargado. Estos cambios sugieren que el flujo de protones requiere únicamente dos aminoácidos para su función de abrir el anillo del azúcar (la His y un residuo ácido), esto es muy similar a lo encontrado en la estructura cristalográfica de la D-xilosa isomerasa para abrir el anillo del sustrato donde una His presenta un grupo ácido muy cercano, que se polariza para transferir un protón e iniciar la apertura del anillo (Blow, D.M., et al. 1992). El aminoácido principal de la catálisis, el Asp 72 de *E. coli* se conserva en todas las secuencias. Trabajos muy recientes muestran que U18997 probablemente forme parte de un operón involucrado en una vía tentativa en el procesamiento de N-acetilgalactosamina como una galactosamina 6-fosfato isomerasa (J. Reizer et al. 1996). L10328 se reporta como una proteína perteneciente a la familia glucosamina/galactosamina-6-fosfato isomerasas (Burland, V.D., et al. 1993). Una breve revisión de la posición de este marco de lectura abierto (ORF) en el genoma de *E. coli* nos muestra que está cercano a un operón implicado en el procesamiento de beta-glucósidos (genes *glb*) presente en diversos organismos. L10328 posiblemente forme parte de este operón como una isomerasa para el procesamiento de beta-glucósidos. Sin embargo, pese a la baja identidad mencionada al principio de esta sección, es posible realizar una comparación entre secuencias de proteínas con baja identidad usando el criterio de conservación de aminoácidos claves a lo largo de la región catalítica. El resto de la estructura conservará un patrón estructural similar que se puede observar bajo la conservación de aminoácidos hidrofóbicos del núcleo proteico. De estas dos enzimas se construyeron las tentativas topológicas basadas en predicciones de estructura secundaria mostradas en la figura 1 donde se puede apreciar que estas enzimas no son hexámeros como la glucosamina-6-fosfato desaminasa de *E. coli*. U18997 y L10328 no presentan la región de dimerización sin la cual la interacción entre los trímeros de dímeros no es posible. También las interacciones de la His 222

no las presentan y que son esenciales para generar la conexión en torno al eje de orden tres.

#### Estabilidad y calidad de los modelos.

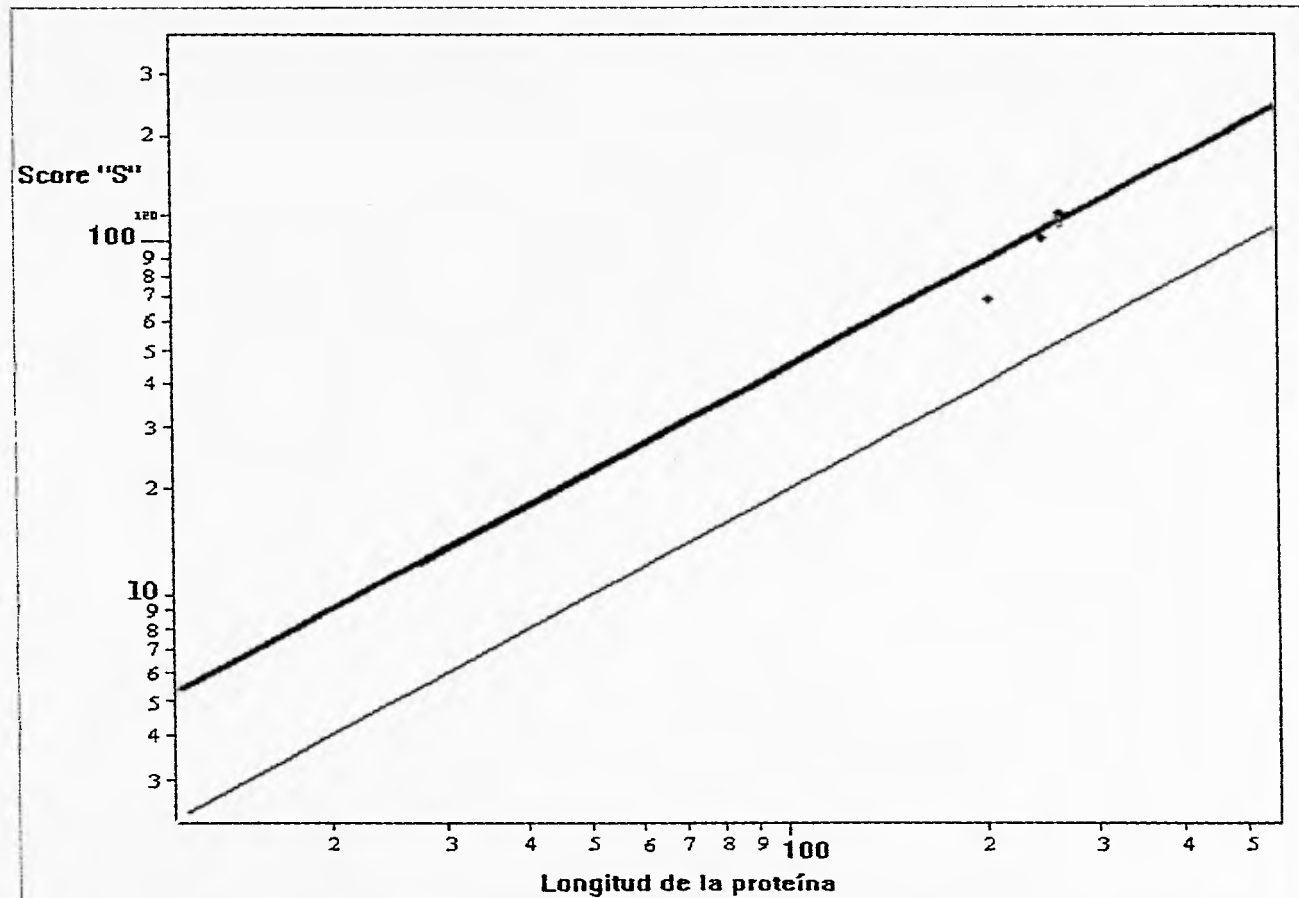
El r.m.s. de los modelos minimizados (tabla 2) mejora con el trabajo de la dinámica causando un mejor empaquetamiento de las moléculas para cada modelo, demostrando su estabilidad en 20 picosegundos. Esto significa que las proteínas mantienen un plegamiento de la misma naturaleza y/o tipo que *E. coli* y por lo tanto pertenecen estructuralmente a una familia de proteínas. En los casos de la enzima humana y de *Candida albicans* se sabe por datos experimentales de cinética enzimática que son enzimas que realizan la misma función catalítica que en *E. coli*. Con base en esto, los resultados de ambos modelos y por la alta identidad de secuencias, podemos decir que se trata de proteínas homólogas derivadas de un gen ancestral que comparten la misma función y el mismo plegamiento. La calidad de los modelos se verificó usando el programa Profiles3D (Lüthy,R.,et.al.,1992). La puntuación "S" (ver apéndice) para cada modelo se compara con la puntuación esperada para ese número de aminoácidos y la puntuación del modelo cristalográfico (Tabla 3). Las puntuaciones "S" ubican a los modelos entre los valores para estructuras cristalográficas y claramente separados de las estructuras de proteínas mal plegadas (gráfica 1), lo cual quiere decir que los modelos están refinados de manera aceptable. Los modelos de *H. influenzae*, humano y *C. albicans* están ubicados, en la gráfica 1, entre valores para proteínas cristalográficas bien refinadas, muy cerca de la línea promedio de las mismas.

Enzima	r.m.s.m	No.atomos	r.m.s.d	Identidad
<i>Haemophilus influenzae</i>	0.609	266(266)	0.609	72.6%
<i>Candida albicans</i>	1.008	247(248)	-	47%
<i>Homo sapiens</i>	1.197	266(266)	1.197	58.27%
<i>Mycoplasma capricolum</i>	0.921	197(206)	0.907	30.66%

TABLA 2.-La desviación r.m.s. de los alineamientos comparativos con *E. coli* en el programa "O" muestran que los modelos son estables al no presentar desviaciones grandes. El número de átomos es el tomado por el alineamiento. El número entre paréntesis es el no. de aminoácidos del modelo. El r.m.s.m. es de los modelos minimizados. El r.m.s.d. son los modelos de la dinámica. Se puede apreciar que el empaquetamiento de las moléculas a mejorado al acomodar los átomos internos durante la dinámica.



## Gráfica de la puntuación "S".



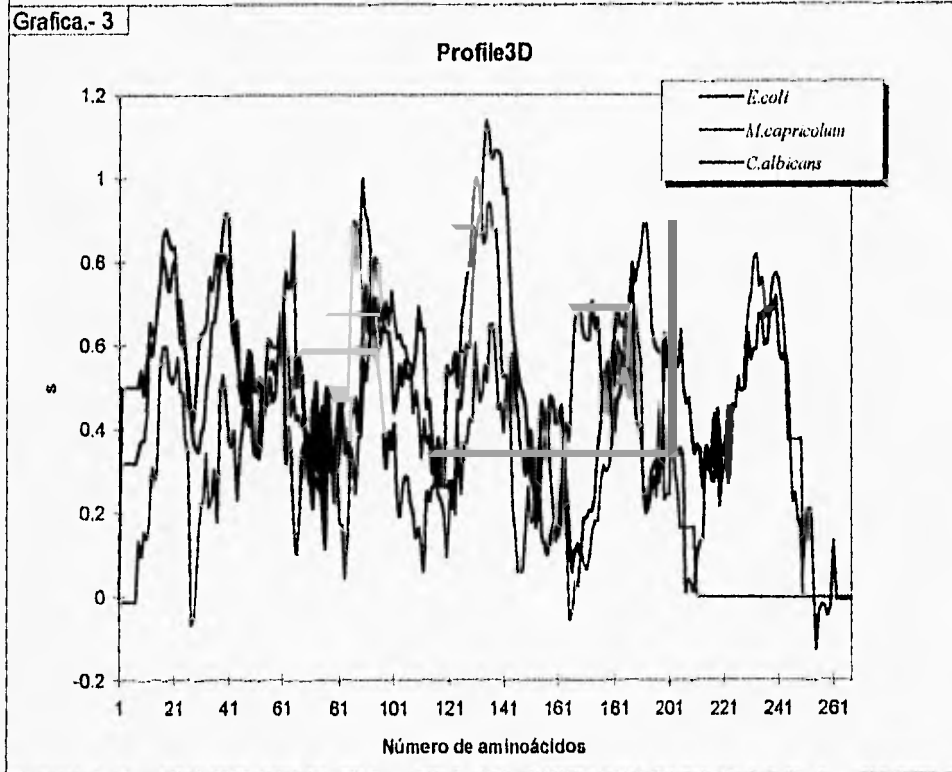
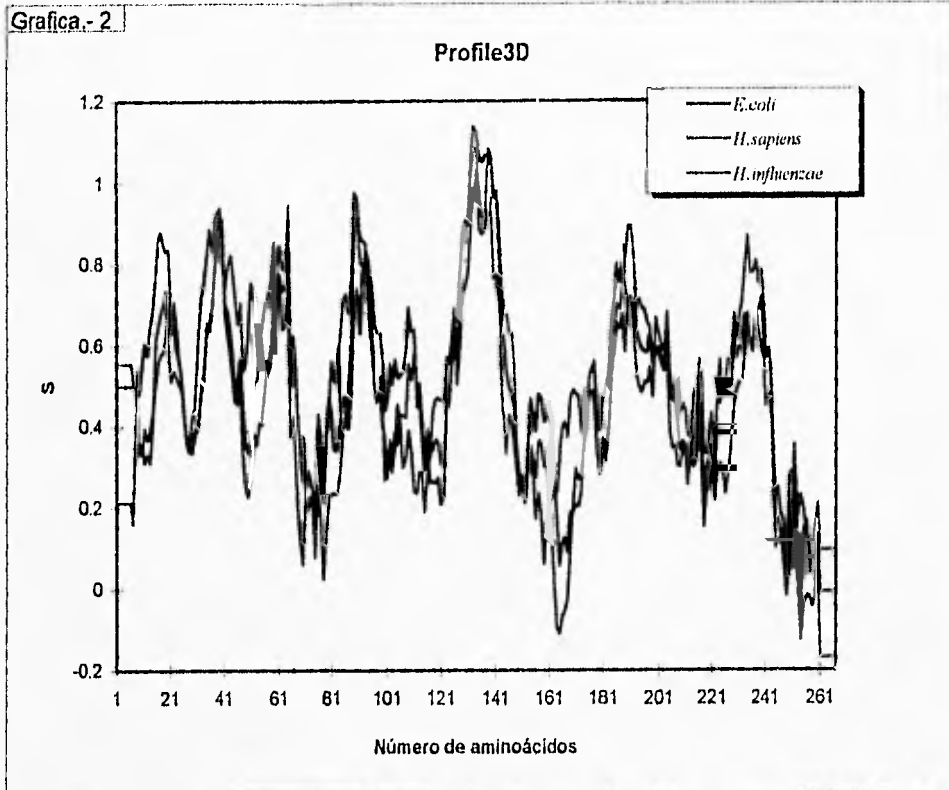
Gráfica 1.- Esta gráfica muestra el puntuación "S" obtenido con el programa Profiles3D para estructuras cristalográficas refinadas a 2 angstroms o menos y con un factor R de menos del 20%, representadas por la línea negra más ancha. La línea más delgada representa estructuras mal plegadas y algunos modelos computacionales. El modelo cristalográfico de *E. coli* está representado por el punto negro, *Haemophilus influenzae* por el punto azul, *Homo sapiens* por el punto verde y *Micoplasma capricolum* por el punto magenta.

El modelo de *M. capricolum* se localiza justo entre la línea media para buenas estructuras cristalográficas y la línea medias de estructuras mal plegadas y algunos modelos computacionales (gráfica 1). Las gráficas de la puntuación (tabla 3) para cada aminoácido muestran también la calidad del refinamiento (gráfica 2,3), normalmente la gráfica de un modelo está por debajo de la gráfica del modelo cristalográfico, en este caso se puede ver que los datos de los modelos están al mismo nivel de la gráfica del modelo cristalográfico. Dicho modelo presenta un valor mayor que el estimado para el número de aminoácidos. Se presentan algunos valores inferiores a cero en los modelos (incluyendo al cristalográfico) y esto sucede en regiones como el C-terminal, que es una región flexible. En el caso particular de *M. capricolum* los primeros 6 aminoácidos que corresponden a la zona de la hebra  $\beta$  A, que se desplaza al sitio de la hebra  $\beta$  G, presenta balores bajos. Los motivos por los cuales esta zona presenta estos valores puede deberse a que es una hebra  $\beta$  con aminoácidos hidrofóbicos y esta expuesta al solvente, o la estructura de esta zona es diferente a la propuesta por el modelo. Sin embargo, las predicciones de estructura secundaria para esta proteína proponen una hebra  $\beta$  del aminoácido 2 al 7, en un caso, y del 2 al 9 en otro. También el hecho de que se presente la hélice  $\alpha$  1 como una estructura estable similar a la de *E. coli*, sugiere que la estructura anterior sea una hebra  $\beta$ . La hélice 1 muestra valores que concuerdan con su estabilidad (a excepción de los últimos 2 aminoácidos de la hélice en el extremo C-terminal).

#### **Alineamiento estructural de los monómeros.**

El alineamiento estructural nos permite comparar los modelos del plegamiento de la cadena principal, con la estructura cristalográfica de *E. coli* y obtener un promedio de la desviación de las estructuras. Cabe mencionar que dicho alineamiento estructural incluye las asas (tabla 2). Con este alineamiento estructural se observan los cambios de estructura por lo diferentes aminoácidos de cada secuencia de aminoácidos. Se hace evidente la similitud de las estructuras viendo la cantidad de aminoácidos considerados para el alineamiento estructural (tabla 2). En el caso de la enzima humana y la de *H. influenzae* la cantidad de aminoácidos considerados es de 266 y de *C. albicans* es de 247 (sólo uno menos del tamaño de la secuencia de *C. albicans*). Para *M. capricolum* se consideraron 197 (sólo 9 menos de su tamaño). El r.m.s. de las comparaciones de los modelos con la estructura cristalográfica de *E. coli* son bajas en general siendo menor de 1.4 para los modelos minimizados y menor a 1.1 para los modelos resultantes de la dinámica.

Gráfica que muestra la calidad de los modelos.



	Puntuación "S"	"S"esperado	No. aminoácidos
<i>E. coli</i>	122.249	120.924±30	266
<i>H. influenzae</i>	114.169	120.924±30	266
<i>Homo sapiens</i>	117.630	120.924±30	266
<i>M.capricolum</i>	68.200	94.828±27	206
<i>C. albicans</i>	110.349	112.678±29	248

Tabla 3.- Puntuación "S" de Profile3D. La desviación se estimó a partir del gráfico de la figura 2 del artículo Lüthy, R. et al. 1992.

#### Hexámeros de la enzima humana.

Se ha probado experimentalmente que la enzima humana es una proteína hexámerica y alostérica igual que la de *E. coli* (Lara-Lemus, R & Calcagno, M.L. Sin publicar). La enzima mutante pro268stop presenta un comportamiento Mikaeliano perdiendo sus propiedades de proteína alostérica. La enzima no es activada por GlcAcN6P y presenta una actividad más baja que la enzima nativa y sin cooperatividad (Lara-Lemus & Calcagno, M.L. sin publicar); sin embargo, conserva su afinidad por el activador. Dado que la enzima mutante pro268stop no es alostérica y la enzima nativa sí, esto nos sugiere que la región extra de 23 aminoácidos en el C-terminal está relacionada con el control de las propiedades catalíticas de la enzima. La estructura del modelo pro268stop humano es muy similar al modelo cristalográfico de *E. coli* y algunas de las estructuras (principalmente asas) sólo se encuentran desplazadas debido a la dinámica (debido a que el modelo humano se trabajó como monómero y no se fijaron las regiones relacionadas con las interacciones entre monómeros). Una dinámica molecular puede tender a modificar localmente la estructura y en particular las estructuras flexibles como las asas debido a que la enzima nativa presenta diferentes interacciones en algunas partes de la molécula, como en las regiones que están en contacto con otros monómeros. Por lo tanto puede ser que algunas partes parezcan ser más inestables, flexibles o deformadas, desplazando la estructura debido a que los aminoácidos de esas zonas no están en el entorno adecuado. Las interacciones hexáméricas de *E. coli* (figura 2) se presentan en la enzima nativa humana conservando

las interacciones complementarias para mantener unidos a los seis monómeros. La mutante pro268stop sin activador es inestable y se desnaturaliza en monómeros inactivos, el modelo no nos permite tener una evidencia estructural pertinente para poder explicar este comportamiento.

Sin embargo, en presencia del activador el hexámero es más estable sin contribuir a la actividad enzimática. La secuencia de *Mesocricetus auratus* (Parrington, J. et al. 1996) nos muestra que la alta identidad entre ésta y la enzima humana nos permite deducir que la enzima de hamster es hexamérica y funcionalmente similar a la humana. Dicha enzima se ha reportado resientemente y sería de gran interés su estudio detallado en el futuro.

#### • **Hexamerización de la proteína de *H. influenzae*.**

El caso particular de *Haemophilus influenzae* cuya secuencia de la proteína es 72.6% idéntica a *E. coli* podemos decir que se trata de una proteína perteneciente a la misma familia. Los resultados de la dinámica son claros, al verificar que dicha enzima no se destruye ni sufre grandes cambios en la estructura durante la simulación de 20 ps de dinámica. Esta estructura es sumamente similar a la estructura cristalográfica de *E. coli* (ver tabla 1 y figura 1 y 2) y mantiene la mayoría de los aminoácidos claves en sus posiciones. Los cambios de algunas asas del modelo se pueden deber a que se trabajó como si fuera un monómero. Las diferencias de estructura se deben, en parte, a regiones que deben estar en un entorno distinto al del solvente (como las regiones de hexamerización). *Haemophilus influenzae* debe ser también un hexámero porque el modelo mantiene el plegamiento de esta región y conserva los residuos necesarios (ver figura 2) para mantener las interacciones del hexámero. La secuencia de las regiones implicadas en las interacciones intermonoméricas se conservan con una identidad de 68% y todos los aminoácidos claves están conservados, con la excepción de un residuo que mantiene el carácter hidrofóbico. Si esta proteína es hexamérica debe ser alostérica y activada por el mismo compuesto, GlcAcN6P, porque los aminoácidos claves del sitio alostérico se conservan en un 100% así como el plegamiento de la región en el modelo. La interacción de los residuos Cys 219 entre dos monómeros de dímeros distintos de la enzima de *E. coli* no se presentan en la de *H. influenzae* y, aunque es una interacción importante, la pérdida de ésta no lleva a la separación del hexámero de *E. coli* (Altamirano, M.M., et al. 1993). Por lo tanto la sustitución de una Cys por una Ser en *H. influenzae* no destruye al hexámero que forma la enzima.

## Interacciones intermonoméricas del trímero de las enzimas hexaméricas.

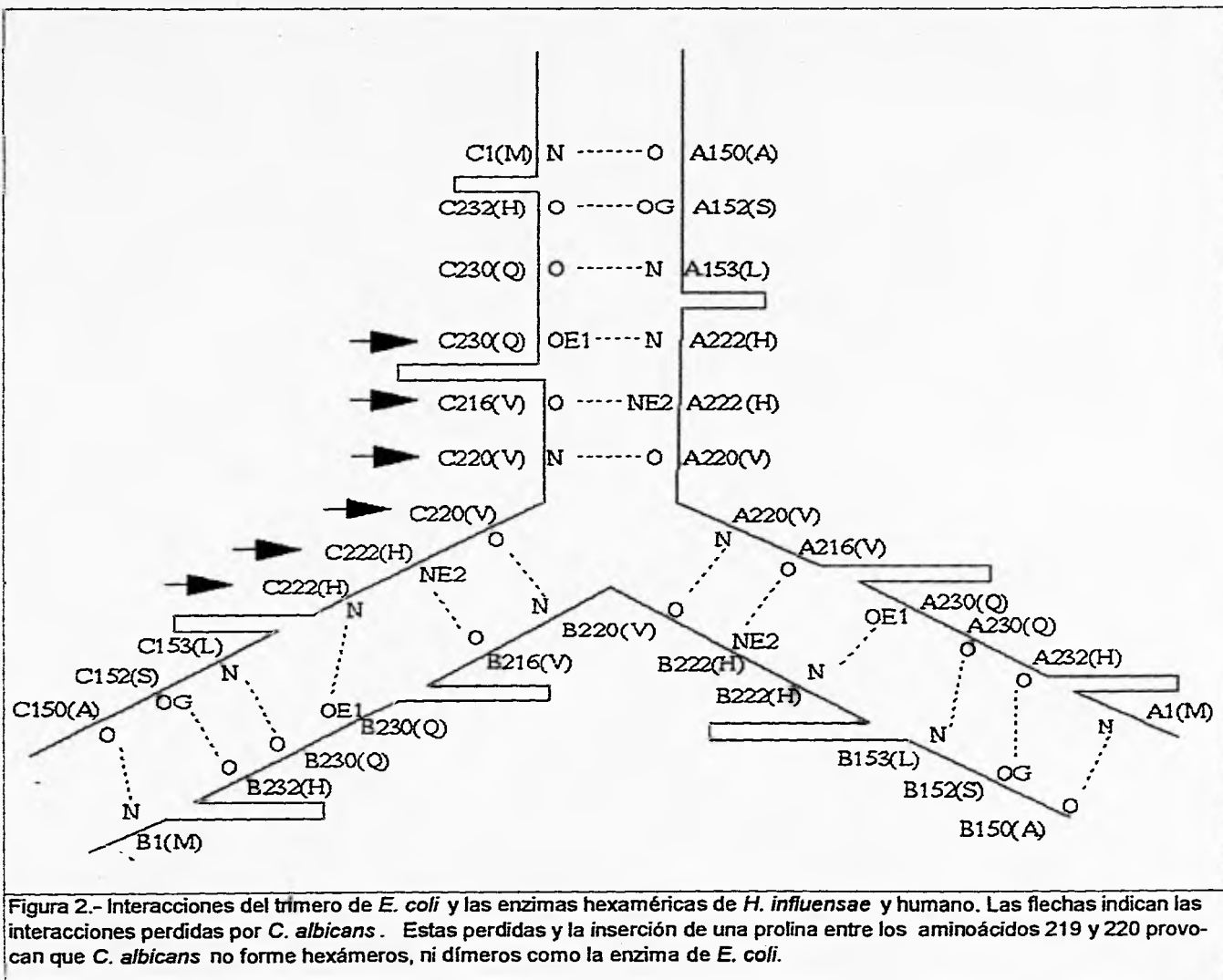


Figura 2.- Interacciones del trímero de *E. coli* y las enzimas hexaméricas de *H. influenzae* y humano. Las flechas indican las interacciones perdidas por *C. albicans*. Estas pérdidas y la inserción de una prolina entre los aminoácidos 219 y 220 provocan que *C. albicans* no forme hexámeros, ni dímeros como la enzima de *E. coli*.

**La GlcN6PD de *Candida albicans* no es hexámerica.**

El modelo de *Candida albicans*, que fue realizado por Richard Garratt (comunicación personal) presenta un plegamiento también muy similar al de *E. coli*, aunque de esta enzima no se ha podido comprobar su condición de monómero o de dímero (no es hexámero) y se sabe que no responde al activador GlcAcN6P, siendo una enzima michaeliana (Datta,A.,et.al.1993). El sitio alostérico de las enzimas hexáméricas está en la interfase entre dos monómeros. Al no ser hexámero, esta enzima no posee sitio de unión para el activador. Más aún, el modelo muestra, que la región del monómero implicado en la interacción con el activador, aunque muy similar en su plegamiento, carece de algunos aminoácidos claves para poder enlazar el activador alostérico: la posición equivalente a la Lys 160 es sustituida por una Val, perdiendo una carga necesaria para el enlace del fosfato del activador. También carece del aminoácido 254, el cual es importante para poder estabilizar el cambio alostérico mediante un enlace de hidrógeno intermonomérico. La similitud entre la enzima de *E. coli* y la enzima de *Candida albicans* en la región de dimerización es baja, además de presentar una pérdida de 18 aminoácidos en el C-terminal provocando que la región para la dimerización quede cortada por la mitad (Tabla 4).

<i>E. coli</i>	EPSTMELKMKTLRYFNEL
<i>Haemophilus influenzae</i>	EAAQQELKVKTVKYFTEL
<i>Homo sapiens</i>	EDATLELKVKTVKYFKGL
<i>Candida albicans</i>	NAAA-GLKSKL

Tabla 4.- Región de dimerización (241-258) entre las enzimas de *E. coli*, *H. influenzae* y humana que son hexámeros y la región correspondiente de *Candida albicans* que no dimeriza, debido a que no presenta suficientes interacciones (corresponde a su C-terminal).

<i>E.coli</i>	GSQKALALQAAVEGC-VNHMWTISCLQLHP
<i>Haemophilus influenza</i>	GHQKALAVQAAVEGS-INHLWTVSALQMHR
<i>Humana</i>	GAHKAFALYKAIIEEG-VNHMWTVSFAFQQHP
<i>Candida albicans</i>	GKSKQFALDKTVNGKPNPKYPSSYLQDHA

Tabla 5.- Zona de contacto entre trímeros (205-233) de las tres enzimas hexáméricas y la región homóloga de *Candida albicans* que no es un hexámero. La inserción entre el residuo 219 y 220 incapacita a *C. albicans* para formar trímeros.

En *Candida albicans* con respecto a *E. coli* sólo se conservan los aminoácidos de las posiciones 247, 248 y 250. La importancia de estos aminoácidos radica en que el puente salino entre el Glu 246 del oxígeno E2 de un monómero, con el nitrógeno Z de la Lys 250 en la enzima de *E. coli* parece ser clave para la estabilidad del contacto entre los dímeros del hexámero mientras que en *C. albicans* no existe. Al perderse este enlace, también se pierde una interacción del oxígeno E1 del Glu 246 con el nitrógeno Z de la Lys 248. La mutación de la Val 249 por Ser en *C. albicans* provoca la pérdida de una interacción hidrofóbica importante en esta región. Por estas mutaciones mencionadas, la dimerización en *C. albicans* no se presenta de la forma en que dimerizan las proteínas que forman hexámeros. *C. albicans* presentan una baja similitud con las secuencias de aminoácidos que forman la región del plegamiento que enlaza a los trímeros en las enzimas alostéricas (Tabla 5). El modelo minimizado nos muestra que el plegamiento se conserva, pero las mutaciones en este sitio en *C. albicans* provocan la pérdida de la capacidad de interactuar con otros monómeros (figura 2) por lo tanto, la posibilidad de formar un trímero no es posible. Esta modificación, posiblemente sea posterior a la existencia de la proteína hexamérica, ya que si se pensara que la enzima de *Candida albicans* nunca hubiese sido un hexámero, la modificación de la hélice 7 (equivalente a la hélice 5 en el dominio de unión a NAD) no se habría realizado en la enzima (figura 1). Las modificaciones en la enzima de *E. coli* (hélice  $\alpha$ -asa-hélice3-10) en la hélice 7 le permiten la formación del hexámero; por lo tanto, *Candida albicans* no tendría que presentar este motivo estructural que se mantiene debido a las prolinas 225, 222 y la insertada entre los residuos 219 y 220. La inserción de una Pro entre los aminoácidos 219 y 220 en *C. albicans* causa que se pierdan tres enlaces fundamentales en la formación de un hexámero, entre los nitrógenos y los oxígenos de los aminoácidos 220 de la cadena principal de los tres monómeros implicados. La interacción es la siguiente: un monómero A se conecta con un puente de hidrógeno de su oxígeno con el nitrógeno del monómero C, el monómero C enlaza su oxígeno con el nitrógeno del monómero B y este último enlaza su oxígeno con el nitrógeno de A. La pérdida de esta interacción se debe a un cambio en la estructura de la cadena principal causado por la inserción. El modelo del hexámero de *C. albicans* demuestra que la formación de un trímero esta obstaculizada por la inserción de la Pro, entre los residuos 219 y 220, los residuos de Pro de cada uno de los monómeros implicados quedan superpuestas imposibilitando una interacción adecuada. Otra pérdida importante en esta región, para la hexamerización de *C. albicans* es el cambio de la His



222 por una Asn que no mantiene el puente de hidrógeno entre el nitrógeno E2 de la His y el oxígeno de la Val 216 de la cadena principal. También se pierde el puente de hidrógeno del nitrógeno de la cadena principal del residuo 222 con el oxígeno E1 del residuo Gln 230 por cambios en los ángulos diedrales. Las interacciones restantes no son suficientes para mantener el enlace de un trimero. Más aún sin los dímeros la interacción entre trimeros no es posible. Esto se ha confirmado con trabajos de *E. coli* mutadas en la región de dimerización que no han logrado formar enzimas triméricas (Calcagno, M. y colaboradores, comunicación personal). Los cambios en ambas regiones imposibilitan a la enzima de *C. albicans* para formar un hexámero. Esta modificación posiblemente sea un cambio adaptativo que lleva a una proteína monomérica extracelular y que puede ser debida a que esta enzima se requiere para el proceso de infección en cepas virulentas de *C. albicans* (Nataranja, K. & Datta, A. 1993).

#### **Monómero de la enzima de *M. capricolum*.**

El modelo de la enzima de *Mycoplasma capricolum* es estable con este plegamiento, pese a los cambios y regiones de las que carece esta enzima (figura 1) las que están relacionadas con la asociación con otros monómeros (dimerización y hexámerización). Además se elimina el motivo estructural hélice  $\alpha$  7/asa/hélice 3-10, la hebra  $\beta$  G y la hélice 8 (figura 3 y tabla 1). Dicha proteína al carecer, de estas regiones, probablemente sea monomérica y sin activación alostérica. Una descripción del modelo nos puede dar una mejor idea de cómo se estructura la proteína. En general, la geometría de la estructura  $\alpha/\beta$  de hoja abierta se conserva y sigue un patrón muy similar a la glucosamina-6-fosfato desaminasa de *E. coli*. El plegamiento se puede describir también como un dominio de unión a NAD (figura 1), modificado en algunas regiones al igual que en *E. coli*, pero con algunas otras diferentes. Se propone a partir del modelo de la proteína de *M. capricolum* que se conservan 6 hebras  $\beta$  paralelas en la hoja central y las 3 hebras  $\beta$  de la hoja  $\beta$  antiparalela. La hoja  $\beta$  central esta rodeada por 6 hélices  $\alpha$ , conservándose de la hélice 1 a la hélice 6. Las hélices 7 y 8 no están presentes. La hebra  $\beta$  A en *E. coli* forma parte de una de las modificaciones (hebra  $\beta$  A / hélice  $\alpha$  1) así como en *M. capricolum*, ya que en ésta el dominio de unión a NAD pierde la hebra  $\beta$  G quedando en su posición espacial la hebra  $\beta$  A. Esta proteína conserva el plegamiento de la región alostérica y presenta una mutación en el residuo 155, equivalente al 160 de *E. coli*, de una lisina por una valina perdiéndose así una carga importante para la fijación del fosfato del activador

## Alineamiento estructural de los modelos computacionales con el modelo cristalográfico de la GlcN6P-D de *E. coli*.



Figura 3.- El alineamiento de estructura de los modelos contra el modelo cristalográfico de la glucosamina 6-fosfato desaminasa de *E. coli* (en amarillo). En esta imagen se pueden apreciar las modificaciones de *M. capricolum* (en magenta) y la ausencia de la región de Hexamerización. La hebra beta G es remplazada en su posición espacial por la hebra beta A. La enzima termina en el aminoácido 209. También se aprecia donde termina *C. albicans* (en rojo) en el aminoácido 248, dieciocho aminoácidos antes que *E. coli*. *M. capricolum* y *C. albicans* no son hexaméricas. El modelo humano (en verde) y de *H. influenzae* (en azul) presentan todas las regiones de hexamerización. Por lo tanto deben ser un hexámeros.

## Canal de entrada del amonio en *E. coli*.

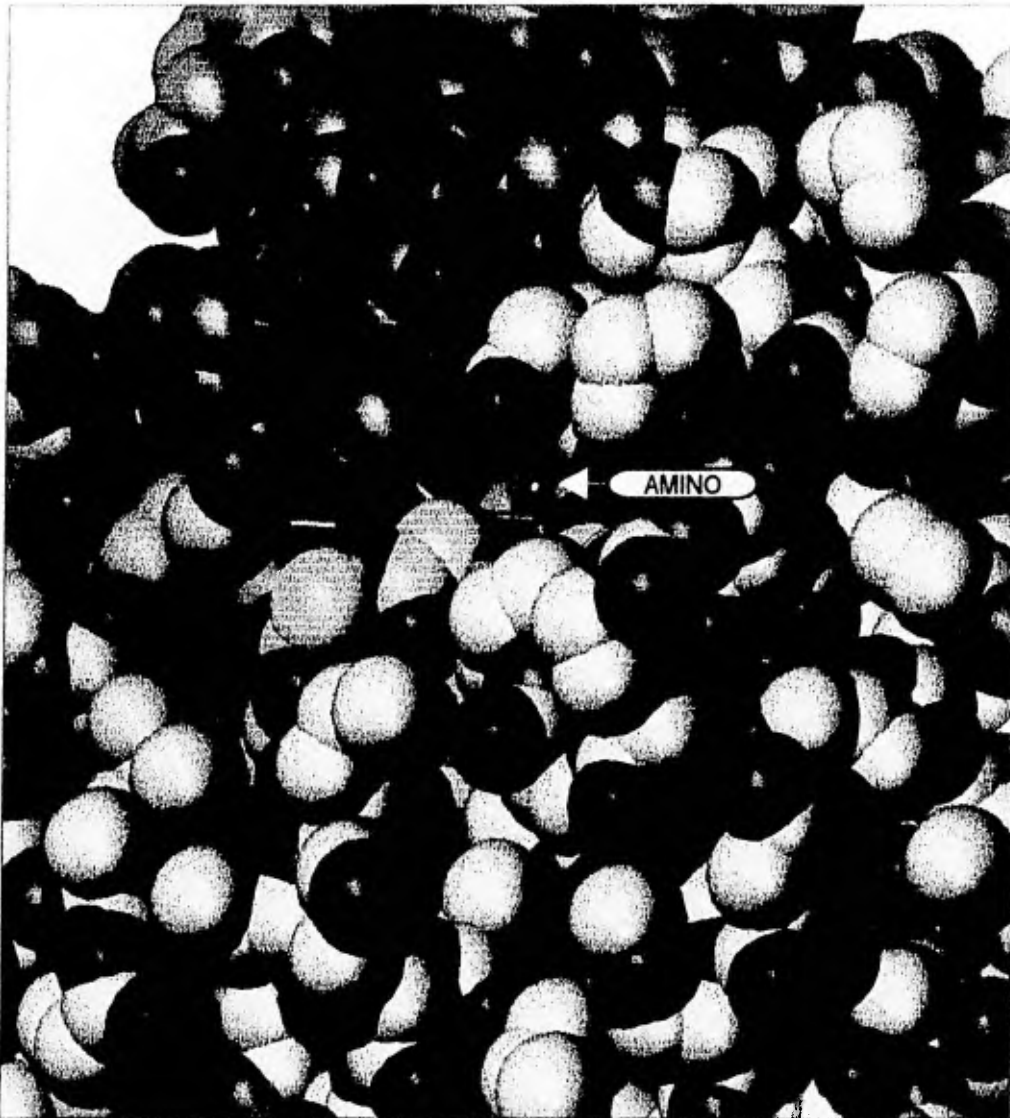


figura 4.- El canal de entrada del amonio al sitio activo en la GlcN6PD de *E. coli* da directo al grupo amina del inhibidor. El inhibidor esta representado en "Ball-stick" (los entaces entre átomos son cilindros y los átomos se representan como esferas) de color verde-limón. El nitrógeno es blanco y el resto de la proteína se representa por esferas de los radios atómicos: en verde la estructura que aísla al substrato (betaC'/hélice5/betaB'). En azul el loop de ligadura del fosfato (aminoácidos 40-45). En amarillo el resto de la proteína. En azul oscuro se representan todos los nitrógenos y en rojo todos los oxígenos.

alostérico (GlcAcN6P). Además, carece de la posición del N-terminal en el mismo sitio que *E. coli*. Éste se desplaza por el cambio de posición de la capa  $\beta$  A, así la regulación alostérica como se conoce en *E. coli* no es posible. Pese a que no se han incluido en el modelo los tres aminoácidos iniciales, por la dificultad de su colocación al no ser aminoácidos homólogos a *E. coli*, es probable que éstos formen un tipo de giro (no modelado) similar a la asa que precede a la capa  $\beta$  G de *E. coli* y que se alojen a un costado de la hebra  $\beta$  A entre la hélice 6, la hebra  $\beta$  E y parte de la asa catalítica, ya que *M. capricolum* presenta un parche hidrófobo en dicha región. La región mencionada es el lugar que ocupa la hélice 3-10 de *E. coli*.

Se a dicho que el modelo es estable, pero existen partes de la molécula que pueden ser mas inestables que otras, la revisión llevada a cabo por el programa Profile3D nos muestra que la hebra  $\beta$  A es una zona mas inestable que el resto de la molécula en sus primeros 6 aminoácidos y esto es debido a que la estructura propuesta está más expuesta al solvente y tiende a desestabilizarse. Esta hebra sustituye a la hebra  $\beta$  G que forma parte del dominio de unión a NAD y es probable que los tres aminoácidos adicionales del C-terminal puedan cumplir una función estabilizadora de la región.

#### **Entrada del amonio.**

Durante las observaciones de la estructura cristalográfica del sitio activo de la enzima de *E. coli* se descubrió que dicho sitio activo, además de presentar la entrada para el sustrato se presenta una segunda entrada mucho mas pequeña localizada a un lado de la entrada del anterior (figura 4). Dicha entrada es un canal formado entre el C-terminal de la hélice 5, la asa que sigue a la hélice 5, el N-terminal de la asa comprendida entre la hebra  $\beta$  B y la hélice 2, además de parte de la hélice 3 y la asa que le precede a esta hélice. La segunda entrada se localiza entre los residuos Thr41, Gly42, Tyr85, Phe88 y Phe 173, adicionalmente, frente a esta entrada se localiza el Glu 83. En el modelo cristalográfico la entrada está muy bien definida y da directo a lo que es el nitrógeno del inhibidor, que es la posición aproximada de ligadura de la amina del sustrato (glucosamina-6-fosfato). La entrada se mantiene abierta aún en presencia del inhibidor y es del tamaño justo para la entrada y salida de una molécula de agua o de un ion amonio. Esta entrada al sitio activo se cree podría ser la entrada y salida del amonio en la reacción catalítica tanto normal como reversa (Miller, G., et al. 1987). El analisis anterior permite comprender, estructuralmente, el mecanismo cinético de intercambio rápido al azar propuesto, que implica la posibilidad de ligadura de amonio y fructosa-6-fosfato en

cualquier orden y como fenómenos independientes (datos sin publicar de M.L. Calcagno). Dicho mecanismo puede deberse a que la fructosa-6-fosfato y el amonio tienen entradas distintas al sitio de catálisis. La posición del Glu 83 probablemente funcione como atractor del amonio hacia la entrada en espera de la fructosa-6-fosfato. La enzima humana y de *Haemophilus influenza* conserva los aminoácidos que forman esta entrada, por lo que podemos suponer que es funcionalmente parecida a la de *E. coli*. Para *Candida albicans* sólo existe una mutación y es en el Glu 83 por una Gln, este cambio es de carga e independiente de los que forman la entrada del amonio y probablemente pierda la capacidad de atraer al amonio. La proteína de *Mycoplasma capricolum* presenta varios cambios en la región y solo se conservan los residuos 41 y 42, éstos dan la posibilidad de una abertura mayor a esta entrada, pues las posiciones equivalente al Glu 83, Phe 88 y la Phe 173 cambian por Gln los dos primeros y por Lys el último, y al ser más pequeños, dejarían una abertura un poco mayor, pero con aminoácidos polares (figura 5). La mutación 173 junto con la 43 (de Gly a Val) sugieren cambios de especificidad para el sustrato en esta región. Una carga positiva como el de las Lys 171, 172 y 173 puede tener contactos con el  $\text{PO}_3^-$  del sustrato y también sugiere que el sustrato no posea un grupo amino que puede tener carga positiva. La pérdida de la carga del residuo 83 puede deberse a esto mismo.

#### **Sitio alostérico de la enzima humana.**

El sitio alostérico conserva los aminoácidos que ligan al fosfato del activador. Sin embargo, la existencia de mutaciones periféricas podrían ser responsables de los cambios observados en la enzima humana. Tales mutaciones son: Ala 150 por Gly, Ala 154 por Val, Ile 159 por Val, Gly 257 por Glu, Asn 256 por Lys, Ala 260 por Leu y Lys 234 por Arg. Observaciones hechas sobre el modelo del hexámero humano de la mutante pro268stop muestran que la región extra del C-terminal pueden también estar implicada en la modulación alostérica y ser responsable de la cinética observada en la enzima nativa. La mutante pro268stop sin esta región pierde sus propiedades alostéricas, lo cual indica que es una zona clave para la regulación alostérica. Las mutaciones periféricas al sitio alostérico, en el modelo de la mutante pro268stop, no interfieren en la ligadura del activador y las más cercanas mantienen el tipo de aminoácido. Existe una doble mutación que evita una interacción muy cercana al sitio alostérico, entre el residuo Arg 2, que cambia a Lys y el glutámico 259 que cambia a una Met. La pérdida de la carga por la mutación del glutámico provoca que se pierda un

## Canal de entrada al amonio del inhibidor en *M. capricolum*.

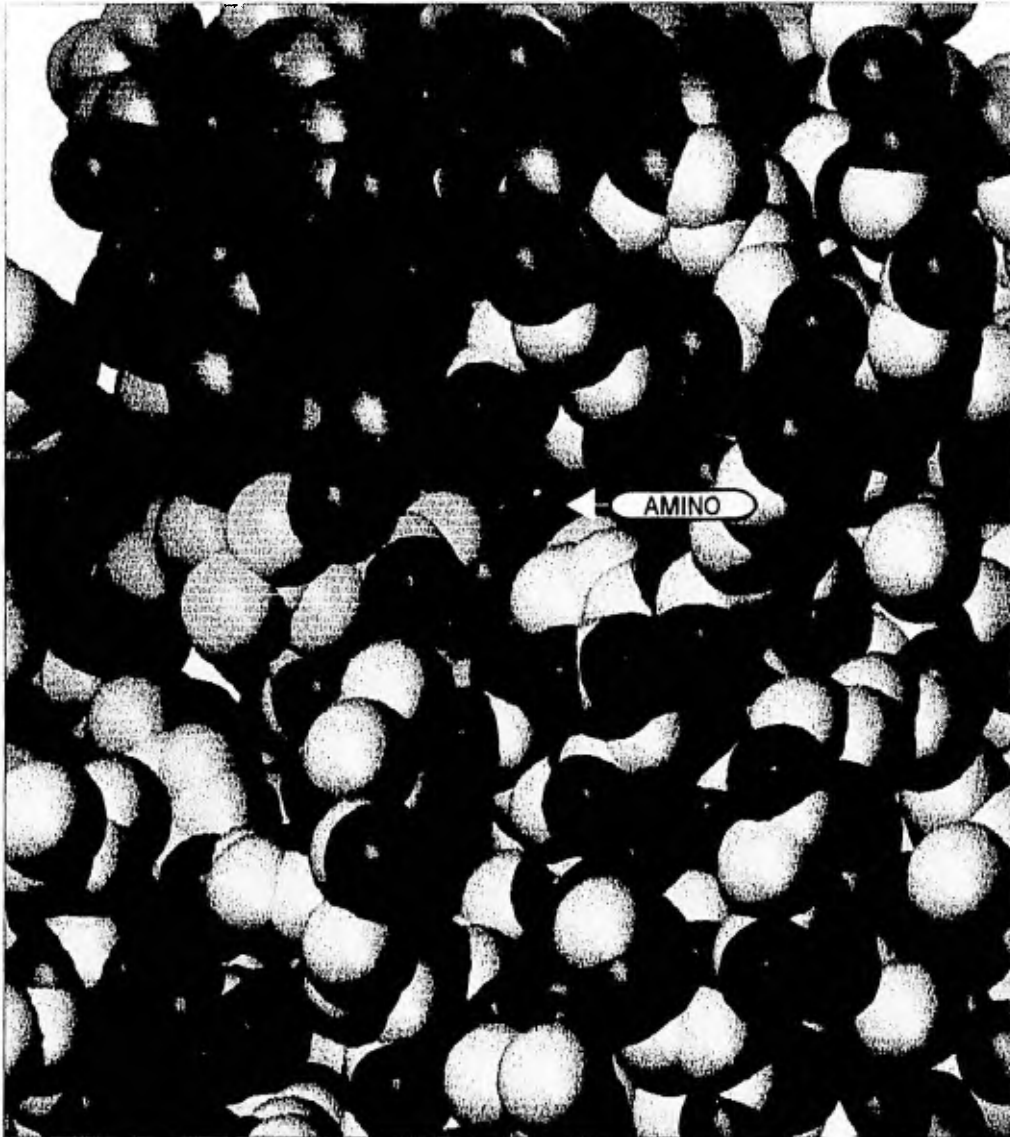


Figura 5.-El canal que conduce al amonio al sitio activo en la GlcN6P-D de *E. coli*, en *M. capricolum* esta bien formado. El compuesto que pueda entrar o salir por este canal en *M. capricolum* no se conoce. Pero se puede predecir que esta enzima es una isomerasa que probablemente lleve a cabo una reacción similar a la desaminación. El amonio del inhibidor es claramente visible (en color blanco) y el resto del inhibidor en verde-limón. El patrón de colores es el mismo que el de la figura 4.

Sitio activo de la GlcN6PD con los ligantes del sustrato mas importantes.

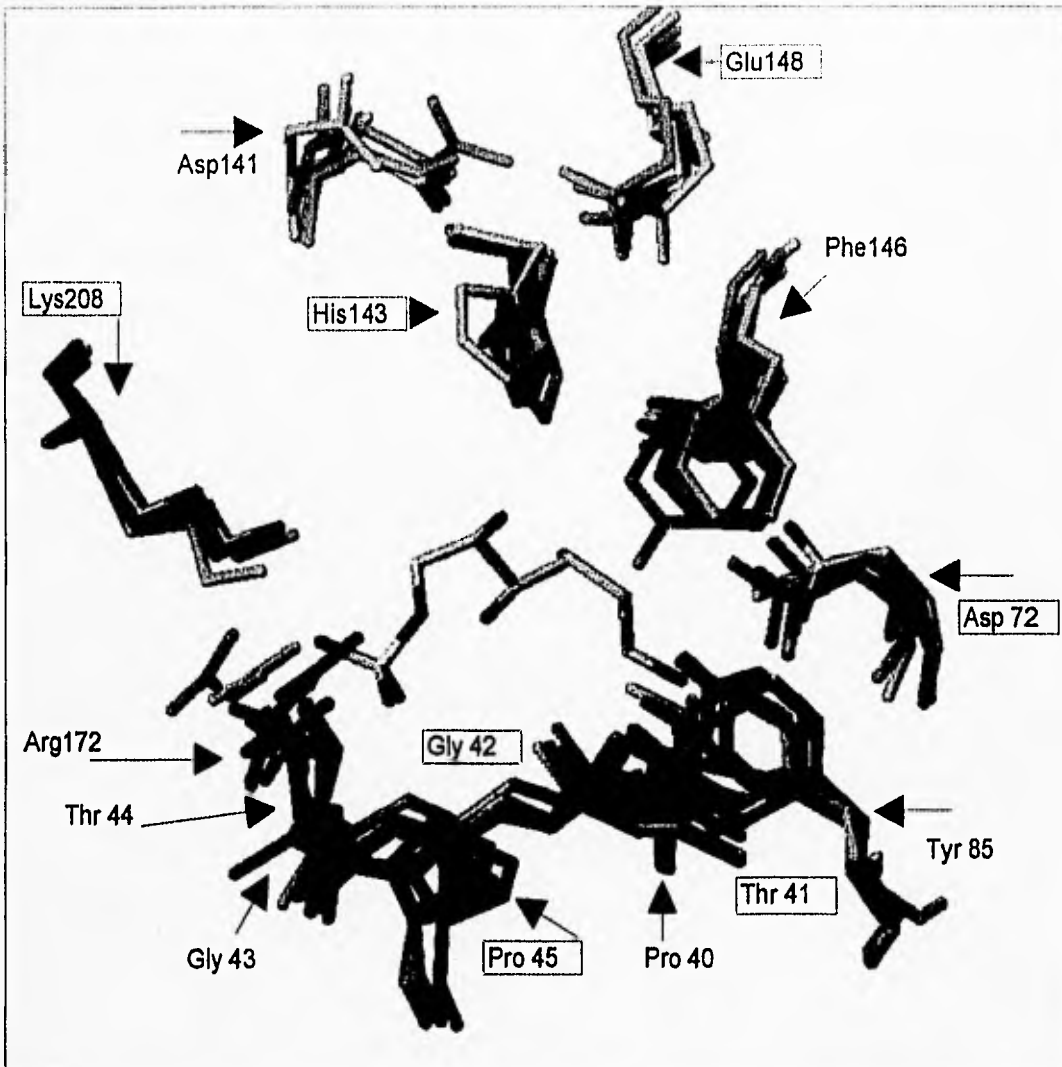


Figura 6.- Sitio activo de los modelos alineados estructuralmente contra la estructura cristalográfica de *E. coli*. Los letreros de los aminoácidos en un rectángulo son posiciones absolutamente conservadas. Los letreros que no están dentro de un rectángulo son aminoácidos que tienen variación. En amarillo se encuentra *E. coli*, en verde la enzima humana, en azul *H. influenzae*, en rojo *C. albicans* y en magenta *M. capricolum*.

punto salino entre el residuo 2 y el 259. Así la lys 2 queda libre en las cercanías del sitio alostérico y la carga total del sitio alostérico varía debido a la presencia de una carga positiva extra.

#### **Sitio activo de la enzima humana.**

El sitio activo (figura 6) de la proteína humana tiene únicamente un cambio de importancia: la posición 43 donde cambia una Gly por una Ser ubicado en la asa de unión del fosfato del sustrato. Los ángulos de Ramachandra para esta Ser están permitidos en esta posición al sustituir una Gly. El cambio puede propiciar una nueva interacción por medio de un puente de hidrógeno con el fosfato del sustrato. Al parecer, también está implicado con la estabilización de la estructura formada por las hebras B', C' y la hélice 5' del sitio activo en su conformación cerrada. La interacción es a partir de un puente de hidrógeno entre el O $\gamma$  de la Ser 43 y la Arg 172 ubicado en el extremo C-terminal de la hélice 5. Una de las mutaciones más importantes en la enzima humana está en la posición 121, la cual cambia de tirosina por fenilalanina, este aminoácido está unido por un puente de H al O-70 de la cadena principal y estabiliza la región que contiene al aminoácido catalítico Asp 72. Esta mutación sigue conservando la forma de la asa pero pierde un enlace de hidrógeno que se establece con dicha asa. Probablemente la ausencia de éste sea la causa principal responsable de las variaciones de la catálisis en la enzima humana. La metionina 71, en la enzima humana, aunque conservada presenta un desplazamiento debido a que se presenta una mutación en el residuo 193 donde los cambios son de una Leu por una Val, esto provoca el desplazamiento de la fenilalanina 134 que a su vez provoca el desplazamiento del residuo 71, al dejar un espacio por el cual éste puede desplazarse. El comportamiento de la catálisis puede también estar implicado con este movimiento, ya que genera un aumento de la cavidad del sitio activo en las vecindades del C1. Las diferencias cinéticas entre la mutante pro268stop y la enzima nativa son grandes, la mutante pasa de un sistema "v" puro con trazas de cooperatividad a un sistema mikaeliano.

#### **Sitio catalítico de la enzima de *C. albicans*.**

En la Posición 43 (figura 6) de *Candida albicans* también se presenta (al igual que la enzima humana) una Ser que probablemente funciona sólo como un nuevo ligante del fosfato y como un estabilizador de la tapa (estructura hebra  $\beta$  B', hélice 5, hebra  $\beta$  C') en su conformación cerrada al estar en un rotámero adecuado para formar un puente de hidrógeno con la Lys 172. En el modelo se presenta dicha interacción. También presenta un cambio en la posición 141 de Asp por Glu, con el



cambio se mantiene la carga negativa necesaria para el sistema de intercambio de protones. *C. albicans* presenta un cambio en la posición 193 de una Leu por Ile, que es un aminoácido más grande y que mantiene en su posición a la Phe 134 evitando el desplazamiento del aminoácido 71.

#### **Sitio catalítico de la enzima de *M. capricolum*.**

Esta enzima parece llevar a cabo el mismo tipo de reacción que la enzima de *E. coli* pero posiblemente con otro sustrato. Se llegó a esta conclusión analizando los posibles efectos de los cambios del sitio activo (figura 6): a) se conservan los tres aminoácidos considerados básicos para llevar a cabo la apertura del anillo. Esto son: Glu 72, His 143 y Glu 148 (Blow, M.D. et al. 1992); b) la Met 71 la cual cambia por una Leu, esta posición define una parte del tamaño de la cavidad del sitio activo en las vecindades del C1 del sustrato de la enzima de *E. coli*, este cambio parece provocar que el sitio activo sea más grande en las proximidades del C1 del inhibidor en la enzima de *M. capricolum*. El cambio del residuo Gly 137 por Ser, puede ser un nuevo ligante del sustrato por medio de un puente de hidrógeno, pues se encuentra muy cerca del O3 del inhibidor. La proximidad entre el fosfato del inhibidor y el residuo permite suponer que el sustrato tiene una conformación algo distinta. La Tyr 85, que interacciona con el amino del inhibidor, cambia a Phe, pero el puente de hidrógeno entre el grupo OH de la Tyr 85 y el grupo amino es sustituido por el puente de hidrógeno con el OH de otro cambio el en aminoácido 146 de Phe a Tyr (figura 5). El residuo 43 cambia de Gly a Val, esta posición es importante a nivel del fosfato del sustrato, la Gly es una de los ligantes a través de su nitrógeno de la cadena principal y el cambio por Val no cambia los ángulos de Ramachandra drásticamente y mantiene la conformación. Este residuo parece que interactúa con el 172, que sufre un cambio que mantiene la carga, (Arg por Lys) manteniendo la estructura de la tapa abierta y de esa forma evitando se cierre al sustrato. Estos cambios son complementarios, ya que el volumen molecular que se pierde con la Arg se gana con la Val, manteniendo la estructura que aísla al sustrato durante la reacción. Con el cambio del residuo 172 se mantiene la carga y por lo tanto la ligadura del fosfato. La estabilidad del loop 40-45, parece estar dada por la Pro 45, la cual se conserva en todas las enzimas. La posición 44 donde se mantiene el grupo OH en las 5 especies (Thr y Ser) parece ser una posición clave para unir el sustrato y en los casos de las enzimas caracterizadas este residuo liga al amonio. La Gly 42 y la Thr 41 también se conservan en las 5 especies. El cambio de la Pro 40 por una Ala en *M. capricolum* parece no influir en la conformación de esta asa.

### **Caracterización del plegamiento del dominio de unión a NAD.**

La GlcN6PD de *E. coli* tiene regiones estructural y funcionalmente homólogas con los dominios de unión a NAD. La GlcN6PD presenta, como los dominios de unión a NAD, dos motivos estructurales de unión a mononucleótido o "Rossmann Folds". De estos dos motivos estructurales en las enzimas que unen NAD, uno es el encargado de unir el pirofosfato del NAD. En la GlcN6PD uno de los dominios de unión a mononucleótido también ejerce una función similar: unir el fosfato de la glucosamina. El segundo motivo de unión a mononucleótido (en las enzimas que unen NAD) funciona como ligante de la segunda ribosa y su asa no interviene en la unión del pirofosfato. En la GlcN6PD no existe una segunda zona de unión para glucosa y su asa está modificada, presentando ahí la inserción del loop catalítico (residuos del 137 al 157). El plegamiento del segundo motivo de unión a mononucleótido presenta cierta homología con la asa del motivo de unión a mononucleótido que une el NAD y se encuentra más conservado que la asa de unión del fosfato de la GlcN6PD de *E. coli* y sus enzimas homólogas. Esto sugiere que la presión de selección sobre la asa de ligadura fue mayor y provocó más cambios para poder unir azúcares fosfatados. El alineamiento estructural de la asa de ligadura del fosfato de la GlcN6PD de *E. coli* con la asa del motivo de unión a mononucleótido de la alcohol deshidrogenasa de hígado de caballo permite acomodar la secuencia de *E. coli* y de sus secuencias homólogas en la secuencia "fingerprint" (de huella dactilar) (Wierenga, R.K., et al., 1986) para el reconocimiento de motivos estructurales  $\beta$ - $\alpha$ - $\beta$  que unen NAD, los cuales comparten el mismo motivo estructural (Rossmann, M.G., et al., 1974). El alineamiento (Tabla 6) del "fingerprint" para motivos estructurales  $\beta$ - $\alpha$ - $\beta$  que unen NAD y proteínas que lo presentan, con el alineamiento para la glucosamina-6-fosfato desaminasa y sus homólogas nos muestra claramente que éstas forman una subfamilia dentro de la familia de proteínas que presentan un plegamiento de unión a NAD. Las secuencias de la glucosamina-6-fosfato desaminasa y sus homólogas tienen aminoácidos invariables y aminoácidos que conservan la mayor parte de los caracteres fisicoquímicos en las mismas posiciones que el "fingerprint" (Tabla 6). Los cambios encontrados que se detallan a continuación pueden deberse a la especialización de enlazar azúcares fosfatados en lugar de dinucleótidos. La asa de unión del fosfato en la enzima de *E. coli* (aminoácidos

del 40 al 45) está cambiado, pero se puede ver claramente que los cambios forman un plegamiento el cual es capaz de unir azúcares fosfatados en la misma posición espacial. Esta asa tiene un residuo altamente conservado en la misma posición: una Gly que es un ligante importante del fosfato y su posición se ha conservado seguramente porque es el único aminoácido capaz de presentar los ángulos de Ramachandra adecuados para mantener la conformación de la asa. Las otras dos Gly que se presentan en el motivo de unión a mononucleótido son sustituidas por prolinas, la del sitio 45 es totalmente conservada y la del sitio 40 tiene tres variaciones (Pro, Ala y Thr). El Asp 72 es otra posición altamente conservada en la GlcN6PD de *E. coli* y sus homólogas. En las enzimas que ligan NAD este Asp catalítico puede variar a Glu. En la desaminasa de *E. coli* y sus homólogas existe una inserción de un aminoácido hidrofóbico a nivel del residuo 71 y forma una pequeña desviación de la hebra beta para colocar en su posición al Asp 72. Entre las posiciones 39 y 40 de la desaminasa de *E. coli* y sus homólogas se presenta un indel con respecto a las enzimas del "fingerprint", lo cual indica una delección al principio de la asa de unión del fosfato. El estudio estructural nos indica que esta delección no es importante, ya que provoca en la alcohol deshidrogenasa una conformación diferente entre la asa de ligadura y la primera hebra beta del motivo de unión a mononucleótido. Otra modificación es que la primera hebra beta en la desaminasa y sus homólogas es más grande por 2 aminoácidos y por lo tanto la primera posición del "fingerprint" no está considerado en este alineamiento. La primera posición es un residuo básico o hidrofílico que en la desaminasa puede ser un aminoácido hidrofóbico o polar.

**Alineamiento del motivo de unión a NAD entre diversas enzimas que fijan NAD y las enzimas de la familia de la glucosamina.**

	beta		loop	hélice	loop variable	beta	
	3 3	4 0	4 6			6 6	7 2
a	- - T C A V F	G L G G V G	L S V I M G C	K A A G A A		R I G V - D	I
b	- - K I T V V	G V G A V G	M A C A I S I	L M K D L A D		E V A L V - D	V
c	- - K I G I D	G F G R I G	R L V L R A A	L S C G A Q		V V A V N - D	P
d	- - D Y L V I	G G G S G G	L A S A R R A	A E L G A		R A A V V - E	S
e	- - Q V A I I	G A G P S G	L L L G Q L I	H K A G I		D N V I L - E	R
f	- - T V G V L	G S G H A G	T A L A A W F	A S R H V P T A		L W A P A - D	H
g	- - H Y T V I	G G G L M G	A G I A Q V A	A A T G H		T V V I V - D	Q
h	- - R V V V I	G A G V I G	L S T A L C I	H E R Y H S		V L Q P L - D	V
i	P F V L G L -	P T G G T P	M T T Y K A D	V E M H K	A G Q V S F K H	V V T E N M D	E
j	P F V L G L -	P T G G T P	L K T Y Q E L	I R L Y Q	A G K V S F K H	V V T E N M D	E
k	Y F T L G L -	P T G G S T P	L G C Y K K L	I E Y Y K	N G D L S F K Y	V K T E N M D	E
l	T F V L G L -	P T G G S S P	E G I Y A K L	I E A N K	Q G R V S F K N	V V T E N M D	E
m	F A K I G L -	P A T G V S P	V D C Y K L I	I K H S K	Q N N I S W K N	I T T E N L D	E
n	N A V I C L -	P T G A T P	L L T Y H Y I	V E K I H	Q Q Q V D V S Q	L T F V K L D	E
o	R V N L A L -	P A G S T P	K G M Y E Y L	T T L V K	- G K P W Y D N	C Y F E N F D	E

Tabla 6.- Alineamiento de la Glc6PD y sus enzimas homólogas contra los dominio de unión a NAD de diversas enzimas. Las cinco primeras enzimas (de a á e) fueron empleadas para el fingerprint (Wierenga, R.K., et. ai. 1985).

- a.- ALCOHOL DEHIDROGENASA DE HIGADO DE CABALLO: ADH
- b.- LACTATO DEHIDROGENASA DE DOGFISH: LDH
- c.- GLICERALDIHEDO-3-FOSFATO DEHIDROGENASA DE LANGOSTA: GAPDH
- d.- GLUTATION REDUCTASA HUMANA: GR
- e.- P-HIDROXIBENZOATO HIDROXILASA DE Pseudomonas Fluorescens: PHBH
- f.- NOPALINA SINTETASA DE Agrobacterium Tumefaciens: NS
- g.- L-3-HIDROXIACIL-COA-DEHIDROGENASA DE CERDO: LHCOADH
- h.- D-AMINASA OXIDASA DE CERDO: DAO
- i.- GLUCOSAMINA-6-FOSFATO DESAMINASA DE E. Coli: GluN6PDEC
- j.- GLUCOSAMINA-6-FOSFATO DESAMINASA DE H. Influenzae: GluN6PDHI
- k.- GLUCOSAMINA-6-FOSFATO DESAMINASA DE HUMANO: GluN6PDHUM
- l.- GLUCOSAMINA-6-FOSFATO DESAMINASA DE C. Albicans: GluN6PDCA
- m.- Posible GLUCOSAMINA-6-FOSFATO DESAMINASA DE M. Capricolum: isomerasa-MC
- n.- Posible ISOMERASA u18997: isomerasa-ECu
- o.- Posible ISOMERASA 1103281: isomerasa-ECI

## Conclusiones.

- ◊ El método empleado para proponer estructuras terciarias (dinámica molecular) funciona satisfactoriamente para tratar de explicar y describir el probable tipo de catálisis de las enzimas homólogas a la GlcN6PD de *E. coli* y proponer en algunos casos su posible estructura cuaternaria.
- ◊ Los modelos proponen que estas enzimas mantienen un plegamiento similar, derivado de un motivo de unión a NAD, modificado para enlazar azúcares fosfatados con dos inserciones en el sitio activo que se mantienen: la asa catalítica (residuos 137 a 157) y el motivo estructural hebra  $\beta$  B/hélice  $\alpha$  5/hebra  $\beta$  C' que aísla al substrato para la reacción catalítica (residuos 158 a 156).
- ◊ Las secuencias de *M. capricolum*, y de *E. coli*: U18997 y L103281 son probables isomerasas de diferentes substratos a la glucosamina-6-fosfato desaminasa debido, a las mutaciones del sitio activo.
- ◊ Las secuencias estudiadas de *M. capricolum*, de *E. coli*: U18997 y L103281, y *C. albicans* son proteínas no hexáméricas debido a la falta de las estructuras que forman el hexámero.
- ◊ La enzima de *H. influenzae* es un hexámero como en la enzima de *E. coli* y *h. sapiens*.
- ◊ El extremo C-terminal de la enzima humana es clave en la modulación de la alostericidad.
- ◊ La enzima de *Mesocricetus auratus* dada su extrema similitud con la enzima humana es un hexámero de la misma función y propiedades catalíticas similares.
- ◊ Estructuralmente se comprueba que el  $\text{NH}_4^+$  tiene una entrada particular durante la reacción reversa. Dicha entrada se mantiene en todos los modelos y puede explicar el mecanismo al azar observado en la glucosamina-6-fosfato desaminasa de *E. coli*.

## Apéndice 1: Homology

Homology es un programa de computadora que ayuda en la construcción de estructuras de proteínas. Las secuencias de aminoácidos pueden ser correlacionadas por medio de alineamientos de las mismas. Un alineamiento de secuencias que con una identidad de al menos 20%, en la cual se incluyen los aminoácidos claves de la catálisis, nos indica que esas proteínas son homólogas y forman parte de una misma familia. Cuando además de la secuencia primaria se conoce la estructura terciaria de una de las proteínas del alineamiento se puede correlacionar ésta estructura terciaria con la secuencia primaria de aquellas proteínas a las que no se les conoce la estructura terciaria. Entonces nosotros podemos extrapolar para construir por medio de Homology un modelo basado en la estructura terciaria. Una parte fundamental del método de construcción de Homology es que asigna coordenadas de la proteína de referencia al esqueleto de carbono (C'-C $\alpha$ -N) o backbone del modelo. Las diferencias entre secuencias como los indeles, que por lo general son más comunes en las zonas periféricas de la molécula (como las asas) pueden ser construidas por búsqueda en estructuras de otras proteínas. La búsqueda de regiones en proteínas reportadas en el Protein Data Bank usa un criterio geométrico para designar las estructuras similares tomando en cuenta el número de aminoácidos que falten o aumenten. La mejor estructura es designada por el menor rms (root-mean-squared) de las distancias entre carbonos  $\alpha$ . El rms está definido por la ecuación siguiente:

$$\sum_{i=1}^N \frac{(x-x_o)^2 + (y-y_o)^2 + (z-z_o)^2}{N} \quad (1)$$

El programa asigna las diez mejores estructuras. El programa asigna coordenadas también a los residuos cambiados manteniendo la misma orientación (siempre que es posible) de las cadenas laterales. Homology puede determinar rotámeros en base a librerías (Pouder y Richards 1987) y elegir las conformaciones preferidas de un aminoácido según su posición o puede ser elegido manualmente (a veces conviene mantener el rotámero cristalográfico). Un aminoácido puede tener rotámeros muy diferentes según la posición que tenga este en el plegamiento. El modelo se va construyendo bajo el hecho de que las proteínas homólogas mantienen un mismo plegamiento y solo hay variaciones en

regiones de baja homología (estas suelen mantener el plegamiento) y otras que son muy variables como los asas y regiones "extras" de la molécula.

Finalmente el modelo obtenido es refinado por medio de una simulación de dinámica molecular por medio del programa Discover para producir un modelo que sea químicamente y conformacionalmente adecuado. Esta técnica presenta la mecánica molecular que incluye las minimizaciones y los cálculos de vibración armónica y la dinámica molecular.

## Apéndice 2 : Discover

Este programa refina modelos basados en estructuras terciarias de cristalografía creados en Homology por medio de mecánica y dinámica molecular. Estas técnicas simulan el movimiento de los átomos, son una aproximación de la trayectoria de una molécula en solución. Los resultados de estos métodos dependen del tipo de "forcefield" que se emplean. Las simulaciones realizadas por Discover se basan en los cálculos de la Energía potencial para una configuración dada de átomos. Estos cálculos junto con las primera y segunda derivada respecto a las coordenadas atómicas proporcionan la información necesaria para la minimización, análisis de vibración armónicos y dinámica molecular. Discover emplea la Energía potencial de superficie la cual se basa en la ecuación llamada aproximación de Born-Oppenheimer donde el movimiento de los electrones se pueden desacoplar de los del núcleo dando dos ecuaciones de las cuales la primera es:

$$H\psi(r;R)=E\psi(r;R) \quad (2)$$

que define la Energía , la cual es una función de las coordenadas del núcleo. Esta es la Energía potencial de superficie.

EL forcefiel es el "empirical fit" de la Energía Potencial de Superficie. El "forcefield" define las coordenadas usadas y los parámetros ajustados en el "empirical fit" de la Energía Potencial de Superficie. La fórmula matemática involucra las coordenadas de las molécula cuando estas son combinadas con el "forcefield" creando la expresión de Energía de la molécula. Como ejemplo se muestra la fórmula para el agua, la cual describe la Energía Potencial de Superficie únicamente para una sola molécula de agua:

$$V(R) = k_{oh}(b - b_{oh}^o)^2 + k_{oh}(b' - b_{oh}^o)^2 + k_{hoh}(\theta - \theta_{hoh}^o)^2 \quad (3)$$

Esta ecuación involucra coordenadas de la longitud de los enlaces y de los ángulos (b, b', y  $\theta$ ), los parámetros de el "forcefield" (constantes de fuerza  $k_{oh}$  y  $k_{hoh}$  y los valores de referencia  $b_{oh}^o$  y  $\theta_{hoh}^o$ ). Una ecuación mas realista para el mismo caso es:



$$V(R) \sum_b D_b [1 - \exp(-a(b - b_0))]^2 \sum_{\theta} H_{\theta} (\theta - \theta_0)^2$$

$$\sum_{\phi} H_{\phi} [1 - \cos(n\phi)] \sum_x H_x X^2$$

$$\sum_b \sum_{b'} F_{bb'}(b - b_0)(b' - b'_0) \sum_{\theta} \sum_{\theta'} F_{\theta\theta'}(\theta - \theta_0)(\theta' - \theta'_0) \quad (4)$$

$$\sum_b \sum_{\theta} F_{b\theta}(b - b_0)(\theta - \theta_0) \sum_{\theta'} \sum_{\phi} F_{\theta\theta'\phi}(\theta - \theta_0)(\theta' - \theta'_0) \cos\phi$$

$$\sum_x \sum_{x'} F_{xx'} X X' \sum_i \sum_{j \neq i} \frac{A_{ij}}{r_{ij}^{12}} - \sum_i \sum_{j \neq i} \frac{B_{ij}}{r_{ij}^6} - \sum_i \sum_{j \neq i} \frac{q_i q_j}{r_{ij}} \quad (5)$$

Donde  $b$  es igual a la longitud de enlace,  $\theta$  es igual al ángulo de la inclinación,  $\phi$  es igual al ángulo de torsión,  $X$  es la distorsión planar de los átomos fuera de plano formado por los átomos  $a$  a los que enlazan.  $\sum_b \sum_{b'}$ ,  $\sum_{\theta} \sum_{\theta'}$ ,  $\sum_b \sum_{\theta}$ ,  $\sum_{\theta} \sum_{\theta'}$ ,  $\sum_x \sum_{x'}$ , son términos que cuentan las interacciones que hay entre los cuatro tipos de coordenadas internas.  $\sum_i \sum_{j \neq i}$  representan las interacciones entre átomos no unidos como la suma de repulsiones y atracciones en términos de Lennard-Jones y Coulomb, todas las cuales son funciones de las distancias entre pares de átomos  $r_{ij}$ . El "forcefield" define la forma de la función de cada término de la ecuación, también define las coordenadas internas como una función de las coordenadas atómicas Cartesianas.

#### Mecánica molecular.

En estos sistemas el tiempo de evolución del sistema es ignorado y se emplea para encontrar geometrías particulares y sus Energías asociadas. Esto incluye encontrar estructuras en equilibrio, estados de transición, Energías relativas y frecuencias de vibración armónicas. Las mecánicas moleculares se pueden separar en dos grupos: las minimizaciones y los cálculos de vibración

armónica.

Minimización.

Este proceso nos permite hallar conformaciones estables, encontrar un punto donde la red de Fuerza en cada átomo desaparezca. Los datos obtenidos de la Energía Potencial de Superficie nos permite realizar esta búsqueda. La minimización se realiza en dos etapas:

1) Una ecuación describe la Energía del sistema como función de sus coordenadas y debe ser definida y evaluada por una conformación dada.

2) La conformación es ajustada a un bajo valor de la función de la Energía. El mínimo puede encontrarse en un ajuste o en cientos de interacciones dependiendo del algoritmo usado, la forma de la función y la talla de la molécula. La eficiencia de la minimización esta dada por el tiempo necesario para evaluar la función y el numero de ajustes estructurales necesarios para alcanzar un mínimo. La búsqueda del mínimo se define por la aplicación de cada algoritmo a la minimización de una ecuación cuadrática como la siguiente función elíptica:

$$E(x,y)=x^2+5y^2 \quad (6)$$

con la cual una minimización comienza con cualquier ecuación análoga. Diferentes algoritmos pueden ser usados. Para la función dada que define la Energía Potencial de Superficie y un punto inicial de inicio, en una minimización se debe determinar la dirección al mínimo y la distancia al mínimo en esa dirección. Una dirección inicial es la derivada de la función en un punto. La derivada de la función de un punto (la derivada es un vector de dos dimensiones respecto a la ecuación (6) esta dada por:

$$\nabla E=(2x,10y) \quad (7)$$

Para optimizar la eficiencia se usan algoritmos como "conjugate gradients" y Newton-Rampson los cuales usan los datos de como cambia la derivada para determinar la dirección. El minimizador de Discover para determinar la dirección al mínimo emplea una estrategia denominada "line search" la cual no es sino la búsqueda del mínimo entre dos puntos mas altos de energía.

Molecular Dynamics.

Las ecuaciones empleadas en una Dinámica Molecular resuelven la clásica ecuación de movimiento para un sistema de N átomos interactuando de acuerdo con el "forcefield" de la Energía Potencial. La búsqueda de conformaciones es un propósito fundamental de las dinámicas moleculares. Utilizan diferentes mecanismos para controlar la temperatura y la presión del sistema que se simula. La Dinámica Molecular nos permite obtener conjuntos estadísticos para los cuales se pueden hacer cálculos termodinámicos, estructurales y propiedades dinámicas. Por otra parte nos permite observar una aproximación del movimiento de los átomos que es fundamental para las interacciones entre las moléculas y ubicar las conformaciones más comunes en los movimientos de los aminoácidos de una proteína o péptido y de compuestos como los sustratos. La trayectoria de la Dinámica se puede construir por representación de la vibración de la molécula, este es un método que restringe los movimientos armónicos al mínimo. En la Dinámica molecular el tiempo de evolución de la molécula es importante pues con el tiempo se encuentra una gama más amplia de conformaciones y el sistema se estabiliza. El "forcefield" que se emplea para resolver la ecuación de movimiento de Newton es:

$$F_i(t) = m_i a_i(t) \quad (8)$$

Donde  $F_i$  es la fuerza,  $m_i$  la masa y  $a_i$  la aceleración del átomo  $i$ . La fuerza del átomo puede ser calculada a partir de la derivada de la Energía Potencial  $v$  con respecto a las coordenadas  $r_i$ :

$$\frac{\delta v}{\delta r_i} = - \frac{\delta^2 v}{\delta t_i^2} m_i \quad (9)$$

Esta ecuación es determinística ya que parte de coordenadas iniciales y velocidades conocidas. Las velocidades y coordenadas para una Dinámica completa se llama la trayectoria. El método que resuelve diferenciales del tipo de la ecuación no. (9) es el método de diferencial finita. Dadas las coordenadas iniciales y las velocidades y otros datos como el tiempo ( $t$ ), se calculan las posiciones y velocidades en el tiempo  $t + \Delta t$ , este depende del método de integración. Los integradores de Discover 2.9.5 emplean el algoritmo Verlet leapfrog.

Temperatura.

Una variable de estado como la temperatura, especifica el estado termodinámico del sistema en las simulaciones dinámicas. Ésta se relaciona con la simulación dinámica a través de la Energía cinética, la cual es calculada a partir de las velocidades atómicas. La temperatura y la distribución de las velocidades en un sistema son relacionadas en la ecuación de Maxwell-Boltzmann:

Esta ecuación nos expresa la posibilidad  $f(v)$  de que una molécula de masa  $m$  tenga una velocidad de  $v$  cuando esta a la temperatura  $T$ . La Presión es otra variable termodinámica que se define como

$$f(v)dv = \left[ \frac{m}{2\pi kt} \right]^{\frac{3}{2}} e^{-mv^2/2kt} 4\pi v^2 dv \quad (10)$$

la fuerza por unidad de área. La presión se puede decir que es un "tensor" de la forma general (McQuarrie 1976):

$$P = \begin{pmatrix} P_{xx} & P_{xy} & P_{xz} \\ P_{yx} & P_{yy} & P_{yz} \\ P_{zx} & P_{zy} & P_{zz} \end{pmatrix} \quad (11)$$

Cada elemento del tensor es una fuerza que actúa en la superficie de un volumen cúbico infinitesimal que tiene bordes paralelos en los ejes  $x$ ,  $y$ , y  $z$ . En una situación isotópica, donde la  $F$  es igual en todas direcciones y no hay  $F$  de viscosidad, el tensor de la presión es diagonal y se puede expresar así:

$$P = \rho \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (12)$$

Donde la cantidad escalar  $\rho$  es la presión. Algunas veces el tensor de la presión es también conocido como el Tensor de stress o stress. El uso general de las simulaciones dinámicas son:

- Búsqueda de conformaciones.
- Generación de la "estadistical ensembles".
- Estudio sobre el movimiento de las moléculas.

### Apendice3: Profiles-3D.

El Profiles-3D versión 2.3.D.1994 sirve para verificar la calidad de un modelo proteico teórico. Se mide la compatibilidad de una secuencia de aminoácidos con su estructura terciaria. Simplifica la estructura terciaria tridimensional a una representación de una dimensión llamada "environment string". El cual es comparado en una gráfica con la secuencia de aminoácidos. Esto significa ubicar cada uno de los aminoácidos dentro de un ambiente determinado que puede ser hidrófobico o polar y sus categorías intermedias. Las categorías de ambiente son formadas de acuerdo al área enterrada de las cadenas laterales de la proteína, la fracción del área de la cadena lateral que esta expuesta a átomos polares y la estructura secundaria local. Existen seis tipos básicos de clases de ambientes y cada uno se divide en tres clases de acuerdo con la estructura secundaria local. El calculo de área enterrada y fracción polar se basa en el algoritmo de Lee & Richards (1971). para determinacion de superficies de accesibilidad a solvente de una proteína. El cual se basa en una esfera imaginaria de solvente que rodea al átomo o molécula. La ecuación que calcula el área de la superficie de accesibilidad al solvente  $A_a$  es para un átomo:

$$A_a = \left(\frac{n_a}{n_t}\right)A_t \quad (13)$$

Donde  $n_a$  es el no. de puntos de la muestra accesibles al solvente,  $n_t$  es el no. total de puntos de la muestra en la esfera de solvente y  $A_t$  es el total de área de la superficie de la esfera de solvente. El área de superficie enterrada de una cadena lateral se calcula como la diferencia entre el área de accesibilidad al solvente de la cadena lateral en la proteína un tripeptido Gly-X-Gly. El valor  $a$  sido calculado para cada uno de los 20 aminoácidos ( Eisenberg et. al.1989). El área de accesibilidad al solvente de la cadena lateral es entonces calculada como la suma de las áreas de accesibilidad al solvente de las átomos que la constituyen incluyendo al C- $\alpha$ .

La fracción de área cubierta por átomos polares es calculada con :

$$F = \frac{N_p}{N_t} \quad (14)$$

en donde  $N_p$  es el no. de puntos de la muestra cubierta por átomos polares (N, O, o agua) para todos los átomos de una cadena lateral.  $N_l$  es el no. de puntos de la muestra para todos los átomos de la cadena lateral.

El "environmental string" no puede ser alineado con la secuencia de aminoácidos sin la misma medida de la compatibilidad de cada uno de los 20 aminoácidos con cada una de las clases de ambiente, por lo tanto se requiere generar una matriz de comparación dependiente de posición. La matriz de 3D-1D es calculada a partir de las probabilidades de encontrar cada uno de los 20 aminoácidos en cada una de las clases de ambiente observados en los bancos de datos de estructuras conocidas y sus secuencias. Cada elemento  $S_{ij}$  de la matriz es un score que indica la compatibilidad del residuo  $i$  con el ambiente  $j$ . Valores altos significa mayor compatibilidad. Valores negativos baja compatibilidad. Los elementos individuales de la matriz son calculados (Fano, 1961) así:

$$S_{ij} = \ln\left(\frac{P(i;j)}{P_i}\right) \quad (15)$$

en donde  $P(i;j)$  es la probabilidad de encontrar un residuo  $i$  en un ambiente  $j$  y  $P_i$  la probabilidad de encontrar el residuo  $i$  en cualquier ambiente.  $P(i;j)$  y  $P_i$  son estimados a partir de los datos experimentales de proteínas de estructuras conocidas de los bancos de datos. Las clases de ambiente se reajustan, la matriz es calculada cada vez y con el máximo  $S_i$ . El score 3D-1D total sumado sobre todos los tipos de residuos y clases de ambiente en los alineamientos experimentales son:

$$S_i = \sum_j \sum_l N_{ij} S_{ij} \quad (16)$$

Para los casos de verificación de estructura de modelos proteicos se contruye el perfil 3D de la proteína y mide su compatibilidad con la secuencia. En este caso solo se calcula el 3D-1D score, mediante calculos de scores 3D-1D locales en ventanas de 5 a 20 residuos. Esto puede ser graficado contra la posición de la secuencia lineal de aminoácidos (Lüthy, 1992). En la verificación de estructuras el score de compatibilidad es difícil de interpretar. En este caso es necesario compensar el efecto de la longitud y comparar con el score del modelo cristalográfico en el que se baso el modelo. La comparación se puede realizar en una gráfica log-log con el 3D-1D score de cada

proteína conocida y la longitud de la secuencia. La regresión lineal de esta gráfica está dada por :

$$Scale = e^{(-0.83 + 1.000(x/L))} \quad (17)$$

Scale es el score esperado correcto de una proteína teniendo la longitud de la secuencia L.

## Apéndice 4: Predicción de estructura secundaria.

Este apéndice solo contiene la predicción múltiple del alineamiento, la predicción para *E. coli* y *M. capricolum* con los programas PHD (Rost & Sander.1993,1994.Levin, Pescarella, Argos & Garnier.1993) y SSP (Solovyev,V.V. & Salamov, A.A.1994).

Predicción de estructura secundaria respecto a las 7 secuencias de GlcN6P y sus posibles afines por el programa PHD.

### Abbreviations: PHDsec

---

sequence:

AA : amino acid sequence

secondary structure:

HEL: H=helix, E=extended (sheet), blank=other (loop)

PHD: Profile network prediction Heidelberg

Rel: Reliability index of prediction (0-9)

detail:

prH: 'probability' for assigning helix

prE: 'probability' for assigning strand

prL: 'probability' for assigning loop

note: the 'probabilities' are scaled to the interval 0-9, e.g.,

prH=5 means, that the first output node is 0.5-0.6

subset:

SUB: a subset of the prediction, for all residues with an expected average accuracy > 82% (tables in header)

note: for this subset the following symbols are used:

L: is loop (for which above " " is used)

": means that no prediction is made for this residue, as the reliability is: Rel < 5

### Abbreviations: PHDacc

---

solvent accessibility:

3st: relative solvent accessibility (acc) in 3 states:

b = 0-9%, i = 9-36%, c = 36-100%.



PHD: Profile network prediction Heidelberg  
 Rel: Reliability index of prediction (0-9)  
 P\_3: predicted relative accessibility in 3 states  
 note: for convenience a blank is used intermediate (i).  
 10st: relative accessibility in 10 states:  
 = n corresponds to a relative acc. of n\*n %  
 subset:  
 SUB: a subset of the prediction, for all residues with an expected  
 average correlation > 0.69 (tables in header)  
 note: for this subset the following symbols are used:  
 "I": is intermediate (for which above " " is used)  
 ".": means that no prediction is made for this residue, as the  
 reliability is: Rel < 4

protein: COLI length 312

```

          . . . . . 1 . . . . . 2 . . . . . 3 . . . . . 4 . . . . . 5 . . . . . 6
AA      | .....MRLIPLTTAEQVGKWAARHIVNRINAFKPTADRPFVLGLPT|
PHD sec |          HHHHHHHHHHEEEEE   HHHHHHHHHHHHHHHH   EEEEE   |
Rel sec | 98998764489999852013566435735899999999999974388998617786258|

```

detail:

```

prH sec | 0000012368899986544211111136899999999999976300000000000000|
prE sec | 0100010000000000001567763100000000000000000000000000147787520|
prL sec | 988987763100000244321112567621000000000000012688888741112478|
subset: SUB sec | LLLLLL. .HHHHHH. . . . . EEE. .LL. HHHHHHHHHHHHHHH. . LLLLLL. EEEE. LL|

```

ACCESSIBILITY

```

3st:    P_3 acc | eeebebeebbbeebbbb bbbbee eebbebbbbebbbbebeeee bbbbbe|
10st:   PHD acc | 99706079007007700000500006647600700070000006707677750000077|
        Rel acc | 575120430133454312061543211042414066406310413141334115244223|
subset: SUB acc | eee. .e. . . . . bee. . . . . b. bb. . . . . e. b. e. bbe. b. . . . . e. . . . . e. . . . . b. bb. . . |

```

```

          . . . . . 7 . . . . . 8 . . . . . 9 . . . . . 10 . . . . . 11 . . . . . 12
AA      | GGTPMTTYKALVEMHKAGQVSKHVVTFNMDEYVGLPKEHPESYYSFMHRNFFDHDVIPA|

```

```

PHD sec |   HHHHHHHHHHHH   EEEEEEEEE   HHHHHHHHHHHH   |
Rel sec |998647999999997618515688756414323579965724678889988633478882|
detail:
prH sec |001168899999998740000000000001232100012146788889888763310003|
prE sec |000000000000000000256788767642111110000000001000000000000|
prL sec |998731000000001158742100121245545678877853211000011235678885|
subset: SUB sec |LLLL.HHHHHHHHHHH.LL.EEEEE.E....LLLLLL.HHHHHHHHHH..LLL.|

```

ACCESSIBILITY

```

3st:   P_3 acc |eeee ebbbbbbeebbeebbbbbeebbeebbeeb eebbbbbb e bbebbebe|
10st:  PHD acc |97665600600070077070007000006077007007747700000464006007066|
       Rel acc |422101402064520550501250523222432230056145211123110432033221|
subset: SUB acc |e....b...bbe..ee.e...e.b.....e.....ee.ee.....b.....|

```

.....13.....14.....15.....16.....17.....18

```

AA      |ENINLLNGNAPDIDAECRQYEEKIRSYGKIHLFMGGVGNDBGHIAFNEPASSLASRTRIKT|
PHD sec |   E           HHHHHHHHHHHH   EEEEE E   EEE           EE|
Rel sec |22311257778941568999999970982788751166541313787888888641021|

```

detail:

```

prH sec |3321110001003567889999999740000000000222110000000000111211|
prE sec |0124432111000000000000000014888764510124543101001000123344|
prL sec |5554356887896422100000001498510112447765334588888888764333|
subset: SUB sec |.....LLLLLL.HHHHHHHHHHH.LL.EEEEE..LLL....LLLLLLLLLL....|

```

ACCESSIBILITY

```

3st:   P_3 acc |ebbbebbe eebeeebe eebeeebeb bbbbbb beb bbbeeeee bee bee|
10st:  PHD acc |70060060647607706647760676060500000051704000769777550765066|
       Rel acc |503114100031043532204513231014175521210411242414233110411322|
subset: SUB acc |e...b.....e.e...ee.....b.bbb.....e...b.e.e.....e.....|

```

.....19.....20.....21.....22.....23.....24

```

AA      |LTHDTRVA..NSRFFDNDVNQVPKYALTVGVGTLDD.AEEVMILVLGSQKALALQAAVEG|
PHD sec |   HHHHHHHHHHHH   EEE   HHHHHHHHEEEEE   HHHHHHHHHH |
Rel sec |12122467642121368778896541552521358999753123204632689999861|

```

detail:

```

prH sec |23444677754444210111000111001134568888775433321357889988874|
prE sec |3211110012332221001000122466410001000000245653100000000000|
prL sec |44433211112223468778887664223655311000121100036763100000125|
subset: SUB sec |.....HHH.....LLLLLLLLL..EE.L...HHHHHHH.....L.HHHHHHHH.|

```

ACCESSIBILITY

3st: P\_3 acc |bbeebbbb b eeebeebbebbbbbbbe beebbbbbeebeebbee|
10st: PHD acc |00770000000550777076007000000000650770000006670000670077|
Rel acc |213415134514112234031204045021401402034478867001140052132253|
subset: SUB acc |...e.b.bb.b....e....e.bb...b..b....eebbb...e..b....e.|

.....25.....26.....27.....28.....29.....30

AA |CVNH.MWTISCLQLHPKAIMVCDEPSTMELKVKTLRYFNEL.....|
PHD sec | HHHH EEEEE HHHHHHHHHHHHEEEEE |
Rel sec |7666531554321355488883534112444655444315676423177667421246|

detail:

prH sec |1111124666544321100000233455656766656654111110000111234431|
prE sec |01000121122211002688885100000122122121123677665541111100000|
prL sec |7777765411113467621001366654422111112211100123357776765567|
subset: SUB sec |LLLLL.HH....LL.EEEE.L.....HHH....EEEE....LLLLL....L|

ACCESSIBILITY

3st: P\_3 acc | beebbbbbeeb bebbbbeebbeeebebebbe beebbbbbee e eeeee|
10st: PHD acc |50667000000604060000077007770706007406700000006664757777|
Rel acc |011120115444111015556633014362512335151335230303211114196875|
subset: SUB acc |.....bbb.....bbbb...e.e.e....e.b..b.....e.eeeee|

.....31.....32.....33.....34.....35.....36

AA |.....|
PHD sec | |
Rel sec |55456999999|

detail:

prH sec |11222000000|
prE sec |11000000000|
prL sec |66667999999|
subset: SUB sec |LL.LLLLLLLL|

ACCESSIBILITY

3st: P\_3 acc |eeeeeeee ee|
10st: PHD acc |777787977599|
Rel acc |784664975069|
subset: SUB acc |eeeeeeee.ee|

Predicción de estructura secundaria de la GlcN6P de *E. coli* con el programa PHD.

```

      ....,....1....,....2....,....3....,....4....,....5....,....6
AA      |MRLIPLTTAEQVGKWAARHIVNRINAFKPTADRPFVLGLPTGGTPMTTYKALVEMHKAGQ|
PHD sec |          HHHHHHHHHHHHHHHHH          EEEEE          HHHHHHHHHHH      |
Rel sec |99424899589999999999999999999999448899998188762999773999999999962893|

```

detail:

```

prH sec |000010007899999999999999863000000000000000116999999999974000|
prE sec |00332000000000000000000000000000000000001588874000000000000000003|
prL sec |98656899210000000000000000368999998411125999883000000000025896|
subset: SUB sec |LL...LLHHHHHHHHHHHHHHHH...LLLLLL.EEEE.LLLL.HHHHHHHHHHH.LL.|

```

ACCESSIBILITY

```

3st:    P_3 acc |b ebebeeeebbebbbe bbb b eeeeee bbbbbbbbeeeebbeebbeb ebe|
10st:   PHD acc |057060777870070007400050589779974000000079966770076007057707|
        Rel acc |013212443651271486088116143767561067754023421541242855215735|
subset: SUB acc |.....ee.ee..e.bbe.bb..b.e.eeeee..bbbbb...e..ee..e.bbe..ee.e|

```

```

      ....,....7....,....8....,....9....,....10....,....11....,....12
AA      |VSFKHVVTFMDEYVGLPKEHPESYYSFMHRNFFDHVDIPAENINLLNGNAPDIDAECRQ|
PHD sec |EEEEEEEEEE          HHHHHHHHHHHHHHH          HHHHHHH          HHHHHHHH|
Rel sec |78999998511111699985827799999999852899499989863788147799999|

```

detail:

```

prH sec |0000000003444200002157888999998754000289889873111468899999|
prE sec |88888998752000000000000000000000000000000000000000000000000000|
prL sec |100000002444547899878411000000001245899600010126888531100000|
subset: SUB sec |EEEEEEEE...LLLLLL.HHHHHHHHHHH...LLL.HHHHHHH.LLL..HHHHHHH|

```

ACCESSIBILITY

```

3st:    P_3 acc |bbbebbbbbbbe bbbbeeeebbbbbb ebbbebbbeebbebbbbb beebabee|
10st:   PHD acc |00070000006500006787770000046600700600670060005077077077|
        Rel acc |356408967261144200884365301551116332614313462581105444353757|
subset: SUB acc |.bbe.bbbb.b..bb...eee.eb...bb...b...b.b...bb.bb...beeb.e.bee|

```

```

      ....,....13....,....14....,....15....,....16....,....17....,....18
AA      |YEEKIRSYGKIHLFMGGVNDGHI AFNEPASSLASRTRIKTLTHDTRVANSRFFDNDVNQ|
PHD sec |HHHHHHH          EEEEE          E          EHHHH          |
Rel sec |999997459589983589975211126899987976540411104313022346999434|

```

detail:

```

prH sec |9999986200000000001232100000000011111143442245333221000332|
prE sec |00000000289885200000134442000001000224533112111211221000000|
prL sec |000001379710013688987444457889988887664213445543344557999656|
subset: SUB sec |HHHHHH.LLLEEEE.LLLLLL...LLLLLLLLLLL.....LLLLL...|

```

ACCESSIBILITY

```

3st: P_3 acc | eeebeebb bbbbbbbbeebbbbeebbeeebee eeebeebbebbb bbe eeee|
10st: PHD acc |47760670050000000077000000690067707747760677066000500759977|
Rel acc |148162606291967111035322770131115326314535176212624254403466|
subset: SUB acc |.ee.b.e.b.b.bbb....e...bb.....e..e..ee.b.ee...b.b.bbe..eee|

```

.....19.....20.....21.....22.....23.....24

```

AA |VPKYALTVGVGTLDDAEVMILVLGSKALALQAAVEGCVNHMTISCLQLHPKAIMVCD|
PHD sec | EEEEE EEEE EEEEE HHHHHHHHHH EHHHH EEEEE |
Rel sec |64136873643561854189986185289999998197699933102241032356654|

```

detail:

```

prH sec |2221101001110002100000000238999999850000000011464432210000|
prE sec |01457875125674102489987400000000000001100033444200012567762|
prL sec |753210137621248664100014865100000001488799956443224454221126|
subset: SUB sec |L...EEE.L..EE.LL..EEEE.LL.HHHHHHHHH.LLLLLL.....EEEE.|

```

ACCESSIBILITY

```

3st: P_3 acc |bbebbbebbbbbeeeebbbbbbeeeebbbeeebebeeeeee bbbbb e bebbbbbe|
10st: PHD acc |00700060000000777600000077660007760806797830000058307000009|
Rel acc |715287264724316341889695365122754413701336611254614106366873|
subset: SUB acc |b.e.bb.bbb.b..e.e.bbbbb.ee...bbe..e...ee...bbb.e..e.bbbb.|

```

.....25.....26.....27.....28.....29.....30

```

AA |EPSTMELKVKTLRYFNEL|
PHD sec | HHEHHHHHHHH |
Rel sec |951111211445334119|

```

detail:

```

prH sec |024444444566555440|
prE sec |00000134322221000|
prL sec |975443111211122459|
subset: SUB sec |LL.....H.....L|

```

ACCESSIBILITY

```

3st:   P_3 acc |eeeebebebebbe beee|
10st:  PHD acc |777707070700640689|
       Rel acc |352315342514212154|
subset: SUB acc |.e...e.e.e.b....ee|

```

Predicción de estructura secundaria de la probable GlcN6P de *M. capricolum* por PHD.

```

protein:       predict phd       length       212

               .....1.....2.....3.....4.....5.....6
AA             |MFYLKIIIFDKIEDLQEYCAQLFIDQVKQNPFAKIGFATGVSPVDCYKLIKHSKQNNIS|
PHD sec | EEEEEEE HHHHHHHHHHHHHHHHHH EEEEE HHHHHHHHHHHH EE|
Rel sec |97877884478269999999999999999999138993887169989399999999997159457|

```

```

detail:
prH sec |000000000057999999999999999999300000000000000399999999998520000|
prE sec |088888862100000000000000000000000688851000000000000000000367|
prL sec |9111101368842000000000000000468993111479989600000000001479621|
subset: SUB sec |LEEEEE..LL.HHHHHHHHHHHHHH..LLL.EEE.LLLLL.HHHHHHHHHHH.LL.EE|

```

ACCESSIBILITY

```

3st:   P_3 acc |bbbbbbbbbeeeebeebbbeebbeebbeeeebbbbeebbeebbeebbeb eebbb|
10st:  PHD acc |200000007777707600760076067876060060076067006600705770700|
       Rel acc |021906644443443322594154413176310155310132105421275611460626|
subset: SUB acc |...b.bbbbee.ee...bbe.bbe...ee...bb.....eb...bbe..ee.e.b|

```

```

               .....7.....8.....9.....10.....11.....12
AA             |WKNITTFNLDEFVNISKDHKQAFIKQMKDNLFDHLDIDPKNINFLDCQTSDPKKECQRYE|
PHD sec |EEEEEEE HHH HHHHHHHHHHHHHHHH EE HHHHHHHHHH|
Rel sec |789686322133369887918999999999741131681112113799888369999999|

```

```

detail:
prH sec |00000003455310011058899999999864421104433332100111578899999|
prE sec |788787632110100000000000000000012310001233210000000000000|
prL sec |1101113544235788878410000000013445478454433578888842000000|
subset: SUB sec |EEEEEE.....LLLLLL.HHHHHHHHHH.....LL.....LLLLLL.HHHHHHHH|

```

ACCESSIBILITY

```

3st:   P_3 acc |bebbbbbbbbbbbbee eebb ebb ebbbebbbbeebbbbbeebbeebbeeb|
10st:  PHD acc |070000000000000077477005600472006000066700000677077077706706|
       Rel acc |440735626016602066055341116130651370511544247133154054462511|

```

subset: SUB acc |be.b.bb.b..bb...ee.ee.b...b...bb..b.b..ebb.bb....ee.eeb.e..|

.....13.....14.....15.....16.....17.....18

AA |NLIRSVGDIDFYISLGINGHMAYNEPNTSFNTDTHVVKLTKETILDMVNKKKFNLSLDDC|

PHD sec |HHHHH EEEEE E HHHHHHHHHHHHHHHHHH |

Rel sec |999973994999937375640113799998797234478889889999988853785677|

detail:

prH sec |9998830000000000221110000000010135567888889999988863102211|

prE sec |00000006999861310024443100000000012111100000001000100000|

prL sec |0000169930000376767644458999888753210000011000000125887787|

subset: SUB sec |HHHHH.LL.EEEE.L.LLL....LLLLLLLLL...HHHHHHHHHHHHHH.LLLLLL|

ACCESSIBILITY

3st: P\_3 acc |eebee bebbbbbbbbbebbbbbb bbebebeb beebbeeb bebb beeebeeb|

10st: PHD acc |7706740600000000070000050070660705066007700506005008790770|

Rel acc |647240329082400120422365110030010501323515443122211004350464|

subset: SUB acc |eeb.e...b.b.b.....e...bb.....e.....b.eeb.....e.e.eeb|

.....19.....20.....21.....22.....23.....24

AA |PTHAMTMGIQTLNWTKKAITVSYGIHKDFS|

PHD sec | EEEE HHHH HHHEEEEE E |

Rel sec |41237633255315062268895353212349|

detail:

prH sec |2442111256653146531111112232100|

prE sec |11257751001321013478887622333330|

prL sec |63310125321136411110000265424569|

subset: SUB sec |...EE...HH..L.H..EEEE.L.....L|

ACCESSIBILITY

3st: P\_3 acc |bebbbbbbbbbeeeebbbbbeeeebbbe|

10st: PHD acc |070000000000707760000606660009|

Rel acc |25098255835624045158775121110223|

subset: SUB acc |.e.bb.bbb.bb.e.ee.bbbbb.....|

Predicción de estructura secundaria de la probable GlcN6P de *E. coli* por SSP.

Pred A: 'probability' for assigning  $\alpha$  helix.

Pred B: 'probability' for assigning  $\beta$  strand.

Pred ic: prediction.

```

pred A:          aaaaaaaaaa aaaaaaaaaa          aa
AA              N  2.9  C  N  2.0  C              N
pred B:                                bbbb
BB                                N3 C
Predic          aaaaaaaaaa aaaaaaaaaa          bbbb          aa
a/acid  MRLIPLTTAEQVGKWAARHIVNRINAFKPTADRPFVLGLPTGGT P M T T Y K
              10          20          30          40          50

pred A:  aaaaaaaaa
AA      5.4 C
pred B:          bbbbbbbbbbb
BB      N  4.3 C
Predic  aaaaaaaaa bbbbbbbbbbb
a/acid  ALVEMHKAGQVSFKHVVT FNMDEYVGLPKEH PESYYSFMHRNFFDHVDIP
              60          70          80          90          100

pred A:          aaaaaaaaaaa
AA              N  4.3  C
pred B:                                bbbbbbbb
BB                                N  2.3 C
Predic          aaaaaaaaaaa bbbbbbbb
a/acid  AENINLLNGNAPDIDAECRQYEEKIRSYGKIHLFMGGVGN DGHIAFNEPA
              110          120          130          140          150

pred A:                                aaaaaaaaaa
AA                                N  2.5
pred B:                                bbbbbbbb          bb
BB                                N  4.0 C          N
Predic          bbbbbbbbbaaaaaabb
a/acid  SSLASRTRIKTLTHDTRVANSRFFDNDVNQVPKYALT VGVGTL LLD AEEVM
              160          170          180          190          200

pred A:  aaaa  aaaaaaaaaa          a
AA      C  N  4.8  C              N
pred B:  bbbbb          bbbbbbbb          bbbb  bbbbbbbb
BB      3.6 C          N  2.0 C          N2 C  N  2.7 C
Predic  bbbbb  aaaaaaaaa bbbbbbbb          bbbb  bbbbbbbb
a/acid  IIVLGSQKALALQAAVEGCVNHMTISCLQLH PKAIMV CDE PSTMELKVK
    
```



	210	220	230	240	250
pred A:	aaaaaaaa				
AA	1.9 C				
pred B:					
BB					
Predic	aaaaaaaa				
a/acid	TLRYFNEL				

Predicción de estructura secundaria de la probable GlcN6P de *M. capricolum* por SSP.

```

pred A:          aaaaaaaaaaaaaaaaaa
AA              N    4.8    C
pred B:  bbbbbbbb                               bbbbbbbb
BB              N  3.9 C                               N  2.3
Predic  bbbbbbaaaaaaaaaaaaaaaaaa                               bbbbbbbb
a/acid  MFYLKIIIFDKIEDLQEYCAQLFIDQVKQNPFAKIGFATGVSPVDCYKLI
              10      20      30      40      50

pred A:          aaaaaaaaaaaaaaaaaa
AA              N    2.5    CN  3.1 C
pred B:  bb
BB              C
Predic  bb          aaaaaaaaaaaaaaaaaa
a/acid  IKHSKQNNISWKNITTFNLDEFVNI SKDHKQAFIKQMKDNLFDHLDIDPK
              60      70      80      90      100

pred A:          aaaaaaaaaa
AA              N    4.2    C
pred B:          bbbbbbbb
BB              N  1.7 C
Predic  aaaaaaaaaa          bbbbbbbb
a/acid  NINFLDCQTSDPKKECQRYENLIRSVDGIDFQYISLGINGHMAYNEPNTS
              110     120     130     140     150

pred A:          aaaaaaaaaa          aaaaaaaaaaaaaaaaaa
AA              N    2.7 C          N    2.8    C
pred B:  bbbbbbbb                               bbbb
BB              N  3.2 C                               N  1.
Predic  bbbbbbaaaaaaaaaa          aaaaaaaaaaaaaaaaaa
a/acid  ENTDTHTVVKLTKETI LDMVNKKKFNSLDDCPTHAMTMGIQTLLNWTKKAI
              160     170     180     190     200

pred A:
AA
pred B:  bbb
BB              9 C
Predic  bbb
a/acid  TVSYGIHKDFSY
              210

```

## Referencias

- Almagro, J.C., Vargas-Madrado, E., Lara-Ochoa, F., Horjales, E. (1995). Molecular modeling of a T-cell receptor bound to a major histocompatibility complex molecule: Implications for T-cell recognition. *Protein Science*. 4:1708-1717.
- Altamirano, M.M., Plumbridge, J. A., Horjales, E., Calcagno, M. L. (1995). Asymmetric allosteric activation of *Escherichia coli* glucosamine-6-phosphate deaminase produced by replacements of Tyr 121. *Biochemistry*. 34: 6074-82 .
- Altamirano, M.M., Plumbridge, J.A., Barba, H.A., & Calcagno, M.L. (1993). Glucosamine-6-phosphate deaminase from *Escherichia coli* has a trimer of dimers structure with three intersubunit disulphides. *Biochem. J.* 295:645-648.
- Altamirano, M.M., Plumbridge, J.A., & Calcagno, M.L. (1992). Identification of two cysteine residues forming a pair of vicinal thiols in glucosamine-6-phosphate deaminase from *Escherichia coli* and study of their functional role by site-directed mutagenesis. *Biochemistry* 31:1153-1158.
- Altschul, S.F., Gish, W., Miller, E.W., Myers, D.J., Lipman (1990). Basic local alignment search tool. *J. Mol. Biol.* 215: 403-410.
- Bairoch A. and Apweiler R. (1996) The SWISS-PROT protein sequence data bank and its new supplement TrEMBL. *Nucleic Acids Res.* 24:21-25.
- Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F. Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. (1977). The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112:535-542.
- Blow, D.M., Collyer, C.A., Goldberg, J.D. & Smart, O.S. (1992). Structure and mechanism of D-xylose isomerase. *Faraday Discuss.* 93, 67-73.
- Biosystems Technologies. Insight II 2.0.0. San Diego.
- Brooks III, C.L., Karplus, B.M., Pettitt. (1988). Proteins. An Interscience publications. USA. 259pp.

- Burland, V.D., Plunkett, G. III, Daniels, D.L. and Blattner, F.R.(1993) DNA sequence and analysis of 136 kilobases of the *Escherichia coli* genome: organizational symmetry around the origin of replication. *Genomics*. **16**:551-61.
- Proteína hipotética de 23.3 kDa L10328 ECOUW82\_82 L=213 comment [similarity] belongs to the glucosamine/galactosamine-6-phosphate Isomerase family
- Calcagno,M.,Campos,P.J.,Mulliert,G.,&Suástegui,J.(1984). Purification, molecular and kinetic properties of glucosamine-6-phosphate isomerase (deaminase) from *Escherichia coli*. *Biochem. Biophys. Acta*. **787**:165-173.
- Carl Branden & John Tooze. Introduction to protein structure.New York, USA.1991.
- Comb,D.G., & Roseman,S. (1958). *J.Biol.Chem.* **232**:802-827.
- Datta A.& Natarajan,K.(1993)Molecular cloning and analisis of the NAG1 cDNA coding for glucosamine-6-phosphate deaminase from *Candida albicans*. *J.Biol.Chem.* **268**:9206-9214
- Das,M.&Datta, A.(1982)Purification and some properties of inducible glucosamine 6 phosphate deaminasa from *Candida albicans*. *Biochem. Int.* **5**:735-741.
- Devereux,J., Haeberli, P. & Smithies, O.(1984). A comprehensive set up analysis programs of VAX. *Nucleic Acids Res.* **12**:387-395.
- Fink, A.L. (1987). Acyl group tranfe: the serine proteases. In enzyme mechinisms.(Page, M.I. & Wolliams, A., ends), pp.159-177, The Royal Society of Chemistry, London.
- Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A., Kirkness,E.F., Kerlavage, A. R., Bult,C.J., Tomb,J.-F., Dougherty,B.A., Merrick,J.M., McKenney,K., Sutton,G., FitzHugh,W., Fields,C.A., Gocayne,J.D., Scott,J.D., Shirley,R., Liu,L.-I., Glodek,A., Kelley,J.M., Weidman,J.F., Phillips,C.A., Spriggs,T., Hedblom,E., Cotton,M.D., Utterback,T.R., Hanna,M.C., Nguyen,D.T., Saudek,D.M., Brandon,R.C., Fine,L.D., Fritchman,J.L., Fuhrmann,J.L., Geoghagen,N.S.M., Gnehm,C.L., McDonald,L.A., Small,K.V., Fraser,C.M., Smith,H.O. and Venter,J.C. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269** (5223): 496-512.
- Fuxreiter,M.,Farkas,Ö.& Náráy-Szabó,G.(1995). Molecular modelling of xylose isomerase catalysis: the roll of electrostatics and charge transfer to metals. *Protein Engineering*. **8**:925-933.

- George, D.G., Barker, W.C., Mewes, H.-W., Pfeiffer, F. and Tsugita, A.(1996).The PIR: International Protein Sequence Database. *Nucleic Acids Res.* **24**:17-20.
- Gillevet P., Ally A., Barton F., Brenner S.E., Clark-Whitehead R., Dolan M., Douglas N., Hsu E., Purzcki M.S., Richter B., Russo S., Sartell J., Smith S.W., Wang C., Williams J., Gilbert W. Submitted (13-JUL-1994) to the EMBL/GenBank/DDBJ databases. Mycoplasma capricolum Genome Project, Harvard University, Cambridge.
- Hoffrén, A.-M., Teeri, T.T., Teleman, O.(1995). Molecular dynamics simulation of fungal cellulose-binding domains: differences in molecular rigidity but a preserved cellulose binding surface. *Protein Engineering.* **8**:443-450.
- Homer, K. A., Patel, R. & Beighton, D. (1993). Effects of N-acetylglucosamine on carbohydrate fermentation by *Streptococcus mutans* NCTC 10449 and *Streptococcus sobrinus* SL-1. *Infect Immun.* **61**: 295-302.
- Horjales, E., Altamirano, M.M., Calcagno, M.L., Dauter, Z., Wilson, K., Garratt, R.C., & Oliva (1992). Crystallization and preliminary crystallographic studies of glucosamine-6-phosphate deaminase from *Escherichia coli*. *J.Mol.Biol.* **226**:1283-1286.
- Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and locating errors in these models. *Acta Crystallogr.A.* **47**:110-119.
- Lara-Lemus y Calcagno, M.L. Datos sin publicar sobre bovino y humano.
- Lara-Lemus, R., Libreros-Minotta, C.A., Altamirano, M.M., Calcagno, M.L. (1992) Purification and characterization of glucosamine-6-phosphate deaminase from dog kidney cortex. *Archives of Biochemistry and biophysics.* **297**:213-220.
- Lüthy, R., Bowie, J.V. & Eisenberg, D.(1992). Assessment of protein models with three-dimensional profile. *Nature.* **356**:83-85.
- Milliert, G., Lara-Lemus, R. & M. Calcagno. (1987). Phmetrical determination of the glucosamine-6-phosphate isomerase deaminase reverse reaction. *Biochem. Int.* **14**:511-?
- Midelfort, C.F., & Rose, I.A. (1977). Studies on the mechanism of *Escherichia coli* glucosamine-6-phosphate isomerase. *Biochemistry.* **16**:1590-1596.

- Miyajima,N. (06-JUN-1994) Submitted to the DDBJ/EMBL/GenBank databases. Kazusa DNA Research Institute, Laboratory of Genome Information; 1532-3 Yanauchino, Kisarazu, Chiba 292, Japan.
- Monod,J., Wyman,J., & Chargeoux,J.P. (1965). *J.Mol.Biol.* 12:88-18.
- Nomura, N., Miyajima, N., Sazuka, T., Tanaka, A., Kawarabayashi, Y., Nagase, T.,Ishikawa, K., Seki, T. and Tabata, S. Prediction of the coding sequences of unidentified human genes. II. The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by analysis of cDNA clones from human cell line KG-1. *DNA Res.* 1 (5): 223-229 (1994).
- Oliva,G., Fontes,M.R.M., Garratt,R.C., Altamirano,M.M., Calcagno,M.L. & Horjales,E. (1995) Structure and catalitic mechanism of glucosamine 6-phosphate deaminasa from *Escherichia coli* at 2.1 Å resolution. *Structure.* 3:1323-1332.
- PHD Program:PHDsec:
- Rost, B. & Sander, C.(1994). Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins.*19:55-72.
- Rost, B. & Sander, C.(1993). Improved prediction of protein secondary structure by use of sequence profiles and neural networks. *Proc. Natl. Acad. U.S.A.* 90:7558-7562.
- Rost, B & Sander, C.(1993). Prediction of protein secondary structure at better than 70% accuracy. *J.Mol. Biol.* 232:584-599.
- Levin, Pescarella,Argos & Garnier.(1993). Quantification of secondary structure prediction improvement using multiple alignment. *Prot.Engng.* 6:849-854.
- PHDacc: Rost, B. & Sander, C.(1994).Conservation and prediction of solvent accessibility in protein families. *Proteins.* 20:216-226.
- Palmer, T. (1985). Understanding enzymes. Ellis Horwood Limited. U.K. 411pp.
- Parrington,P.,Swann, K., Shevchenko, V. I., Sesay, A. K. & Lai, F. A. (1996). Calcium oscillations in mammalian eggs triggered by a soluble sperm protein. *Nature.* 379:364 - 368.
- Peri,K.G., Goldie,H. & Waygood,E.B. (1990). Cloning and characterization of the N-acetylglucosamine operon of *Escherichia coli*. *Biochem.Cell.Biol.* 68:123-137.
- Plumbridge,J.A.,(1991). Repression and induction of the nag regulon of *Escherichia coli* K-12; the role of nagC and nagA in maintenance of the uninduced state. *Mol. Microbiol.* 5:2053-2062.

- Plumbridge, J.A., (1989). Sequence of nag BACD operon in *Escherichia coli* K-12 and pattern of transcription within the nag regulon. *Mol. Microbiol.* **3**:505-515.
- Plumbridge, J.A., Cochet, O., Sousa, J.M., Altamirano, M.M., Calcagno, M.L. & Badet, B. (1983). *J. Bacteriol.* **175**:4951-4956.
- Plunkett, G. (22-DEC-1994). TITLE: Direct Submission. JOURNAL: Unpublished Submitted, NCBI gi: 606081, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. GenBank entry U18997. ECOUW67\_71.
- Reizer, J., Ramseier, T.M., Reizer, A., Charbit, A. & Saier, M.H. Jr. (1996). Novel Phosphotransferase genes revealed by bacterial genome sequencing: a gene cluster encoding a putative N-acetylgalactosamine metabolic pathway in *Escherichia coli*. *Microbiology.* **142**: 231-250.
- Rogers, M.J., Ohgi, T., Plumbridge, J., & Soll, D. (1988). Nucleotide sequences of the *Escherichia coli* nagE and nagB genes the structural genes for the N-acetylglucosamine transport protein of the bacterial phosphoenolpyruvate; sugar phosphotransferase system and for glucosamine-6-phosphate deaminase. *Gene.* **62**:197-207.
- Rossmann, M.G., Moras, D. & Olsen, K.W. (1974). Chemical and biological evolution of a nucleotide-binding protein. *Nature.* **250**: 194 -199. □ Rudd, K.E. (1993) *ASM News.* **59**:335-341.
- Stigler, R.-D., Rüker, F., Katinger, D., Elliott, G., Höhner, W., Henklein, P., Ho, J.X., Keeling, K., Carter, D.C., Nugel, E., Kramer, A., Porstmann, T., & Schneider-Mergener, J. (1995). Interact between a FAB fragment against gp41 of human immunodeficiency virus 1 and its peptide epitope: characterization using a peptide epitope library and molecular modeling. *Protein Engineering.* **8**:471-479.
- SSP Program: Solovyev, V.V. & Salamov, A.A. (1994). Predicting  $\alpha$ -helix and  $\beta$ -strand segments of globular proteins. *ABIOS.* **106**:661-669.
- Singh, B. & Datta, A. (1979) Regulation of glucosamine-6-phosphate deaminase synthesis in yeast. *Biochim. Biophys. Acta.* **583**: 28-35.
- Vogler, A. P. & Lengeler, J.W. (1989). Analysis of the nag regulon from *Escherichia coli* and *Klebsiella pneumoniae* and of its regulation. *Mol. Gen. Genet.* **29**: 97-105.

- White, R.J. (1968). Control of amino sugar metabolism in *Escherichia coli* and isolation of mutants unable to degrade amino sugars. *Escherichia coli. Biochem.J.* **106**:847-885.
- Wierenga, R.K., Terpstra, P. & Hol, W. G. J. (1986). Prediction of the occurrence of the ADP-binding  $\beta\alpha\beta$ -fold in proteins, using an amino acid sequence fingerprint. *J.Mol.Biol.* **187**: 101-107.
- Weidanz, J.A., Campbell, P., Delucas, L.J., Jin, J., Moore, D., Roden, L., Yu, H., Heilman, E. & Vezza, A.C. (1995). Glucosamine-6-phosphate deaminase in normal human erythrocytes. *Br. J. Haematol.* **91**:72-79.
- Weidanz, J.A., Campbell, P., Delucas, L.J., Roden, L. & Vezza, A.C. (1995). Glucosamine-6-phosphate deaminase in *Plasmodium falciparum*. *Br. J. Haematol.* **91**:578:586.