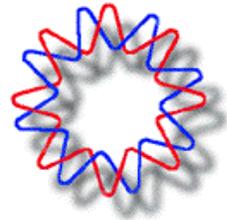




UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

INSTITUTO DE BIOTECNOLOGÍA



- Estudios cristalográficos de ThiN de *Thermotoga maritima*,
Pirococcus furiosus y *Sulfolobus solfataricus*, una probable enzima
análoga a la Tiamina Fosfato Sintasa de *Escherichia coli* -

Tesis que para obtener el grado de
Maestría en Ciencias Bioquímicas
presenta:

Yagul Pedraza Pérez.

Cuernavaca, Morelos. Octubre 2010



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



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.

Este trabajo se realizó bajo la cotutoría de los Drs. Juan Enrique Morett Sánchez y Eduardo Horjales en el Departamento de Medicina Molecular y Bioprocesos del Instituto de Biotecnología. Durante este proyecto el alumno Yagul Pedraza Pérez recibió una beca del CONACyT y un complemento de beca de la DGEP, UNAM.

AGRADECIMIENTOS

A Dios por absolutamente todo.

A mi Familia por todo su apoyo.

A mis Amigos por compartir.

Al Instituto de Biotecnología de la UNAM por su programa de Maestría y Doctorado en Ciencias Bioquímicas.

Al CONACyT y a la DEGEP por la beca otorgada durante este trabajo.

A los Drs. Juan Enrique Morett Sánchez y Eduardo Horjales por su cotutoría.

Al Dr. Enrique Rudiño Piñera por su apoyo durante la realización de este trabajo.

Al Comité Tutorial conformado por la Dra. Rosario Adelaida Muñoz Clares, el Dr. Jesús Aguirre Linares, el Dr. Juan Enrique Morett Sánchez, el Dr. Eduardo Horjales y el Dr. Enrique Rudiño Piñera.

A Sonia Rojas Trejo y Leopoldo Guereca Gurrola por su apoyo técnico en determinado momento del desarrollo de este trabajo. A Leticia Olvera Rodriguez y Maricela Olvera Rodriguez por su apoyo técnico en la clonación de los genes para la sobre expresión de las proteínas.

Al los miembros de jurado: Dra. Adela Rodríguez Romero, Dr. Carlos Daniel Amero Tello, Dr. Joel Osuna Quintero, Dr. Enrique Rudiño Piñera, Dr. Juan Enrique Morett Sánchez por sus valiosas aportaciones a esta tesis.

A la Dra. Adela Rodríguez Romero por las facilidades otorgadas en la colecta de datos en el Laboratorio Universitario de Estructura de Proteínas (LUEP) en la Ciudad de México.

A los integrantes de los laboratorios de los Drs. Juan Enrique Morett Sánchez y Eduardo Horjales y a todos mis compañeros de maestría por el tiempo compartido.

A la Unidad de Cómputo del Instituto y a Alma Lidia Martínez Valle, Manuel Hurtado Ramírez, Roberto Rodríguez Bahena y Jérôme Verleyen.

A la encargada de la Unidad de Biblioteca Shirley Ainsworth.

CONTENIDO

Resumen	1
Introducción	3
Enzimas análogas	4
Antecedentes	9
Marco Teórico	13
Cristalografía de rayos X y estructura de proteínas	13
Cristalización de proteínas	14
Diagrama de fases	15
Cristales y simetría	16
Difracción de cristales de proteína	17
Colecta de datos	18
Métodos para resolver el problema de fase	19
Afinamiento de la estructura	20
Hipótesis y Objetivos	21
Metodología	22
Purificación de las proteínas	22
Expresión de la proteína con seleniometioninas	23
Digestión con trombina para remover la marca de histidina	23
Pruebas de cristalización	24
Colecta y procesamiento de datos de difracción	24
Resultados y Discusión	26
Purificación de las proteínas	26
Expresión de la proteína con seleniometioninas	27
Digestión con trombina para remover la marca de histidina	27
Pruebas de cristalización, colecta y procesamiento de datos de difracción	27
ThiDN_Tm	27
ThiDN_Pf	31
ThiN_Ss	35
Conclusiones	42
Perspectivas	44
Referencias	45

1. RESUMEN

Las enzimas análogas son aquellas que realizan la misma función pero que tienen un origen evolutivo diferente. Debido a que es difícil determinar si dos proteínas son análogas por medio de un análisis de secuencia, es necesario hacer una comparación estructural entre ambas. En el laboratorio del Dr. Enrique Morett se utilizó un método bioinformático, denominado correlación negativa, que permite predecir si dos proteínas son análogas con base en su presencia o ausencia dentro de los genomas totalmente secuenciados. Con esta estrategia se encontraron varios candidatos a ser análogos dentro de la vía de síntesis de tiamina, entre ellos la Tiamina Fosfato Sintasa (TPS) de *Thermotoga maritima*. La TPS de *Bacillus subtilis*, de estructura conocida, tiene un plegamiento (alfa/beta)₈ del tipo barril TIM; en cambio, la TPS de *T. maritima* pertenece a un grupo cerrado de proteínas que no tiene un homólogo con estructura conocida, por lo que resulta de interés conocer el plegamiento de esta proteína y determinar si efectivamente es una análoga de la TPS de *Bacillus subtilis*. Dentro del grupo de proteínas al que pertenece la TPS de *T. maritima* están las de *Pyrococcus furiosus* y *Sulfolobus solfataricus* cuyo ADN se encontraba disponible en el laboratorio, por lo que se clonaron los genes para sobreexpresar las proteínas. Las TPS de *T. maritima* y *P. furiosus* se encuentran fusionadas con otra proteína de la misma vía de síntesis de tiamina, llamada ThiD y que tiene estructura conocida, formando una proteína híbrida que se denominó ThiDN. La enzima de *S. solfataricus* no se encuentra fusionada.

El objetivo del presente trabajo fue obtener cristales de las proteínas ThiDN de *T. maritima*, *P. furiosus* y ThiN de *S. solfataricus* para hacer estudios de difracción de rayos X que permitieran evaluar la posibilidad de determinar la estructura de estas probables enzimas análogas. Se purificaron las proteínas a homogeneidad usando una marca de histidinas y filtración en gel. Las pruebas de cristalización se hicieron por el método de difusión de vapor en gota colgante. En una primera etapa se usaron los kits de cristalización comerciales para evaluar las condiciones recurrentes de pH, fuerza iónica, concentración y tipo de precipitante, y con ello diseñar nuevas matrices para optimizar la obtención de cristales útiles para estudios de difracción.

Se obtuvieron cristales de las tres proteínas y se colectaron patrones de difracción de varios de ellos. Para los cristales difractados las proteínas ThiDN de *T. maritima* y *P. furiosus* se intentó resolver el problema de fase por medio de Reemplazo Molecular usando un modelo de búsqueda basado en la estructura conocida de ThiD; sin embargo, sólo se obtuvo una solución parcial. Con los métodos de SAD/MAD y SIR/MIR no se logró ubicar los átomos pesados debido, muy probablemente, a una alta movilidad de los dominios de estas proteínas. Para los cristales difractados de la proteína ThiN de *S. solfataricus* se intentó resolver el problema de fase por medio de SIR/MIR. Los derivados de átomos pesados, sólo o en combinación con otros, no presentaron señal suficiente para determinar las posiciones de los mismos.

En conclusión, en este trabajo, se obtuvieron cristales de las proteínas ThiDN de *T. maritima*, ThiDN de *P. furiosus* y ThiN de *S. solfataricus* con calidad suficiente para hacer estudios cristalográficos y se colectaron patrones de difracción de varios de ellos. El procesamiento de estos datos no permitió resolver el problema de fase.

2. INTRODUCCIÓN

El metabolismo de todos los seres vivos está sustentado por una extensa red de enzimas que son capaces de llevar a cabo las reacciones químicas necesarias de síntesis y degradación de las moléculas para el correcto funcionamiento de las células. Durante millones de años estas enzimas se han ido especializando para llevar a cabo, de manera cada vez más *eficiente*, alguna reacción química en particular. La optimización de la reacción catalítica no está orientada a ser más rápida o específica cada vez, sino más *apta* dentro del entorno bioquímico global de la célula, según las necesidades metabólicas de esta.

Existe por ejemplo, la succinato deshidrogenasa del ciclo de Krebs, que es un complejo enzimático asociado a membrana y un componente de la cadena respiratoria de todos los organismos aerobios, cataliza la reacción de succinato a fumarato en el ciclo del ácido tricarboxílico y transfiere los electrones directamente a quinonas en la membrana (Janssen, *et al*, 1997; Oyedotun, Lemire, 2004). Sin embargo, esta enzima es notablemente lenta, tiene una actividad molecular de 1150 moléculas de producto por minuto (Lehninger, 1997). Para realizar una función óptima en la célula compensa su baja actividad con una alta concentración dentro de la mitocondria.

En el otro extremo, existen algunas enzimas que alcanzan velocidades de reacción muy altas, al límite sólo de la velocidad de difusión de los sustratos, como por ejemplo, la anhidrasa carbónica, con una actividad molecular de 36 000 000 moléculas de producto por minuto (Lehninger, 1997). Esta metaloenzima cataliza la reacción reversible de dióxido de carbono y agua en bicarbonato y protones. Se encuentra presente en especies de los tres dominios de la vida; archa, bacteria y eucaria. Participa en un variado número de procesos viales como gluconeogénesis, ureagénesis, lipogénesis, biosíntesis de ciertos aminoácidos, y la biosíntesis de nucleótido de pirimidina, entre otros (Jude, *et al*, 2006). Su función principal en plantas y algas es proporcionar dióxido de carbono utilizado en la fotosíntesis mientras que en animales ayuda a regular el pH y a expulsar en dióxido de carbono.

Otras enzimas catalizan reacciones químicamente muy simples pero en extremo importantes, como la catalasa, que se encarga de transformar el peróxido de hidrógeno en oxígeno molecular y agua evitando el daño celular al prevenir la formación de radicales libres (Barynin, *el al*, 2001), o bien, participan en reacciones muy complejas como el caso de algunas enzimas de la ruta de síntesis de Tiamina que presentan una química sin precedentes (Begley, *et al*, 1999; Settembre, *et al*, 2003;

Jurgenson, *et al*, 2009).

Las proteínas son muy versátiles en cuanto a su evolución, pues su estructura tridimensional es tolerante a los cambios que permite en la secuencia de aminoácidos antes de presentar cambios estructurales significativos, de tal forma que dos secuencias pueden divergir notablemente de su ancestro común sin perder la función inicial restringiendo los cambios de los aminoácidos implicados en la catálisis (Todd, *et al*, 2002). De hecho, hay muchas familias de proteínas en donde se ha encontrado una conservación absoluta de los residuos funcionales, por ejemplo, la familia de las enzimas dependientes de piridoxal fosfato, a nivel de secuencia han divergido mucho, sin embargo, todas ellas tienen unos residuos de ácido aspártico (Asp) y lisina (Lys) invariantes en el sitio activo que son los encargados de unir al piridoxal fosfato (Todd, *et al*, 2002).

Pese a la alta especificidad que existe en las enzimas por un determinado sustrato y la eficiencia catalítica para una determinada reacción química, existe también una notable flexibilidad en el sitio activo que permite a varias enzimas con actividad catalítica muy parecida o incluso la misma reacción química, tener un sitio activo diferente en donde distintos grupos funcionales, no conservados respecto a su posición en la estructura primaria, tienen el mismo papel mecánico en la catálisis, es decir, dos o más soluciones a la misma *interrogante* catalítica que han evolucionado dentro de una sola familia de proteínas (Todd, *et al*, 2002). También puede darse el caso de que una enzima tenga más de un tipo de catálisis (Morett, *et al*, 2008; Yankovskaya, *et al*, 2003; Lim, *et al*, 2009), o inclusive diferentes funciones dependiendo de dónde se encuentre localizada como las llamadas “moonlighting proteins” (Jeffery, 2003).

Todo esto no sólo remarca la increíble versatilidad de las enzimas, sino que trae consigo una dificultad técnica, pues con la secuenciación de genomas completos la asignación de función para una enzima se basa, principalmente, en la similitud con otras secuencias ya que es imposible determinar experimentalmente la función de miles de genes reportados constantemente (Hegyi, Gerstein, 1999), esto ocasiona que las bases de datos vayan acumulando errores de anotación (Hegyi, Gerstein, 1999; karp, *et al*, 1999).

Enzimas Análogas

Existe también el caso de las enzimas análogas, las cuales pueden realizar la misma reacción química que otra sin estar relacionada evolutivamente, es decir, que tiene un origen distinto y por lo

tanto otra estructura (Koonin, Mushegian, 1996; Krishna, Grishin, 2004). Así que para determinar correctamente si dos o más enzimas son verdaderas análogas es necesario analizarlas desde su estructura tridimensional, pues la posibilidad de que dos enzimas tengan diferente origen evolutivo sólo puede descartarse cuando éstas tienen un plegamiento diferente (Galperin, *et al*, 1998).

Según su origen evolutivo las proteínas se pueden clasificar en cuatro grupos:

Proteínas **homólogas**. Son proteínas que han evolucionado a partir de un ancestro en común.

Proteínas **ortólogas**. Son proteínas homólogas que llevan a cabo la misma función en distintos organismos.

Proteínas **parálogas**. Son proteínas que tienen un ancestro en común pero distinta función. Estas proteínas surgen por eventos de duplicación génica y posterior divergencia de función.

Proteínas **análogas**. Son proteínas que llevan a cabo la misma función pero tienen un origen evolutivo distinto, es decir, no están relacionadas filogenéticamente (Koonin, 1996)(Fig. 2-1).

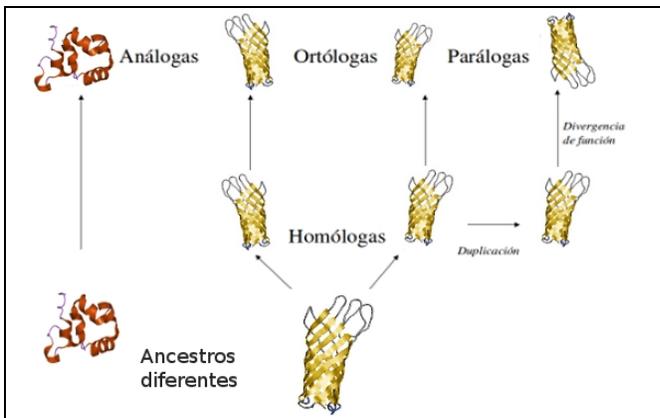


Figura 2-1. Esquema que muestra la relación evolutiva entre las proteínas. Cuando dos proteínas tienen un ancestro en común se denominan Homólogas, dos proteínas que llevan a cabo la misma función en distintos organismos se denominan Ortólogas. Si el gen que codifica para una proteína se duplica evolucionará de manera independiente y dará origen a una proteína con función diferente, llamada Paróloga y está presente en el mismo organismo. Dos proteínas que tienen un origen evolutivo diferente pero la misma función se denominan Análogas y tienen plegamiento diferente.

Hace unas décadas, la presencia de enzimas no relacionadas filogenéticamente pero que realizaban la misma función se consideraba rara y excepcional. Actualmente, con el incremento de los genomas secuenciados totalmente y los esfuerzos por reconstruir las rutas metabólicas completas se han ido encontrando más casos de enzimas análogas, tanto en las rutas del metabolismo central, como en otras rutas, por ejemplo, en la manipulación de ácidos nucleicos y en la transducción de señales (Galperin, *et al*, 1998).

Un caso claro y muy estudiado de enzimas análogas es el de las serin-proteasas. Estas enzimas hidrolizan algún enlace peptídico de un polipéptido para producir dos péptidos más pequeños. Esta

reacción procede en dos pasos: en el primero se produce un enlace covalente entre el oxígeno de la serina catalítica y el carbono 1 del sustrato, generando un estado intermediario en donde el C1 tiene una geometría tetragonal, en contraste con la forma plana del enlace peptídico. Durante este paso, un péptido permanece unido a la enzima mientras que el otro producto de la reacción es liberado. En el segundo paso de la reacción, el intermediario unido a la enzima es hidrolizado por una molécula de agua para liberar el segundo péptido con un grupo carboxilo completo y para restaurar el grupo OH de la serina catalítica (Branden, Tooze, 1999; Hedstrom, 2002).

Miembros que pertenecen a la misma familia de proteasas usualmente están relacionados evolutivamente, pero hay excepciones. En dos casos, la quimiotripsina y la subtilisina, tanto la secuencia como la estructura terciaria son diferentes y, sin embargo, los residuos catalíticos de ambas enzimas están en la misma posición espacial relativa. Este es un caso claro de convergencia evolutiva, en donde partiendo de ancestros no relacionados se ha llegado a la misma solución estructural para un mecanismo catalítico particular (Branden, Tooze, 1999).

Otro caso de enzimas análogas lo podemos encontrar en la anhidrasa carbónica (CA, por sus siglas en inglés), que cataliza la conversión rápida de dióxido de carbono y agua a bicarbonato y protones, en donde se han reportado hasta la fecha cuatro plegamientos diferentes para esta enzima pertenecientes a cuatro familias denominadas alfa-CA, beta-CA, gama-CA y zeta-CA (Tripp, 2001; Lane 2005; Supuran, 2008), representados en la figura 2-2.

Las alfa-CA se encuentran principalmente en mamíferos y bacterias, tienen un átomo de Zn coordinado por tres histidinas y son estrictamente monoméricas. Las beta-CA, presentes en plantas, algas y arqueas, tienen dos cisteínas y una histidina coordinando al Zn, son estrictamente dímeros o multímeros múltiplos de dos. Las gama-CA son triméricas y también tienen tres histidinas coordinando al Zn, sólo que dos de ellas pertenecen a un monómero y la otra al monómero siguiente y se les considera dentro de las CA más antiguas. Estas tres familias tienen dos cosas en común, la presencia de un átomo de Zn en el sitio activo y un mecanismo catalítico de dos pasos: primero se da un ataque nucleofílico del Zn (en forma de Hidróxido) al dióxido de carbono y posteriormente una regeneración del sitio activo por medio de una molécula de agua. Sin embargo, en cada familia existen diferencias en los aminoácidos que participan en este mecanismo (Tripp, *et al*, 2001).

Por otra parte, la zeta-CA tiene Cd en vez de Zn, coordinado por dos cisteínas y una histidina

(similar a las beta-CA) y fue descubierta en una diatomea marina, a la fecha este el único caso conocido en la naturaleza de una enzima capaz de usar Cadmio (Lane, *et al*, 2005; Xu, *et al*, 2008) (Fig. 2-2).

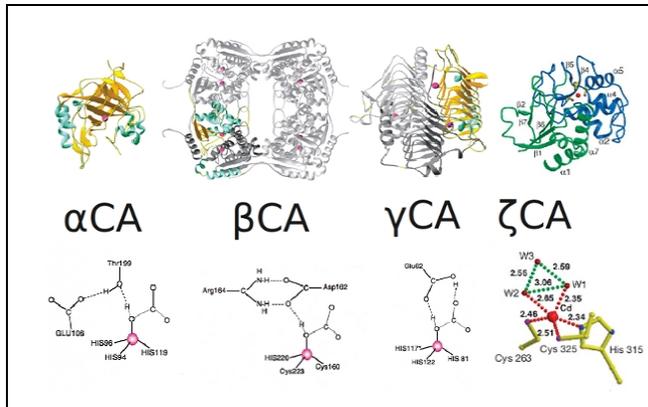


Figura 2-2. Comparación del plegamiento y del sitio activo de las enzimas análogas de la anhidrasa carbónica (CA). Adaptado de Liljas, 2000 y Xu, 2008.

De esta manera podríamos comparar un número cada vez mayor de casos en donde una determinada actividad catalítica se encuentra presente en más de un plegamiento. Actualmente el listado más completo de enzimas análogas lo presenta Omelchenko y colaboradores (Ref. 1, trabajo no publicado) en una actualización del trabajo hecho por Galperin y colaboradores (Galperin, *et al*, 1998). En este trabajo se encontraron 79 casos en los que, por medio de una comparación estructural, se ha podido demostrar que proteínas pertenecientes a un determinado número EC (Enzyme Commission) tienen diferente plegamiento. En 57 de estos casos existen dos plegamientos para una reacción química en particular, 17 de estos casos contienen tres plegamientos para el mismo número EC, 4 casos presentan cuatro plegamientos diferentes y 1 caso tiene cinco plegamientos para el mismo número EC (Ref. 1).

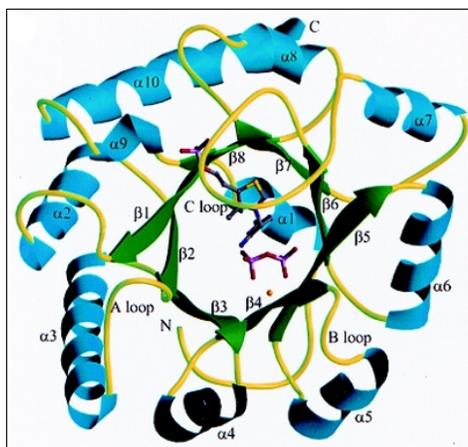


Figura 3-2. Estructura cristalográfica de la TPS de *Bacillus subtilis*. Al centro se puede apreciar la tiamina fosfato, una molécula de pirofosfato y un átomo de magnesio. Adaptado de Chiu, 1999.

Es necesario mencionar que la distribución de estos casos de enzimas análogas reportados no es homogénea en los 6 tipos de catálisis determinados por la EC y que está en función de la cantidad de estructuras conocidas para cada número EC, de tal manera que correlaciona una alta presencia de enzimas análogas en los números EC que tienen más proteínas con estructura conocida.

Para conocer el número total de enzimas análogas en la naturaleza sería necesario determinar el plegamiento de las proteínas de todas las familias de enzimas conocidas, lo que representa una ardua labor. Debido a que las enzimas análogas tienen diferente plegamiento, también tienen diferentes propiedades fisicoquímicas, por ejemplo, algunas son más resistentes al calor o a determinados solventes, y esta capacidad también puede ser explotada en procesos industriales en los que ciertas enzimas resultan sensibles. En la industria farmacéutica, son de gran interés las proteínas análogas como blancos potenciales para el diseño de fármacos altamente específicos, ya que las enzimas de los organismos patógenos, al tener una estructura diferente a la estructura de la proteína del hospedero, se pueden bloquear a nivel de sitio activo con algún fármaco sin que la función de la enzima del paciente se vea afectada (Temperini, *et al*, 2008).

Existen diversos consorcios que han venido trabajando durante años en proyectos de determinación de estructura de todas las proteínas diferentes de un determinado organismo, es decir, de su Proteoma. Con ello se pretende tener un banco de datos de estructura de proteínas cada vez más completo que permita entender mejor los mecanismos bioquímicos que existen dentro de la célula. Evidentemente, en estos proyectos las proteínas con mayor interés en determinar su estructura son aquellas que pertenecen a familias sin homólogos cercanos con estructura conocida.

3. ANTECEDENTES

El estudio de las proteínas, a nivel de su estructura primaria, está basado en análisis de secuencias, y los algoritmos están diseñados para analizar secuencias homólogas. No existen algoritmos que permitan predecir, a partir de su secuencia, qué proteínas son análogas, de tal manera que estas últimas se han venido descubriendo como consecuencia de caracterizar rutas bioquímicas o de manera accidental, y en un principio se pensaba que eran casos poco comunes en la naturaleza.

Recientemente, con la disponibilidad de genomas totalmente secuenciados, se ha visto un fenómeno de ausencia de genes en ciertos organismos. En un estudio de genómica comparativa entre los dos primeros genomas secuenciados de *Haemophilus influenzae* y *Mycoplasma genitalium*, se encontró once casos donde funciones celulares indispensables están codificadas por genes no ortólogos (Koonin, Mushegian, 1996). Lo que sugiere que el fenómeno de enzimas análogas es más distribuido de lo que se pensaba en un principio. El mecanismo más probable para la evolución de las enzimas análogas parece ser el reclutamiento de enzimas existentes que pueden tener nuevas funciones debido a la virtud de cambiar su mecanismo catalítico o la especificidad por el sustrato (Galperin, *et al*, 1998). En un trabajo de actualización de estos resultados se ha podido probar la relación de analogía para 79 casos (Ref. 1).

Por otra parte, en un estudio hecho en el laboratorio del Dr. Enrique Morett, se utilizó un método bioinformático de correlación negativa para buscar sistemáticamente enzimas análogas, en la vía de síntesis de tiamina, en varios genomas totalmente secuenciados. En este trabajo se encontraron 7 candidatos a ser análogos de los cuales 3 conforman una tripe correlación negativa. Se verificó la función por complementación génica de cuatro casos y se caracterizó bioquímicamente uno de ellos, denominado ThiN, candidato análogo de la Tiamina Fosfato Sintasa (TPS, por sus siglas en inglés), presente en bacterias y algunos eucariontes, que cataliza la formación de tiamina fosfato y que es codificada por el gen *thiE* (Morett, *et al*, 2003). También se ha reportado el caso de que la familia de proteínas *YjbQ* es capaz de complementar la función de TPS en cepas de *Escherichia coli* que han sido suprimidas del gen *thiE*, formando así un tercer análogo de esta enzima (Morett, *et al*, 2008).

La tiamina es una vitamina sintetizada por la mayoría de los procariontes y por algunos eucariontes como levaduras y algunas plantas. En todos los casos un tiazol y una pirimidina son sintetizados en ramas separadas de la ruta metabólica y unidos por la TPS para formar tiamina

fosfato. La formación de tiazol requiere al menos seis proteínas y la formación de la pirimidina requiere dos. La levadura, por su parte, sólo utiliza un gen para sintetizar tiazol y otro para sintetizar pirimidina (Jurgenson, *et al*, 2009).

La síntesis del tiazol (4-metil-5-β-hidroxieltiazol o THZ) se lleva a cabo en tres pasos, primero el gliceraldehído 3-fosfato y el piruvado son unidos por la 1-Desoxi-D-Xilulosa Fosfato Sintasa (Dxs) para dar 1-deoxi-D-xilulosa fosfato. En segundo lugar, la proteína ThiS sufre un proceso de adenilación catalizado por la proteína ThiF formando un complejo ThiS/ThiF seguido de una transferencia de azufre por las proteínas ThiI y IscS para dar como resultado una proteína ThiS con un tiocarboxilato en su región C-terminal, ThiS-COSH. El tercer paso consiste en la formación de dehidroglicina a partir de la glicina catalizado por la proteína ThiO (en *B. subtilis*), o a partir de tirosina catalizado por ThiH (en *E. coli*). Estos tres compuestos son unidos por ThiG en una sola molécula, un tautómero (isómero estructural que difiere sólo en la posición de un grupo funcional) del thiazol fosfato carboxilato. La proteína ThiI isomeriza este tautómero para dar el thiazol fosfato carboxilato (THZP) que se unirá a la pirimidina (Jurgenson, *et al*, 2009)(Fig. 3-1).

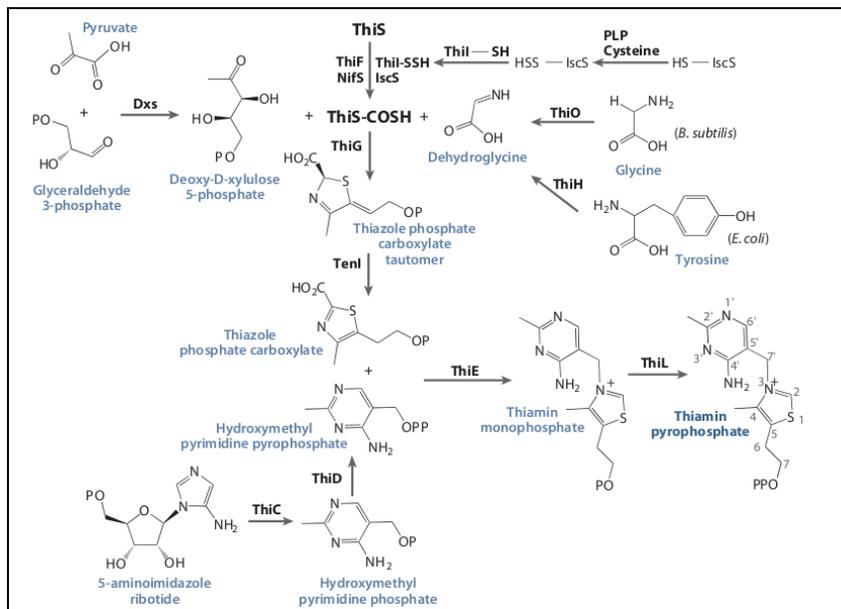


Figura 3-1. Esquema de la ruta completa de biosíntesis de tiamina en bacterias. Tomado de Jurgenson, *et al*, 2009.

La síntesis de la pirimidina (4-amino-5-hidroxi-2-metilpirimidina fosfato o HMP) es llevada a cabo por la proteína ThiC en un complejo rearrreglo del 5-aminoimidazol ribótido para dar HMP que luego es pirofosforilado por la ThiD en hidroximetil pirimidina pirofosfato (HMPP).

Finalmente, el HMPP y el THZP son condensados por la Tiamina Fosfato Sintasa (TPS) para dar tiamina fosfato que posteriormente es fosforilada por la ThiL a tiamina pirofosfato, que es la forma biológicamente activa. (Jurgenson, *et al*, 2009). Actualmente se conoce la estructura cristalográfica de casi todas estas enzimas y han sido estudiadas dentro el contexto bioquímico y genético de la síntesis de tiamina (Settembre, *et al*, 2003; Jurgenson, *et al*, 2009).

Las enzimas análogas que se encontraron en el trabajo del Dr. Morett son: para la TPS, la proteína codificada por el gen TM_0790 de *Thermotoga maritima* (ThiN); para la ThiL, una proteína denominada Thi80; para ThiG, la proteína Thi4; y para ThiC, la proteína TenA. También se mostró que las proteínas Thi4, ThiH y ThiO forman una triple correlación negativa, es decir, que cada una de ellas puede *desplazar* funcionalmente a las otras dos. Con esta metodología se pueden comparar enzimas para predecir su función sin necesidad de tomar en cuenta la similitud de su secuencia, resultando así un método útil para la búsqueda sistemática de enzimas análogas. La proteína ThiN, en *T. maritima* y algunos otros microorganismos, como por ejemplo *Pyrococcus furiosus*, se encuentra fusionada con la ThiD y entonces es denominada ThiDN. Esta ThiDN de *T. maritima* fue purificada y se le detectó actividad de síntesis de tiamina *in vitro* (Morett, *et al*, 2003).

La TPS de *B. subtilis* tiene un plegamiento de barril (alfa/beta)₈ tipo Triosa Isomerasa Monofosfato (TIM), su estructura ha sido determinada en complejo con tiamina, representada en la figura 3-2, y se ha estudiado su mecanismo de catálisis (Chiu, *et al*, 1999; Reddick, *et al*, 2001), en cambio, la ThiN de *T. maritima* pertenece a una familia de proteínas que no tiene un homólogo con estructura conocida.

Por esta razón y con la finalidad de comprender mejor el funcionamiento de las enzimas análogas y en particular del caso de la TPS y la ThiN se planteó un proyecto de colaboración entre los laboratorios del Dr. Enrique Morett y del Dr. Eduardo Horjales para determinar la estructura de la enzima ThiDN de *T. maritima* por cristalografía de rayos X. En el laboratorio del Dr. Morett se tenía disponible ADN genómico de *Pyrococcus furiosus* y *Sulfolobus solfataricus* así que también se clonaron los genes de esta enzima presente en dichos microorganismos.

En la figura 3-3 se muestra un alineamiento de secuencia de las proteínas ThiDN_Tm,

ThiDN_Pf y ThiN_Ss. El porcentaje de similitud correspondiente a los dominios N es de 51 %.

ThiDN_Tm	-----MVLVVGFDPSGGAGIIQDVKLSALGVKTHAVISALTVQENRVSFNFRDWE	54
ThiDN_Pf	MKRLVKALTITAGSDSGGAGIEADLKTFTAFGVHGLVAITSVTAQNTQAVTAIHDIPPE	60
ThiN_Ss	-----	
ThiDN_Tm	EMRKEIEVLTPPRVIKVG---LSAPETVKRLREMFP--DSAIWNVVLESSSGFGFQDP	108
ThiDN_Pf	VVAEQIRAVAEIDIGVDAAKTGMLSNAEIIKAVAKTVKNYDFPLVVDVPMIAKSGAPLLRE	120
ThiN_Ss	-----	
ThiDN_Tm	EEVKKFVEY---ADYVILNSEEAKLG-----EYNNFIVTGGH	143
ThiDN_Pf	DAVDALIKYILPLATLVTPNRFEAELSGMKIRSVEDAKIVAKKIAEETGAEAVIVKGGH	180
ThiN_Ss	-----	
ThiDN_Tm	EKGN--TVKVKYRD---FVFEI PRVPGEFHGTGCAFSSAVSGFLAMSYPVEEAIRSAMELL	199
ThiDN_Pf	IEGEEAVDVLVHNGTFREYRAPKVEGCTHGTGCSFSAAITANLAKGLGLEEAIEVAKKFI	240
ThiN_Ss	-----	
ThiDN_Tm	KK-----ILERSSGVVETEKLRLRDWYRDTLNTLDEILPEFLEIG--HLTVPVQGVN	250
ThiDN_Pf	TLGIAMGHRIGHGHCVPVNSAWIEIPA EKWR IYEELTNAVREFESINP-VRLIPEVGTNF	299
ThiN_Ss	-----MQDESEREYVLKRLKEAVDIFVSNERAYLLIPEVRTNI	38
	: : : * : : * . :*** *	
ThiDN_Tm	SY--ALPWAKNEFEVKGKFPGRIRLKEGKAVAVSCASFKDRSHTARMAVTMMRYHPHMRCV	308
ThiDN_Pf	VYSLPLPYARSTKDVAGVKGRIVKYGNVKA VGPVEFGASDHLARAVLTYMRFYPEYRSA	359
ThiN_Ss	GY--AVSNAADANDVAATPGRLTTAFKKVIYCMPLAFGASDHVARVILTMKHDRNIRSA	96
	* .. * . :*. . **: .. * . * * : * * . . * . .	
ThiDN_Tm	VNVRYEREYVER---AKKRGKLVFHYDRSKEPKEVQEKEGQSMVWMIEQAIAELKSPDP	364
ThiDN_Pf	INIRYSREIIEEII EIAQERGFVVSFYDRREEPEEIKAKEGATIPWGIETAIKRIKERPD	419
ThiN_Ss	INLKYYREVIEK-----LPSQDLCIFDRSSEPKEVKYREHSTMNFMVDSCIRKLSKVPN	150
	:: ** :*. . . : ** .**::: * : : : : * *	
ThiDN_Tm	VIYDEGWGKEAMIRVFGRNPKVLEKIKLMVRE-----	398
ThiDN_Pf	IIYHLGDVGKEPMLLVFGRNPREVLEKIKMLI-----	451
ThiN_Ss	YIVDLGDYGKEPSLFILDRDPVTVVNSLELLKYISQDIL	190
	* . * ***. : :*: * * : * :	

Figura 3-3. Alineamiento de secuencias de las proteínas ThiDN_Tm, ThiDN_Pf, ThiN_Ss. La interrupción (gap) más grande se encuentra en la posición 133 de la ThiDN_Tm, ubicada en el dominio D, y que corresponde a la pérdida de una hélice alfa en relación con las demás ThiD. El alineamiento fue hecho con CLUSTAL W, Versión 2.0.12, disponible en <http://www.ebi.ac.uk/Tools/clustalw2/index.html>

4. MARCO TEÓRICO

Cristalografía de rayos X y estructura de proteínas.

La cristalografía de rayos X es un método que permite determinar el arreglo tridimensional de los átomos dentro de un cristal. Los rayos X son una forma de radiación electromagnética de alta energía y pequeña longitud de onda, del orden del tamaño de los átomos, de tal manera que puede interactuar con ellos. Cuando un haz de rayos X incide sobre un cristal difracta en direcciones específicas y genera un *patrón de difracción* que depende de la simetría interna del cristal. A partir del ángulo y de la intensidad de estos rayos difractados se puede reproducir una imagen tridimensional de la densidad de los electrones dentro del cristal y a partir de esta densidad se puede determinar la posición promedio de los átomos, sus enlaces químicos, su desorden relativo y alguna otra información dependiendo de la calidad de los cristales y la medida de su difracción.

Para determinar la estructura de una proteína primero se debe obtener un cristal de ella que difracte con la calidad suficiente para hacer estudios de difracción de rayos X, posteriormente se debe coleccionar un patrón de difracción completo y resolver el *problema de fases*, esto es, conocer de qué punto del espacio (dentro del cristal) proviene cada rayo X difractado. Después se debe generar un mapa de densidad electrónica a partir de la cual se construirá el modelo tridimensional de la proteína (Fig. 4-1).

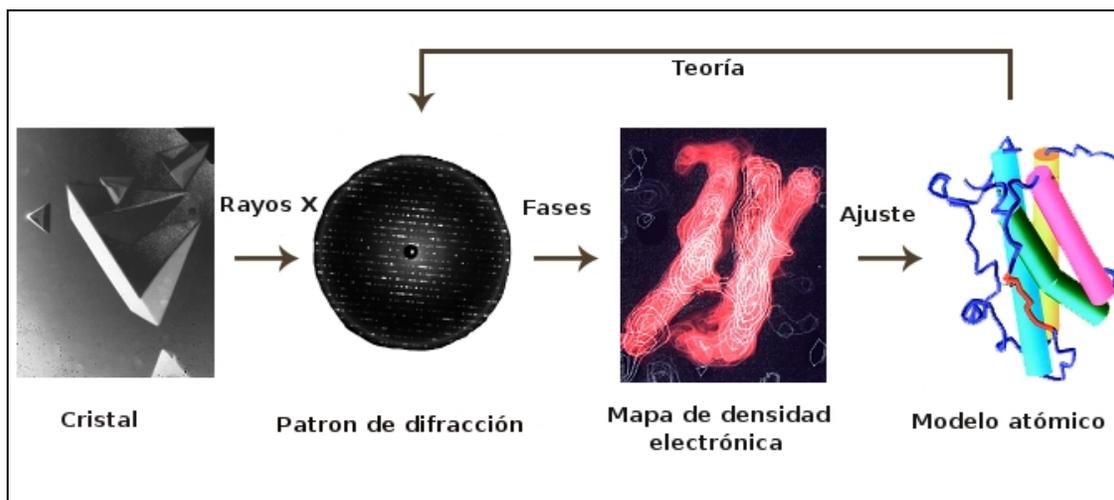


Figura 4-1. Esquema del flujo de trabajo en la determinación de la estructura de una proteína por medio de cristalografía de rayos X. Primero se debe obtener un cristal de la proteína de interés (a), del cual se debe obtener un patrón de difracción (b), una vez que se resuelve el problema de fase se puede calcular un mapa de densidad electrónica (c) que permitirá construir un modelo de la proteína (d).

Cristalización de proteínas.

Para obtener un cristal de proteína es necesario tener una muestra de la proteína altamente pura y que haya sido manipulada lo menos posible ya que durante el trayecto de su purificación algunas proteínas pueden perder algunas de las propiedades necesarias para su función biológica. Una vez que la proteína ha sido purificada en homogeneidad, se debe colocar en un sistema cerrado tal que conforme alcance su estado de equilibrio vaya ordenando a las moléculas de proteína para formar el cristal. La cristalización es en gran medida un procedimiento de ensayo y error en el que la proteína es lenta y ordenadamente precipitada (Drenth, 1999).

Dentro de los métodos de cristalización de proteínas más usados se encuentra el de difusión de vapor en gota colgante. En este método, se coloca una pequeña muestra de la proteína pura, típicamente uno o dos microlitros de proteína a alta concentración, en un cubre objetos previamente siliconizado para favorecer la cohesión molecular y formar una gota semiesférica en vez de una gota esparcida por el cristal. Esta gota con proteína se mezcla con un volumen igual de líquido precipitante, de esta manera ambos quedan diluidos por la mitad. Dentro del mismo sistema se coloca un volumen mucho mayor (típicamente un mililitro) de líquido precipitante que funciona como reservorio y contra el cual se equilibrará la gota con proteína-precipitante. Una variante de este método es el de difusión de vapor en gota apoyada en donde la muestra de proteína-precipitante se coloca en un “puente”, permitiendo así que la gota sea mucho más grande retardando el tiempo que tardará en equilibrarse con el reservorio. Otra variante del método de difusión de vapor es el denominado tipo sándwich en donde la muestra de proteína-precipitante se coloca en medio de dos porta objetos ligeramente separados, de esta manera se reduce significativamente la superficie expuesta por la que la gota equilibrará con el reservorio (Drenth, 1999)(Fig. 4-2).

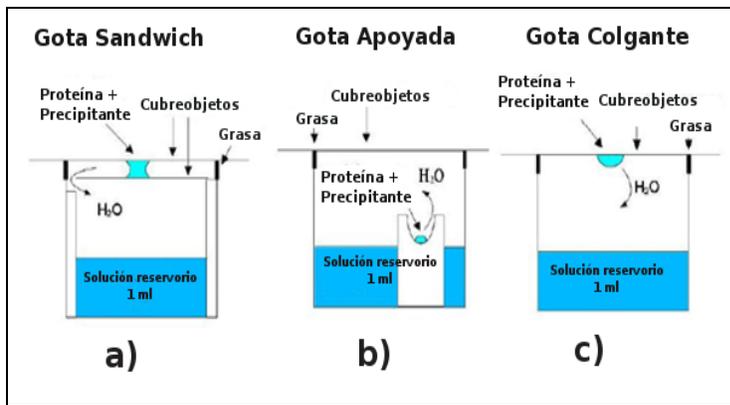


Figura 4-2. Esquema de los métodos de cristalización de difusión de vapor en gota sándwich (a), en gota apoyada (b) y gota colgante (c).

Otros métodos para cristalizar involucran diálisis, en donde se puede usar un gradiente de fuerza iónica como agente precipitante a través de una membrana de diálisis, aceites para reducir aún más la velocidad de equilibrio del sistema, etc., de tal forma que cada método favorecerá una velocidad diferente de equilibrio del sistema (Drenth, 1999; Bergfors, 1999).

La función del líquido precipitante es siempre llevar a la proteína del estado soluble a uno de supersaturación para que se pueda dar el crecimiento de cristales; su naturaleza y concentración serán diferentes para cada proteína y no es posible saber de antemano cuál precipitante y a qué concentración cristalizará una proteína determinada.

Diagrama de fases.

Para obtener cristales grandes, es necesario encontrar condiciones donde aparezcan pocos cristales en la gota; de esta manera, la proteína soluble estará disponible para que los cristales crezcan, en vez de ser usada para formar más cristales. Para conocer estas condiciones, es necesario valerse de lo que se conoce como *diagrama de fases* (Fig. 4-3). En este diagrama se representa un gradiente de concentración de precipitante contra un gradiente de concentración de proteína. En él destacan cuatro zonas o *fases*: la *fase soluble* es la región de la gráfica en donde la proteína se mantiene soluble dentro de un rango de concentración de precipitante. La *fase de crecimiento* (o *metaestable*) es la zona en la que se ve más favorecido el crecimiento del cristal. En la *fase de nucleación*, la concentración de ambos componentes es tal que lo que se ve favorecido es la formación de núcleos de cristalización. Por último, está la *fase de precipitado*, que es donde las concentraciones de precipitante y proteína son tan altas que esta última precipita. El diagrama es específico para cada proteína y para cada precipitante, y no es extrapolable a otros casos.

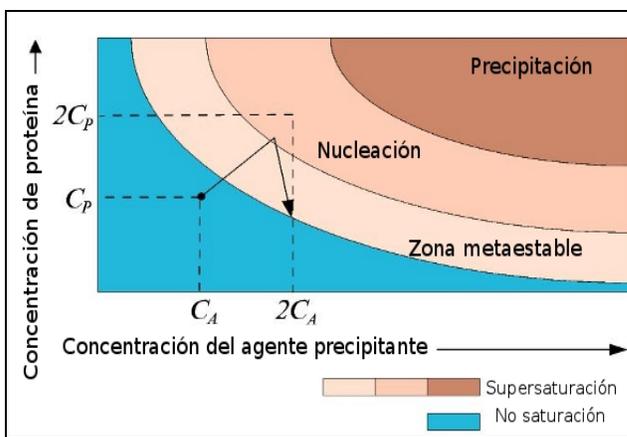


Figura 4-3. Esquema del diagrama de fase. La formación de un cristal de proteína comienza en la zona de baja saturación, cuando la proteína está en solución (C_p , C_A). Durante el equilibrio de la gota la concentración de precipitante aumenta, así como la de la proteína, entonces el sistema pasa a la condición de nucleación ($2C_p$, $2C_A$). Al formarse los núcleos, la concentración de proteína libre disminuye y el sistema pasa a la fase metaestable donde los núcleos comienzan a crecer hasta que llegan a un equilibrio.

Obtener un diagrama de fase completo para una determinada proteína puede llegar a ser muy

laborioso y se consume mucha proteína, sobre todo para determinar los límites entre la zona soluble y la zona metaestable (Asherie, 2004), sin embargo, se puede hacer lo que se denomina un *diagrama de fases de trabajo* (Saridakis, Chayen, 2003). En este diagrama sólo se determinan los límites entre la zona de nucleación y la metaestable. Para conseguir esto se hace un seguimiento de las concentraciones de precipitante, desde las que favorecen la aparición de muchos microcristales hasta donde la gota permanece soluble, a una alta concentración de proteína.

Los cristales de proteína, a diferencia de los cristales de sal o de moléculas pequeñas, son muy sensibles a la manipulación, a la deshidratación y a los cambios de temperatura, por lo que se debe tener mucho cuidado al trabajar con ellos. Para manipularlos se usan *loops* o asas fabricados con materiales orgánicos que minimizan el impacto al tocarlos.

Cristales y simetría.

Una característica muy importante de los cristales es que tienen simetrías, de tal manera que realizando operaciones de rotación o traslación determinados se puede reconstruir el cristal. La repetición periódica por la que se describe la estructura interna de los cristales viene representada por un conjunto de traslaciones en las tres direcciones del espacio, de tal forma que el cristal puede considerarse como un apilamiento, en tres dimensiones, de bloques idénticos llamados *Celdas Unitarias*. Su tamaño viene determinado por la longitud de sus tres aristas (a , b , c), y la forma por el valor de los ángulos entre dichas aristas (alfa, beta, gama: α , β , γ)(Fig. 4-4).

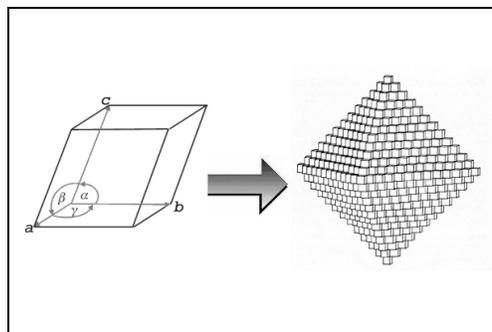


Figura 4-4. Celda Unitaria que por simple traslación reproduce el cristal. Su tamaño está determinado por las aristas a , b , c y su forma por el valor de los ángulos α , β , γ .

Sin embargo, las proteínas no tienen formas que se puedan definir con estos valores, entonces ¿cómo es que las proteínas forman cristales? Dentro de la celda unitaria existen elementos o *motivos* a lo que se les puede aplicar operadores de simetría para reconstruir la celda, estos motivos se conocen como *Unidad Asimétrica* y en cristalografía de proteínas esta unidad asimétrica está compuesta por la proteína, ya sea en estado monomérico o multimérico (Fig. 4-5).

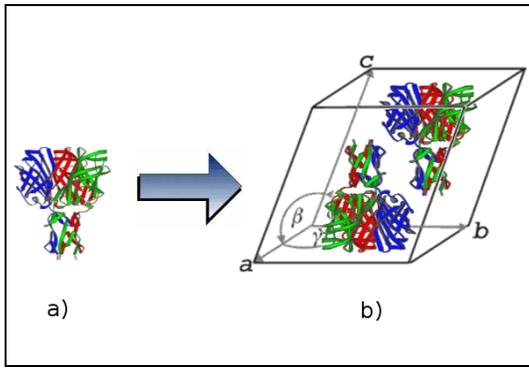


Figura 4-5. Unidad asimétrica **a)** formada por un monómero de proteína, a la que aplicándole los operadores simétricos del grupo espacial reproduce el contenido de la celda unitaria **b)**.

El conjunto de operadores simétricos que son aplicados a la unidad asimétrica para reconstruir la celda se denomina *Grupo Espacial*. En cristalografía de proteínas existen 230 grupos espaciales y su determinación correcta es crucial para la resolución de la estructura.

Difracción de cristales de proteína.

Cuando los rayos X alcanzan un átomo interaccionan con sus electrones exteriores. Éstos reemiten la radiación electromagnética incidente en diferentes direcciones. Los rayos X reemitidos desde átomos correspondientes dentro del cristal interfieren entre sí destructiva o constructivamente y esto genera los puntos del patrón de difracción. Solo cuando los rayos difractados interactúan constructivamente se pueden ver los puntos de difracción y esto sucede cuando se cumple la *Ley de Bragg* $n\lambda = 2d \text{sen}\theta$ (Fig. 4-6).

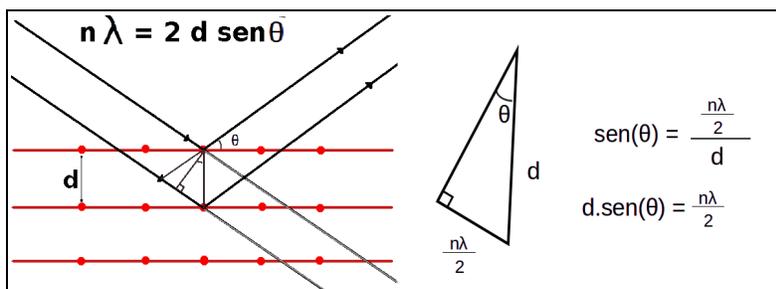


Figura 4-6. Representación geométrica de la ley de Bragg.

En un experimento de difracción conocemos el valor de la longitud de onda y el ángulo del rayo difractado, pero desconocemos el valor de “d” que depende de la disposición de los planos dentro del cristal que a su vez dependen de la simetría que relaciona a la Unidad Asimétrica: la proteína. Para

calcular el mapa de densidad electrónica a partir del patrón de difracción se utiliza una *transformada inversa de Fourier*, la descomposición de una señal en componentes de frecuencias diferentes, que necesita de dos datos: La intensidad de la difracción y la fase de los rayos difractados. La primera se mide experimentalmente con un detector de rayos X, sin embargo la segunda no se puede determinar tan fácilmente, ello representa un problema: *el problema de fases*.

Colecta de datos.

La difracción de cristales con rayos X se puede realizar en dos fuentes de radiación conocidas como ánodo rotatorio y sincrotrón. La primera, también llamada “*de casa*”, es menos intensa y tiene una longitud de onda fija, en cambio la fuente sincrotrón es mucho más intensa y además se puede modular la longitud de onda de los rayos X.

Durante la medición de difracción de rayos X, el cristal es montado en un goniómetro (instrumento de medición con base circular que permite colocar a un cristal en un ángulo definido) y se rota gradualmente mientras es bombardeado con rayos X, produciendo un patrón regularmente espaciado de puntos llamado *patrón de difracción*. Un patrón de difracción completo consta de una serie de *marcos* (frames) o imágenes en dos dimensiones tomados a diferentes valores de rotación del cristal utilizando un detector sensible a los rayos X. La cantidad de estos marcos necesarios para determinar la estructura de una proteína depende de la simetría del cristal, es decir, del grupo espacial. Una vez colectados todos los marcos se integran los valores de la intensidad (I) y de la desviación estándar de la intensidad (I/sigma(I)) de cada uno de los puntos de difracción en un mismo archivo.

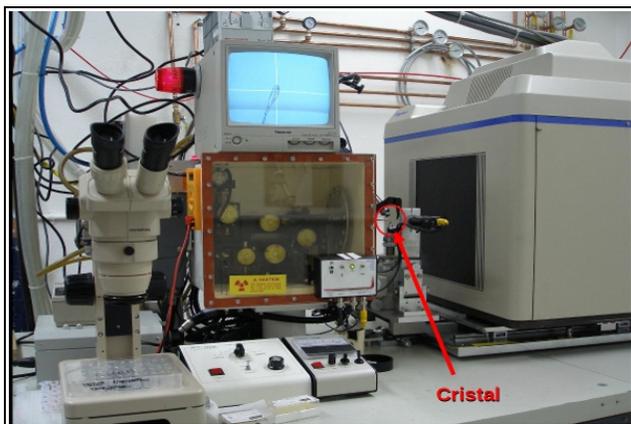


Figura 4-7. Fotografía del equipo de difracción de rayos X instalado en el Instituto de Biotecnología de la UNAM.

Métodos para resolver el problema de fase.

La técnica más utilizada para resolver el problema de fase se denomina Reemplazo Molecular (RM). En este método es necesaria la existencia de una estructura homóloga a la que se quiere determinar y se utiliza como modelo de búsqueda para calcular las fases. El modelo es rotado y trasladado dentro de una celda que tiene los parámetros calculados experimentalmente a partir de la difracción del cristal. Para cada valor de rotación y traslación se calcula un patrón de difracción teórico y se compara con el experimental, hasta que coincidan. Los valores de fases calculadas de esta manera son aproximadas, pero permiten generar un mapa de densidad electrónica inicial y conforme se construye el modelo de la proteína las fases se recalculan y mejora el mapa, permitiendo construir un modelo más detallado que a su vez aporta mejores valores de fase.

También existe la técnica llamada Dispersión Anómala a Múltiple longitud de onda (MAD) y una variante que utiliza una sola longitud de onda (SAD). Para resolver el problema por este método, se usa la señal que emiten ciertos átomos pesados a determinadas longitudes de onda de rayos X, esta señal se denomina “señal anómala”. Una ventaja que presenta este método es que no es necesario usar un modelo de búsqueda, es decir: las fases se calculan a partir de las posiciones de los átomos pesados y la proteína se construye *de novo*. En una fuente sincrotrón se puede cambiar la longitud de onda, y esto permite hacer experimentos de MAD, pues es necesario coleccionar, al menos, tres conjuntos de datos a diferentes longitudes de onda de los rayos X; una donde el átomo pesado absorbe los rayos X, una menor y otra mayor a este valor. Si la señal del átomo es buena, en el punto donde absorbe los rayos X, y se pueden determinar su posición dentro del cristal, es suficiente para determinar las fases de la proteína y poder trazar sus átomos. Mientras que para MAD se necesitan coleccionar tres conjuntos de datos, en SAD, sólo es necesario uno, en donde el átomo pesado absorbe los rayos X. Sin embargo, para poder extraer la señal anómala de algún átomo, se necesitan datos de mejor calidad que para el caso de RM, principalmente: multiplicidad de datos mayor de 4, el doble de grados coleccionados y alta resolución para usarla en los últimos pasos de afinamiento. El átomo más usado para MAD/SAD es el selenio; que se sustituye por el azufre de las metioninas. Para ello es necesario expresar la proteína en un organismo que no pueda sintetizar la metionina o inhibir su vía de síntesis y usar un medio que contenga seleniometionina; de esta manera, la seleniometionina es incorporada a las proteínas en vez de la metionina con azufre. Las proteínas con seleniometionina son más sensibles a la oxidación y usualmente son más hidrofóbicas y por lo tanto menos solubles (Doublié, 1997).

Otras técnicas para resolver el problema de fase son las de Reemplazo Isomorfo Múltiple (MIR) y Reemplazo Isomorfo Simple (SIR) que se basan en el siguiente principio: si se tienen dos cristales, uno de la proteína sola (cristal nativo) y otro de la proteína con átomos pesados unidos o incorporados, y se colectan datos de difracción de ambos, la diferencia en las intensidades de las difracciones será reflejo de la contribución de los átomos pesados a dicha intensidad. Esto porque los diferentes átomos contribuyen a la intensidad en proporción al cuadrado del número de electrones que tiene. Esta diferencia se puede usar para identificar a los átomos pesados, a partir de un *mapa de Patterson*, y puesto que estos siempre serán mucho menos que los de la proteína, es más fácil tratar de encontrar sus fases y con ello sus posiciones. Una vez que se conocen las fases de los átomos pesados, estas se aplican a la difracción del cristal nativo para obtener las fases de los átomos de la proteína. Usualmente es necesario tener más de una derivada metálica (cristal con átomos pesados unidos) para poder resolver las fases por esta técnica, sin embargo, si una derivada es lo suficientemente buena, es decir, que por sí misma aporte información suficiente para conocer las fases de la proteína, entonces no se necesitan más.

Afinamiento de la estructura.

Una vez que se ha resuelto el problema de fase se procede al Afinamiento de la estructura. En este paso se calculan mapas de densidad electrónica en donde se ajusta el modelo. También se añaden las moléculas de agua y otros ligantes que se puedan ver en la densidad electrónica que no pertenezca a la proteína. Para evaluar que el modelo está correctamente ajustado a la densidad electrónica se monitorean los valores de R y R_{free} que son indicativos del margen de error y se permiten valores alrededor del 20%.

5. HIPÓTESIS Y OBJETIVOS

La **hipótesis** que se plantea en este trabajo es que el grupo de proteínas al que pertenece la ThiDN_Tm de *Thermotoga maritima*, ThiDN_Pf de *Pyrococcus furiosus* y la ThiN_SS de *Sulfolobus solfataricus* tiene un plegamiento diferente al de la TPS de *Bacillus subtilis*.

El **objetivo general** es obtener cristales de buena calidad de las proteínas ThiDN de *Thermotoga maritima*, ThiDN de *Pyrococcus furiosus* y ThiN de *Sulfolobus solfataricus* y hacer estudios de difracción de rayos X para, posteriormente, determinar su estructura.

Los **objetivos específicos** son:

–Obtener cristales de las proteínas ThiDN de *Thermotoga maritima*, ThiDN de *Pyrococcus furiosus* y ThiN de *Sulfolobus solfataricus* para hacer estudios de difracción de rayos X.

–Obtener patrones de difracción de los cristales de las proteínas ThiDN's y ThiN.

–Procesar los datos obtenidos de los patrones de difracción y evaluar la viabilidad para determinar la estructura cristalográfica de las proteínas.

6. METODOLOGÍA

Cronológicamente, se comenzó a trabajar con la ThiDN de *T. maritima* (ThiDN_Tm), que era la proteína para la cual ya se tenían disponibles los genes clonados, los ensayos de complementación de función y las medidas de actividad de TPS hechas. Sin embargo, debido a las complicaciones metodológicas de tipo cristalográfico que se fueron dando con esta proteína, se decidió trabajar con la ThiDN de *P. furiosus* (ThiDN_Pf) y posteriormente con la ThiN de *S. solfataricus* (ThiN_Ss).

Purificación de las proteínas.

El protocolo de purificación fue esencialmente el mismo para las tres proteínas. Cada una de ellas se sobre expresó en la cepa de *Escherichia coli* BL21, usando como vector al plásmido pET-28a, que añade una secuencia con 6 histidinas en el amino terminal de la proteína. Una colonia transformada con el vector fue preinoculada en matraces de 50 ml de medio LB y el cultivo fue incubado a 37°C durante una noche. Con este cultivo se inoculó un litro de medio mínimo M9 iniciando con una $DO_{600} = 0.1$ que se dejó crecer a 30 °C hasta una $DO_{600} = 0.45$ para inducir la producción de la enzima con 0.5 mM de IPTG durante 12 hrs. El cultivo fue concentrado, resuspendido en 25 ml de amortiguador A (300 mM NaCl, 50 mM Fosfatos pH 7.4) con inhibidor de proteasas y DNAsa, y fue lisado por sonicación. El lisado fue centrifugado, el sobrenadante filtrado por membranas de 0.22 μm y cargado en una columna de afinidad de Níquel; la elución se llevó a cabo con imidazol usando un gradiente de amortiguador B (500 mM de Imidazol, 300 mM de NaCl, 50 mM de Fosfatos pH 7.4) de 0 a 20 %, con una meseta y después de 20 a 100%. Posteriormente, las fracciones que contenían a la proteína fueron concentradas usando tubos Amicon y cargadas en una columna de cromatografía de filtración en gel Superdex200 de 320 ml y usando como amortiguador 150 mM de NaCl y 50 mM de amortiguador de fosfatos a pH 7.8. Posteriormente, el valor de pH fue cambiado al de las condiciones de cristalización para las enzimas ThiDN_TM (pH 6.5) y ThiDN_Pf (pH 8.5).

Adicional a este protocolo, se intentó purificar a las proteínas aprovechando sus propiedades termoestables. El producto de la sonicación fue centrifugado y el sobrenadante incubado durante 40 minutos a 80°C. Este material se dejó atemperar antes de centrifugarlo; el sobrenadante obtenido fue concentrado, filtrado y aplicado a la columna de filtración en gel. Cuando se purificó las proteínas de esta manera, el perfil de elución presentó varios picos de formas multiméricas con tamaños mayores a 500 kDa, dando un rendimiento mucho menor de proteína útil para pruebas de cristalización.

Expresión de la proteína con seleniometoninas.

El procedimiento para obtener proteína con seleniometoninas es esencialmente el mismo descrito anteriormente. La diferencia radica en inhibir la vía de síntesis de metionina en *E. coli* y añadir seleniometonina al medio de cultivo mientras se lleva a cabo la sobre expresión de la proteína de interés. Una vez que el medio de cultivo ha llegado a una $DO_{600} = 0.40$ se añaden los siguientes aminoácidos: valina (50mg/l), isoleucina (50mg/l), leucina (50mg/l), lisina (100mg/l), fenilalanina (100mg/l) y treonina (100mg/l), después de 15 minutos se agrega la seleniometonina (60mg/l). Cuando el cultivo alcanza una $DO_{600} = 0.45$ se induce la expresión de la proteína y se continúa con el procedimiento de purificación descrito (Doublíé, 1997).

Digestión con trombina para remover la marca de histidina.

El plásmido pET-28a, además de la secuencia para añadir un marca de 6 histidinas en la proteína, añade una secuencia de corte con trombina que permite retirar la marca de histidinas. Para realizar la digestión se usa trombina unida a biotina. Primero se lleva a cabo una prueba de diluciones de trombina para determinar cuál es la mejor concentración de la enzima que se usará en la digestión final. Para estas pruebas se usaron las diluciones 1:25, 1:50, 1:100 y 1:200 con el buffer de almacenamiento de la enzima y se configuraron las siguientes reacciones de prueba:

- 5 μ l 10x *Buffer* de reacción.
- 10 μ l de la Enzima a la concentración deseada. (1mg/ml)
- 1 μ l de la Trombina diluida (1:25, 1:50, 1:100, 1:200).
- 34 μ l Agua Tetradestilada.

Las reacciones se incubaron a 20°C y se tomaron muestras a las 2, 4, 8 y 16 horas para evaluar el grado de digestión por medio de SDS-PAGE.

Para realizar el escalamiento de la reacción se multiplican por 100 los valores de aquella configuración que haya rendido una mayor digestión de la proteína. Para detener la reacción se añaden 44.8 μ l de agarosa-streptoavidina y se incuba durante 30 minutos. Posteriormente se centrifuga a 500 g por 5 minutos en los filtros que retienen la agarosa-streptoavidina, esto para retirar la trombina (Novagen, thrombin kits).

Pruebas de cristalización.

Las pruebas de cristalización se hicieron principalmente por el método de difusión de vapor en gota colgante aunque también se intentaron en determinado momento el método de gota apoyada, de diálisis y *microbatch*. Las principales variaciones que se hicieron en los métodos de cristalización fueron en el tamaño de la gota de proteína-precipitante y en la proporción de estos dos componentes.

El procedimiento para obtener cristales de proteína tiene varias etapas, primero se hace un escaneo de condiciones de cristalización usando matrices incompletas disponibles comercialmente. En este trabajo se usaron el *kit Crystal Screen I formulation*, *Crystal Screen II* y el *Crystal Screen Cryo*, de la compañía Hampton Research. Estas matrices exploran estadísticamente un amplio rango del tipo y concentración de los agentes precipitantes. Las condiciones de cristalización son monitoreadas constantemente y se anotan aquellas en donde se obtiene precipitado, crecimiento cristalino, microcristales, cristales o si permanece translúcido (proteína en estado soluble). De aquí se busca la recurrencia de pH, fuerza iónica y tipo de precipitante, para diseñar nuevas matrices que exploren alrededor de estas condiciones de cristalización y de esta manera hacer una optimización gradual hasta obtener los cristales deseados. Las condiciones de cristalización se colocaron a 4°C, 18°C y/o 37°C, resultando más exitosas las que estaban a 18°C.

Colecta y procesamiento de datos de difracción.

Dado que los costos de difracción de rayos X son elevados, cada vez que se realizan se evalúa la posibilidad de analizar los cristales que existan, con características deseables para estudios de difracción, pertenecientes a otros proyectos, ya sea para coleccionar datos completos o simplemente para hacer pruebas de difracción de material cristalino. Debido a los contactos y proyectos de colaboración con otros laboratorios de estructura de proteínas, el laboratorio cuenta con las siguientes facilidades para hacer estudios de difracción de cristales: El propio equipo de difracción de rayos X de ánodo rotatorio, instalado en el Instituto de Biotecnología en Cuernavaca, Morelos; el Laboratorio Universitario de Estructura de Proteínas (LUEP) en la Ciudad de México, con una fuente de rayos X de ánodo rotatorio; la línea X6A con radiación Sincrotrón, de National Sincrotrón Light Source en Upton, New York; el Laboratorio de Biofísica Molecular de la Universidad de Oxford, UK, con fuente de rayos X de ánodo rotatorio; el European Synchrotron Radiation Facility en Grenoble, Francia; el Instituto de Física de São Carlos en Brazil, con fuente de rayos X de ánodo rotatorio; el Advanced Photon Source en Chicago, USA. Por lo que los cristales de este proyecto fueron difractados en una o más de estas fuentes de radiación.

El procedimiento de difracción de cristales de proteína comienza con la selección de candidatos. Después se lleva a cabo un cuidadoso empaque para su transporte. En el laboratorio se usan principalmente dos tipos de empaque, uno de ellos es colocar el porta objetos donde creció el cristal en una “cámara” cerrada en donde se deposita el líquido precipitante formando un sistema parecido al de difusión de vapor en gota apoyada y se coloca en un recipiente térmico forrado de fibra de asbesto para reducir las vibraciones externas que pudieran dañar los cristales; este método se usa para viajes cortos y/o para cristales resistentes. La otra forma de transportar cristales es montarlos en asas con base para cabeza goniométrica y congelarlos en Nitrógeno líquido, este método es para viajes largos y/o cristales sensibles. Después se traslada a la fuente de rayos X, ya sea de ánodo rotatorio o Sincrotrón.

Para coleccionar los datos de difracción el cristal se coloca sobre una cabeza goniométrica y se centra en el colimador del haz de rayos X. Se colecta un marco de lectura a 0° y otro a 90° que sirven de prueba para determinar la simetría del cristal y sobre eso definir el rango que se coleccionará, el grado de oscilación y el tiempo de exposición del cristal a los rayos X. Una vez determinados estos parámetros se configura la colecta completa de datos.

Posteriormente se realiza el procesamiento de datos que consiste en la integración de la posición e intensidad de todos los puntos de difracción en un sólo archivo y un escalamiento de las intensidades que permite calcular parámetros estadísticos que el *software* necesita durante el procesamiento de datos. Para ello se usan los siguientes programas: Mosflm (Leslie, 2006), XDS (Kabsh, 1993), CCP4 (Collaborative Computational Project, 1994). En los intentos de Reemplazo Molecular se usaron Crystallography & NMR System, más conocido como CNS (Brunger, *et al*, 1998), AMoRe (Navaza, 1994), EPMR (Kissinger, 1999) y MolRep (Vagin, 1997).

7. RESULTADOS Y DISCUSIÓN

Purificación de las proteínas.

La proteína ThiDN_Tm, en la columna de cromatografía de filtración en gel, eluye con un pico mayoritario en la fracción de 165 ml correspondiente a ~ 3.6 veces el peso del monómero, que es 47.4 kDa. Esto sugiere que la partícula en solución es un tetrámero. En cambio, la proteína ThiDN_Pf presentó un perfil de elución con varios picos; el mayoritario fue en 162 ml, correspondiente a ~ 5 veces el peso del monómero, que es 51.3 kDa. Este dato contradice, en cierto modo, a lo esperado, pues el mayor porcentaje de identidad entre los dominios ThiD de estas dos proteínas, con respecto a la ThiD de estructura conocida, se encuentra en los primeros 60 aminoácidos, que constituyen alrededor del 70 % de la zona de dimerización, por esta razón se espera que las proteínas ThiDN_Tm y ThiDN_Pf puedan formar complejos múltiples de 2. La respuesta a esta cuestión se resolvió con una prueba de Dispersión Dinámica de Luz (*Dynamic Light Scattering*) hecha a la proteína ThiDN_Pf a la concentración usada para las pruebas de cristalización, en el amortiguador de la cromatografía de filtración en gel. En esta prueba se determinó que la partícula en solución tiene un radio de 5.97 nm, compatible con la mejor solución parcial obtenida en el Reemplazo Molecular, que fue un tetrámero ligeramente alargado de longitud promedio de ~ 12.5 nm. La proteína ThiN_Ss eluye como un sólo pico en 222 ml, que es compatible con un monómero, cuyo peso molecular es 23.8 kDa. El rendimiento neto de proteína útil para ensayos de cristalización fue de alrededor de 7 mg de proteína por cada litro de cultivo, semejante para las tres proteínas.

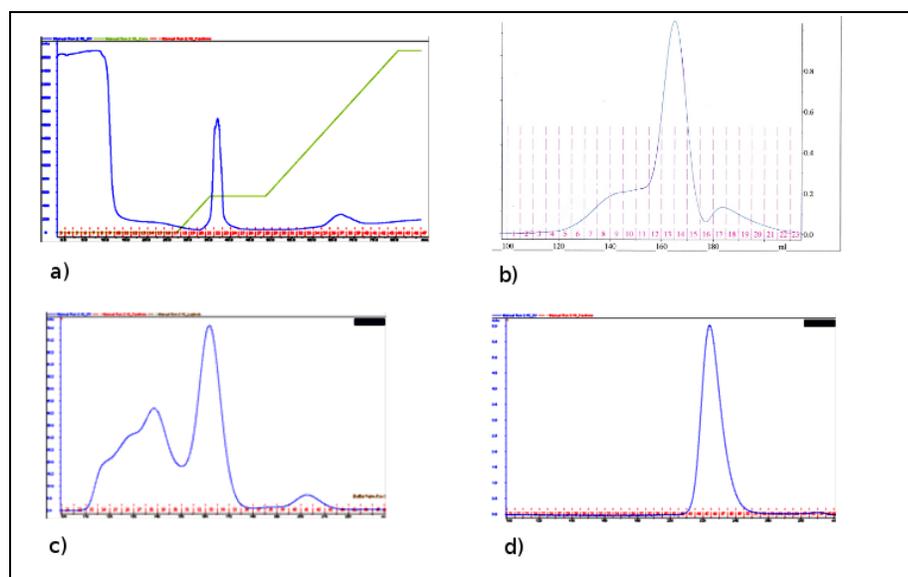


Figura 7-1. Perfil de elución de **a)** columna de afinidad de Niquel típica de las proteínas ThiDN_Tm, ThiDN_Pf y ThiN_Ss; **b)** columna de filtración en gel de la proteína ThiDN_Tm; **c)** de la proteína ThiDN_Pf; **d)** de la proteína ThiN_Ss.

Expresión de la proteína con seleniometoninas.

La expresión de las proteínas ThiDN_Tm y ThiDN_Pf con seleniometoninas produjo alrededor de 7 mg de proteína por litro de cultivo, semejante al rendimiento de la proteína con azufre, lo que permitió realizar pruebas de cristalización. En el caso de la ThiDN_Tm se obtuvieron cristales de tamaño adecuado en condiciones muy parecidas a las que dieron cristales de la proteína nativa, adecuados para hacer estudios de difracción. Dichos cristales tenían un 89 % de integración de Seleniometonina, determinado por la técnica de MicroPIXE (Proton-Induced X-ray Emission), que es capaz de determinar con gran precisión el tipo y la cantidad de un determinado compuesto presente en una muestra (Garman, 1999). A pesar de que se exploraron diversas condiciones de cristalización, los cristales obtenidos de la proteína ThiDN_Pf con seleniometoninas siempre fueron muy pequeños, no útiles para hacer estudios de difracción de rayos X, por lo que no fue posible coleccionar un patrón de difracción para esta proteína. Por otra parte, la proteína ThiN_Ss con seleniometoninas no se expresó.

Digestión con trombina para remover la marca de histidina.

En las preparaciones donde estuvo involucrada la digestión con trombina, el rendimiento de proteína útil para ensayos de cristalización fue muy bajo porque la proteína se precipitaba conforme se digería. Se exploraron algunos amortiguadores para tratar de obtener más solubilidad en la proteína; estos incluían alta fuerza iónica, glicerol o ambos, diseñados en base a la solubilidad de algunas pruebas de cristalización. El amortiguador en donde la proteína resultó ser más soluble fue el que contenía 30 % de glicerol, 150 mM de NaCl, 20 mM de TRIS y 25 mM de CaCl₂. En las digestiones hechas con este amortiguador también se observó proteína precipitada, aunque en menor grado, y se pusieron algunas pruebas de cristalización.

Pruebas de cristalización, colecta y procesamiento de datos.

ThiDN_Tm

Para la proteína ThiDN_Tm se usó, como soluciones precipitantes, el kit *Crystal Screen II formulation* de Hampton Research (Ref. 2) para tener un primer acercamiento a las condiciones en las que pueda cristalizar la proteína. Este kit explora 48 condiciones de cristalización, variando el pH, el tipo y la concentración de precipitante y de sal. Se colocaron pruebas de cristalización, por el método de gota colgante, con cada una de las condiciones del *kit* a 3 temperaturas: 4°C, 18°C y 37°C. Estas pruebas se monitorearon constantemente y se exploraron variantes de aquellas en donde había crecimiento microcristalino, evaluando condiciones recurrentes de pH, fuerza iónica y tipo de

precipitante.

Después de un proceso de optimización de condiciones, se obtuvieron cristales de proteína a los que se les midió su difracción. El cristal que mejor difractó, y del cual se logró coleccionar un conjunto de datos, creció a 18°C en 300 mM Acetato de Sodio, 9% PEG MMe 5000 y 125 mM Fosfato de Sodio a pH de 6.5, con una concentración de proteína de 20 mg/ml. Pudo ser crioprotegido en 35 % de glicerol y transportado en nitrógeno líquido hasta la fuente de rayos X de ánodo rotatorio, en el LUEP en la Ciudad de México. Este cristal difractó a 3.1 Å de resolución, pertenece al grupo espacial P1 y tiene parámetros de celda $a = 60.4 \text{ \AA}$, $b = 84.39 \text{ \AA}$, $c = 110.92 \text{ \AA}$, $\alpha = 107.47^\circ$, $\beta = 96.02^\circ$, $\gamma = 93.82^\circ$, y con un tetrámero en la unidad asimétrica. Se logró coleccionar 159° de rotación antes de que el cristal decayera, esto representa el 88 % de los datos necesarios para este grupo espacial (Fig. 7-2). En la tabla 7-1 se muestran los parámetros estadísticos de la integración del mejor cristal coleccionado de ThiDN_Tm, ThiDN_Pf y ThiN_Ss en el LUEP.

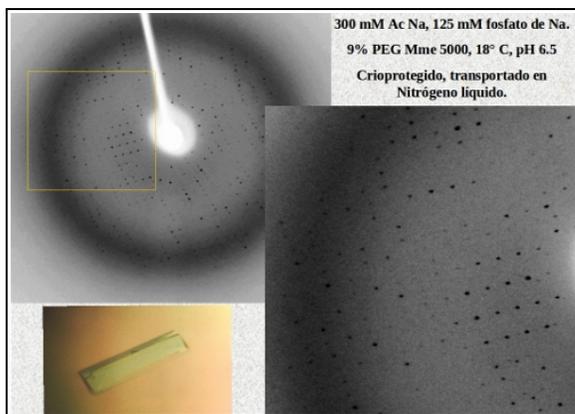


Figura 7-2. Cristal y patrón de difracción de la proteína ThiDN de *Thermotoga maritima*. El cristal creció en 300 mM de Acetato de Sodio, 125 mM de Fosfato de Sodio a pH 6.5, 9% de PEG Monometil-eter, a 18°C. Este cristal fue difractado en una fuente de rayos X de ánodo rotatorio.

Dado que se conoce un homólogo con estructura conocida para uno de los dominios, se intentó hacer Reemplazo Molecular con el programa CNS, usando un modelo de búsqueda basado en la estructura de la ThiD de *Salmonella typhimurium*, ThiD_St (Cheng, *et al*, 2002). El sistema cristalográfico P1 no tiene definido un origen, y por lo tanto, tampoco una función de traslación. Sin embargo, con un tetrámero en la unidad asimétrica, es necesario buscar la posición de cuatro partículas en el cristal, es decir, se deben resolver las funciones de cuatro rotaciones y, además, cuatro traslaciones.

	ThiDN_Tm	ThiDN_Pf	ThiN_Ss
Grupo espacial	P1	P6,22	P4,2,2
Celda: a, b, c (Å)	60.40 84.39 110.92	174.49 174.49 80.11	115.390 115.390 64.470
Alfa, Beta, Gama (°)	107.41 96.02 93.81	90.00 90.00 120.00	90.00 90.00 90.00
Rango de Resolución	40.57 – 3.0 Å	50 – 3.5 Å	81 – 2.01 Å
Grados de colecta	287 °	59 °	46 °
Rango de Oscilación	1 °	1 °	1 °
Integridad de datos	97.60%	99.00%	98.00%
Reflexiones únicas	42349	32864	29937
Redundancia	2.8 (2.7)	6.9 (6.8)	2.9 (2.5)
Mosaicismo	1 °	0.7 °	0.4 °
R _{merge}	0.10 (0.39)	0.2 (0.5)	0.16 (0.35)
I/sigma	4.7 (2.1)	3.4 (1.5)	4.9 (2.3)
Unidad asimétrica	Tetrámero	Monómero	Dímero

Tabla 7-1. Parámetros estadísticos de la integración de los datos colectados de las proteínas ThiDN_Tm, ThiDN_Pf y ThiN_Ss. Estos datos fueron colectados en una fuente de rayos X de ánodo rotatorio en el Laboratorio Universitario de Estructura de Proteínas, en la Ciudad de México. En paréntesis se muestran los datos de la última capa de resolución.

El modelo de búsqueda basado en la ThiD_St representa alrededor del 50 % de la proteína, pero en términos cristalográficos, abarca sólo el 12% de la unidad asimétrica, que es un tetrámero. Además, el modelo no está homogéneamente distribuido, sino que corresponde solamente con un dominio, de tal manera que la información de fases proviene sólo de una región y no de toda la unidad asimétrica. Sumado a la baja resolución e integridad de datos convierte al Reemplazo Molecular con estos datos en algo complejo. Las soluciones de Reemplazo Molecular sólo permitían calcular mapas de densidad electrónica discontinuos, difíciles de interpretar.

Debido a la baja resolución a la que difractó este cristal en un ánodo rotatorio, se decidió coleccionar datos en una fuente de radiación de sincrotrón. El usar esta fuente permitía, además, intentar la estrategia para resolver el problema de fase por el método de Dispersión Anómala con Múltiple longitud de onda (MAD) utilizando selenio en forma de seleniometionina como dispersor anómalo. La ThiDN_Tm tiene 13 metioninas por monómero, esto significa que se debe determinar la posición de 52 metioninas en la unidad asimétrica.

Después de optimizar las condiciones de crecimiento de los cristales de proteína con seleniometioninas y haber hecho pruebas de difracción de varios de ellos, se logró coleccionar un conjunto de datos a 2.95 Å de resolución. A pesar de que se coleccionó un rango más que suficiente de grados, el análisis de los datos mostró que el cristal es demasiado sensible a la radiación y que sólo son útiles los primeros 287° que representan un 79 % de integridad, ya que para poder extraer la señal

anómala en un grupo espacial P1 se necesitan 360°, de tal forma que nuevamente nos enfrentamos a un conjunto de datos con baja resolución e integridad, que no permitieron determinar la posición de los átomos de selenio con la certeza suficiente para usar sus fases y con ello determinar la posición de los demás átomos de la proteína.

Sin embargo, los datos podían ser usados para intentar nuevamente un Reemplazo Molecular ya que, al procesarlos de esta manera, la calidad y la integridad era mucho mejor que los colectados en el ánodo rotatorio y eso podría marcar una diferencia significativa. Para enfrentar el problema de que el sistema cristalino P1 no tiene definido un origen, la estrategia a seguir fue buscar, dentro de los picos de rotación más altos, cuatro que fueran compatibles con los *ejes de simetría no cristalográficos* (simetrías locales que relacionan objetos dentro de la celda y que no forman parte de los ejes cristalográficos), y después buscar “manualmente” las funciones de traslación variando paulatinamente la distancia entre ellos respecto a un origen arbitrario, que en este caso fue el centro de masa aproximado de lo que sería el tetrámero. De esta manera, se encontraron tres rotaciones diferentes entre sí, que se ajustan a las simetrías no cristalográficas, cada una por sí sola, a pesar de que representa un octavo de la unidad asimétrica, describe de manera razonable lo que existe en el cristal. La solución de rotación para un cuarto monómero no fue tan clara. Se hicieron diversos intentos por encontrar las traslaciones de los cuatro monómeros dentro de la celda, pero no se encontró una solución que mejorara de manera significativa, de tal modo que la máxima información que se puede extraer de estos datos es la posición relativa de los monómeros dentro de la celda unitaria, pero no la ruta completa de los carbonos alfa ni de las cadenas laterales de aminoácidos. (Fig. 7-3)

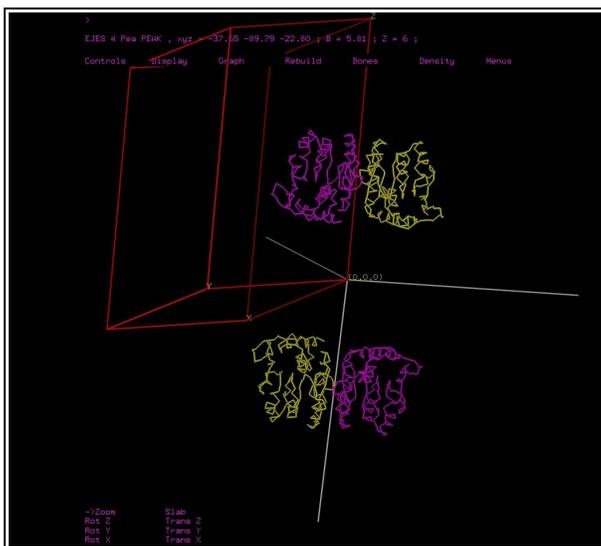


Figura 7-3. Solución parcial del Reemplazo Molecular del cristal de ThiDN_Tm. En amarillo y magenta se muestra la posición relativa de los dominios ThiD dentro del cristal, en rojo se muestra la celda unitaria y en blanco los ejes de simetría no cristalográfica.

Por estos motivos se decidió cambiar de modelo de estudio a la proteína ThiDN de *P. furiosus* (ThiDN_Pf).

ThiDN_Pf.

Las condiciones iniciales de cristalización de la proteína ThiDN_Pf se hicieron usando los *kits Crystal Screen I, II* y *kit Crystal Screen Cryo*. Y algunas pruebas de cristalización se hicieron con proteína que fue sometida a digestión con trombina. Estas condiciones fueron monitoreadas constantemente y se anotaron aquellas en donde se veía crecimiento cristalino para hacer variantes y combinaciones e ir optimizando la obtención de cristales.

De las pruebas de cristalización con proteína digerida se obtuvieron algunos cristales pequeños y uno de ellos se pudo difractar en una fuente de ánodo rotatorio. El cristal difractó a 8 Å de resolución y se logró coleccionar 60° con una oscilación de 2° por marco. A esta resolución no es posible estar completamente seguros de cuál es el grupo espacial, sin embargo, sabemos por un análisis de auto-rotación, que tiene un eje de orden 6 y dos de orden 2, es decir, es un cristal con grupo espacial P622 o alguno de sus derivados y una celda aproximada de $a=b=175$ Å, $c=80$ Å, $\alpha=\beta=90^\circ$, $\gamma=120^\circ$. Obtener cristales con características deseables para estudios de difracción con la proteína sometida a digestión con trombina no fue posible ya que el rendimiento de proteína útil para cristalizar era muy bajo debido a que se perdía por precipitación durante la digestión y al momento de concentrarla para hacer las pruebas de cristalización.

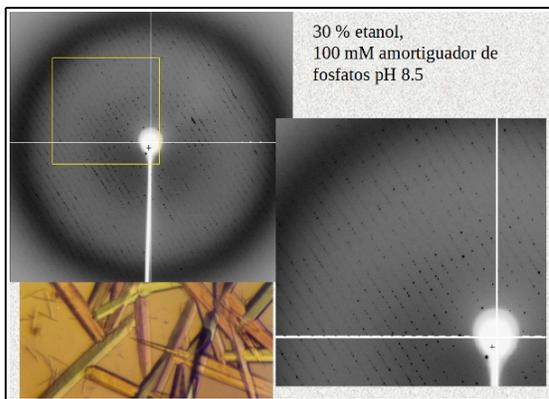


Figura 7-4. Cristales y patrón de difracción de la proteína ThiDN de *Pyrococcus furiosus*. Estos cristales crecieron en 30 % de Etanol, 100 mM de amortiguados de fosfatos a pH 8.5. Los cristales fueron difractados en una fuente de rayos X de Soncrotrón.

En las condiciones de cristalización con la proteína que no fue sometida a digestión con trombina se obtuvieron cristales en 30% de etanol con 100 mM de amortiguador de fosfatos a pH 8.5 como únicos componentes (Fig. 7-4). De las pruebas de difracción de estos cristales, se logró coleccionar

dos conjuntos de datos en una fuente sincrotrón, uno de ellos fue con un cristal de la proteína en estado nativo y el otro de uno remojado en 100 μM de Tiamina monofosfato durante una hora. Estos cristales difractaron alrededor de 3 Å de resolución, con un grupo espacial probable $P6_422$ o $P6_222$ y una celda de $a=b=174.49$, $c=80.11$, $\alpha=\beta=90$, $\gamma=120$, con un monómero en la unidad asimétrica (Fig. 7-5).

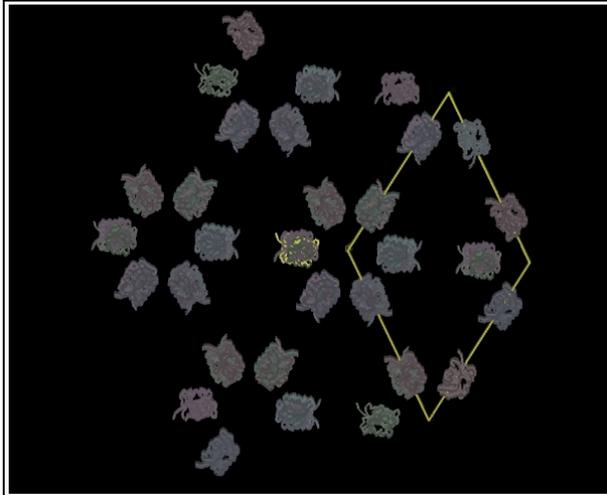


Figura 7-5. Solución parcial del Reemplazo Molecular del cristal de ThiDN_Pf. En gris se muestran las posiciones relativas del dominio correspondiente a ThiD. En amarillo se muestra la celada unitaria.

Con estos datos también se intentó hacer un Reemplazo Molecular usando un modelo de búsqueda basado en la estructura de ThiD_St. En esta ocasión, el modelo representa el 50 % de la unidad asimétrica. La mejor solución encontrada fue de un monómero en la unidad asimétrica, que por las simetrías del cristal puede generar un tetrámero, constituido por un dímero de dímeros. La mejor solución de Reemplazo Molecular de la ThiDN de *P. furiosa* tiene unos valores de R alrededor de 0.45, y unos valores de FOM de alrededor de 0.60. Sin embargo, aún con estos valores, no fue posible determinar la posición correspondiente al dominio ThiN a partir de los mapas de densidad electrónica. El principal problema sigue siendo la baja resolución.

Nuevamente nos encontramos ante la necesidad de tener alguna señal anómala para poder construir el dominio ThiN de la proteína, así que se intentó obtener cristales con seleniometionina, sin embargo, para esta proteína sólo se obtuvieron cristales muy pequeños no aptos para difracción de rayos X. Otra estrategia para obtener una señal anómala, es remojar cristales nativos en una solución que contenga las mismas sustancias que la gota donde creció el cristal más algún átomo pesado. En estos casos, se ha reportado al yodo como un átomo pesado muy útil como dispersor anómalo (Boggon, *et al*, 2000; Dauter, *et al*, 2000).

Para medir señal anómala es muy importante la calidad de los puntos de difracción. También es necesario tomar en cuenta el tiempo que los cristales permanecen remojados en la solución con el átomo pesado; para el yodo debe ser el menor posible, sólo de unos segundos. Otro aspecto importante es que el cristal no sufra decaimiento durante la colecta de datos, y para eso se debe minimizar, lo más posible, el tiempo de exposición con el compromiso de tener la mayor integridad de datos. Una estrategia útil en estos casos es hacer rotar la celda del cristal, durante la colecta, sobre su eje de rotación de máxima simetría, que en este caso es el de orden 6; de esta manera sólo se necesitan 15° teóricos para una buena integridad de datos de un cristal nativo, y 30° para un derivado anómalo. Una forma de hallar este eje es alinear un eje del cristal con el eje de rotación del mismo. Dado que una celda es la mínima expresión del cristal que por simple translación lo reconstruye, algunas veces, las caras del cristal están orientadas de acuerdo a las caras de la celda, en otras ocasiones sucede que algunas caras tienen un crecimiento preferencial, y se refleja de manera macroscópica, de tal manera que la forma del cristal no tiene nada que ver con la forma de la celda.

Con esto en mente, se colectaron datos de difracción de dos cristales en una fuente de radiación de sincrotrón. Para el primer cristal se tuvo mucho cuidado de que el tiempo de remojado fuera mínimo, mientras que para el segundo cristal se tuvo más énfasis en que la orientación fuera con el eje del cristal paralelo al eje de rotación. El análisis de las auto-rotaciones de estos datos permitió confirmar que el grupo espacial es $P6_422$. Las auto-rotaciones son operaciones de rotación que se hacen a los datos de difracción (en donde no está involucrado ningún modelo), de tal manera que, por ejemplo, la aparición de un pico de rotación en 60 grados significa que una rotación de 60 grados en torno a uno de los ejes de la celda reconstruye el patrón de difracción, indicando que el cristal presenta una simetría de orden 6. El análisis sistemático de los picos de auto-rotación permite conocer el grupo espacial sin sesgo alguno.

Para medir con precisión la señal anómala es necesario obtener un valor de $I/\sigma(I)$ de por lo menos 4, es decir, una señal del cristal 4 veces más alta que la señal del ruido y una integridad de datos anómalos por lo menos de 90 %, en comparación con el $I/\sigma(I)$ para datos usados en reemplazo molecular, que puede ser de 1.5 y la integridad de datos de 75% en casos extremos. Para estos cristales, los parámetros de evaluación restringen el rango de datos útiles para ser procesados por SAD hasta una resolución de 3.7 Å, después de esto, ya muestran un decaimiento del cristal. El análisis de estos datos permitió saber que, en este caso, es más importante el tiempo de remojado que la orientación del cristal ya que la señal anómala del segundo cristal tiene más ruido. También es muy

importante el tamaño del cristal ya que entre más grande sea, más alta es la señal de los átomos de la proteína y del yodo con respecto al ruido de fondo y menor es el daño por radiación, que es imprescindible para resolver la estructura cristalográfica de la proteína.

Para tener un diagrama de fase de trabajo para la proteína ThiDN_Pf y obtener cristales más grandes, fue necesario incrementar su solubilidad para tener una concentración más alta. Para ello se recurrió nuevamente a las pruebas de cristalización montadas con anterioridad, de ellas se observó que en las condiciones con pH altos las gotas seguían translúcidas, indicando que la proteína aún se encontraba soluble. De esta manera se decidió trabajar con la proteína a pH 8.5 desde que es colectada de la columna de afinidad de Níquel hasta que se ponen las pruebas de cristalización. También se observó que el glicerol disminuía la nucleación, es decir, también le brindaba más solubilidad a la proteína. Con estas condiciones, manteniendo a la proteína en un amortiguador de fosfatos a pH 8.5, con 150 mM de NaCl y 5% de glicerol se logró una concentración de 60mg/ml y con la muestra se hicieron dos matrices, de 9 puntos cada una, para explorar una zona del diagrama de fase. Las matrices tienen tres concentraciones de proteína (10, 26, 60mg/ml) y tres concentraciones del precipitante (15, 20 y 25 % de etanol para una matriz e isopropanol para otra).

En estas dos matrices no se obtuvieron cristales, sin embargo, sirvieron de pauta para diseñar nuevas condiciones de cristalización. Además, el hecho de lograr un incremento en la solubilidad de la proteína representa una gran ventaja, ya que las condiciones de cristalización se pueden hacer con más proteína por experimento, esto favorece la aparición de cristales de mayor tamaño. Las condiciones de cristalización derivadas de estos resultados exploran un rango de etanol desde 30 hasta 35 % para una concentración de proteína de 10 mg/ml; otro rango de etanol desde 13 hasta 18 % para una concentración de proteína de 25 mg/ml. Cada una de estas con 0, 5 y 10 % de glicerol y a su vez con Tiamina-fosfato, Thiazol-fosfato, Thiazol-fosfato + pirofosfato + Magnesio, pirofosfato + Magnesio y sin ligante. Además, una serie idéntica, pero sustituyendo el etanol por isopropanol. (Fig. 7-6)

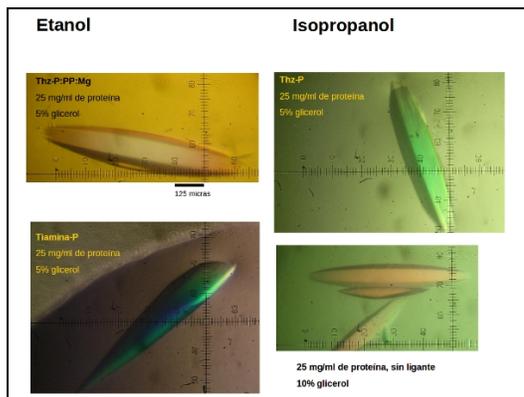


Figura 7-6. Optimización de cristales de la proteína ThiDN_PF. Estos cristales crecieron en condiciones con etanol e isopropanol a 18°C.

Como resultado de estos experimentos se obtuvieron cristales, cuyas dimensiones van desde 80 hasta 160 micras de largo, que son útiles para estudios de difracción de rayos X. Los cristales más grandes se obtuvieron con 25 mg/ml de proteína y los hay con los diferentes ligantes. Una vez que se tuvo instalado el equipo de rayos X en el laboratorio, se hicieron pruebas de difracción de los cristales obtenidos en estas condiciones. Dado que estos cristales no crecen en condiciones crioprotectoras, fue necesario hacer un rastreo de soluciones que cumplan esta función. Se eligieron cinco soluciones crioprotectoras compatibles con condiciones de cristalización con etanol, que tenían los mismos reactivos en donde creció el cristal más; a) 30% de glicerol, b) 30% de trealosa, c) 30% de MPD, d) 35% de PEG400, e) 30% de isopropanol+12%glicerol o bien 30% de etanol+12%glicerol, según fueran gotas que crecieron en isopropanol o en etanol.

A pesar de que los cristales eran más grandes y no fueron transportados, difractaron muy pobremente, alrededor de 6 Å de resolución. Los cristales que crecieron en isopropanol, difractaron mejor que los cristales en etanol, sin embargo, no se logró mejorar los resultados del cristal difractado anteriormente.

ThiN_Ss

Las pruebas de cristalización de la proteína ThiN de *Sulfolobus solfataricus* no habían rendido resultados alentadores. Varios meses después de las pruebas de cristalización iniciales se observaron cristales en la condición 32 del *Crystall Screen Cryo*, la cual contiene 1.5 M de sulfato de amonio y 25 % de glicerol. Muy probablemente estas condiciones se encontraban en equilibrio en la zona metaestable del diagrama de fases, por lo que no se vio favorecido la aparición de núcleos y por consiguiente de cristales, hasta que se produjo un cambio que favoreció la aparición de algunos

núcleos, que crecieron hasta formar los cristales. Uno de estos cristales se difractó en el LUEP en la Ciudad de México, llegó a 2.0 Å de resolución y se pudo colectar un conjunto de datos con una integridad del 90 %. El grupo espacial al que pertenece este cristal es $P4_22_12$ y tienen una celda de $a=b=115.390$ Å, $c=64.470$ Å, $\alpha=\beta=\gamma=90^\circ$. (Fig. 7-7)

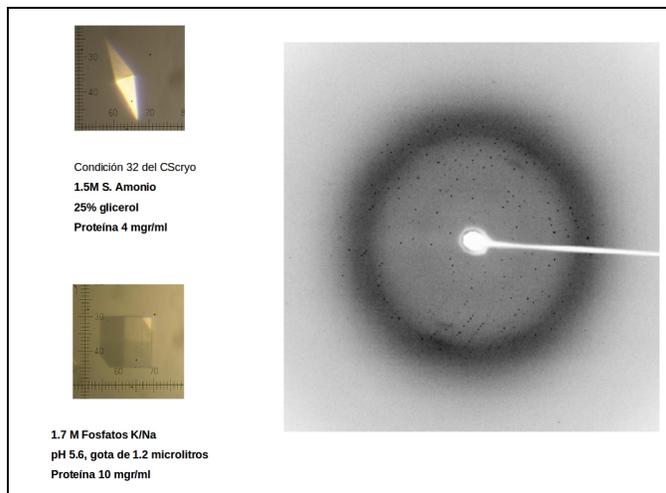


Figura 7-7. Cristales y patrón de difracción de la proteína ThiN de *Sulfolobus solfataricus*. Este patrón de difracción fue obtenido en la fuente de rayos X de ánodo rotatorio del LUEP.

A partir de las condiciones donde aparecieron estos cristales, se exploraron variantes de la concentración de sulfato de amonio desde 1 hasta 1.8 M a cuatro concentraciones de proteína: 3, 18, 20 y 25 mg/ml, manteniendo fijo el porcentaje de glicerol. El objetivo de estas variantes era desplazar el equilibrio de las condiciones iniciales hacia la zona de nucleación del diagrama de fases. En estas pruebas se obtuvo un ligero precipitado en algunas gotas. El sulfato de amonio es volátil y esto genera dificultades para controlar los valores reales de concentración dentro de la gota en equilibrio y por consiguiente el pH. Una estrategia útil en casos como este es sustituir el sulfato de amonio por un regulador de fosfatos que combinara el fosfato de sodio monobásico con el fosfato de potasio dibásico. Los fosfatos son más estables y la combinación de estos compuestos permite trabajar con altas concentraciones de amortiguador de fosfatos, además de poder regular el pH de manera más fácil y reproducible, únicamente modificando los porcentajes de cada uno.

Usando este criterio, se hicieron algunas pruebas de cristalización para ver el comportamiento del amortiguador de fosfatos comparado con el sulfato de amonio, manteniendo la concentración de proteína fija en 10 mg/ml. Estas pruebas exploraron 1.6, 1.8 y 2 M de sulfato de amonio a pH 6.5 y pH 5.5, con 0 y 25 % de glicerol, y una otra serie idéntica pero sustituyendo el sulfato de amonio por el amortiguador de fosfatos, y se colocaron a dos temperaturas: 18 y 30°C. En las condiciones que

contenían el amortiguador de fosfatos, se obtuvieron microcristales y precipitado cristalino, así que se continuó explorando rangos más finos de glicerol, pH y concentración de fosfatos, además de diferentes ligantes como Tiamina-fosfato, Tiazol-fosfato, Tiazol-fosfato + pirofosfato + magnesio y pirofosfato + magnesio. Los mejores cristales se obtuvieron a 18°C en 10% de fosfato de Potasio dibásico y 90 % de fosfato de Sodio monobásico a una concentración de 1.7 M, a 18 grados centígrados con 0 % de glicerol, y algunos de ellos en condiciones con Tiamina-fosfato. Estos cristales tienen entre 80 y 120 micras de largo, cabe notar que la forma que tienen estos cristales es diferente a los cristales que crecieron en sulfato de amonio (Fig. 7-7).

La proteína ThiN_Ss sólo tiene el dominio para el cual no existe ningún homólogo con estructura conocida, por lo que el problema de fase sólo se puede resolver con técnicas como SAD/MAD o SIR/MIR. Sin embargo, cuando se intentó purificar la proteína con seleniometioninas esta no se expresó, por lo que la única técnica para resolver el problema de fase que se podía usar era MIR o SIR.

En la técnica de MIR, la primera etapa consiste en hacer una selección de los átomos pesados a usar, basada en varios parámetros, sobre todo, en su compatibilidad con el pH de la condición de cristalización. Posteriormente se hacen pruebas de *remojado* con cristales no útiles para estudios estructurales, como por ejemplo, cristales pequeños o encastrados. Se evalúan el tiempo de remojado y la concentración del átomo pesado a los cuales el cristal no se destruye. Después de esta etapa se comienzan a hacer pruebas con cristales grandes y se evalúa la calidad de difracción y el daño por radiación, ya que los cristales remojados con átomos pesados suelen ser más sensibles.

La selección de átomos pesados candidatos a usar se hizo con los *kits* disponibles comercialmente por Hampton Research (Ref. 2). Se hicieron pruebas de “supervivencia” de cristales para acetato de plomo ($\text{Pb}(\text{CH}_3\text{COO})_2$), hexacloroiridato de potasio (K_2IrCl_6), tungstato de sodio (Na_2WO_4), cloruro de holmio (HoCl_3), nitrato de mercurio ($\text{Hg}(\text{NO}_3)_2$), ácido *p*-hidroximercuribenzoico ($\text{C}_7\text{H}_5\text{HgO}_3\text{Na}$), tetracloro platinato de potasio (K_2PtCl_4), hexacloro platinato de potasio (K_2PtCl_6) y yoduro de sodio (NaI).

La primer prueba de remojado se hizo con el iridio. En este caso, se colocó el cristal en una gota que contenía 20 mM del átomo pesado y se monitoreó constantemente mientras el cristal no se

desintegraba. Durante 24 horas el cristal presentó estabilidad, después de esto, se cuarteó. Para hacer las pruebas de estabilidad de los cristales con plomo, tungsteno, holmio y nitrato de mercurio se probaron dos concentraciones: 1 y 10 mM, y se evaluaron tres tiempos para cada uno: 10 min, 1 hr y 12 hr. Los cristales remojados en plomo y tungsteno se disolvieron en 10 mM en menos de 10 minutos, sin embargo, sobrevivieron en 1 mM más de 1 hr pero no llegaron a las 12 hr. El cristal remojado en holmio sobrevivió más de 12 hrs tanto en 1mM como en 10mM. En el caso del cristal remojado en nitrato de mercurio, este sobrevivió en 10 mM por más de 10 min, pero no llegó a 1 hr, en cambio, sí sobrevivió más de 12 hr en 1 mM. Para el hexacloro platinato y tetracloro platinato se probaron 1 mM, 100 μ M y 10 μ M, y ambos cristales fueron estables en las tres concentraciones por más de 12 hrs. En el caso del ácido *p*-hidroximercuribenzoico sólo se probó la concentración de 1 mM. Para el yoduro de sodio se probaron tres concentraciones de NaI: 1 M, 500 mM y 300 mM, y el tiempo se mantuvo constante en 10 segundos.

En base a estos resultados, se colectaron los siguientes datos de difracción: El primero fue un cristal remojado en 20 mM de K_2IrCl_6 durante 20 minutos y se expuso a los rayos X durante 60 minutos por marco, este cristal difractó a 1.86 Å de resolución. Un segundo conjunto de datos se colectó con un cristal remojado en 20 mM de K_2IrCl_6 durante 3.5 hrs, este cristal fue expuesto a los rayos X 80 minutos por marco y difractó a 2.98 Å de resolución. La baja resolución no fue el único problema en estos datos, sino que además, presentaron un cambio de celda, de $a=b=113.14$ Å, $c=63.82$ Å, $\alpha=\beta=\gamma=90^\circ$ a $a=b=159.91$ Å, $c=65.4$ Å, $\alpha=\beta=\gamma=90^\circ$. Los programas que utilizan la señal de derivados metálicos pueden aceptar pequeñas diferencias en las celdas entre los datos nativos y los derivados (ya que es lo esperado, al incorporar átomos pesados al cristal), pero una diferencia tan grande convierte a este conjunto de datos en algo complicado para trabajar. Este cambio de celda no significa que se haya expandido de 115.39 Å a 159.91 Å, sino que, una variación en las partículas dentro de la celda hace que ya no se puedan relacionar por simple traslación de la misma manera, es decir, que el cambio en la celda no es de tamaño sino de relación de operadores simétricos.

Para la derivada con acetato de mercurio, se colectaron datos de un cristal remojado en 10 mM de $Hg(NO_3)_2$ durante 15 minutos. El cristal tuvo un tiempo de exposición a los rayos X de 120 minutos y difractó a 3.5 Å de resolución. Para la derivada con Plomo, el cristal se remojó 30 minutos en 1 mM de $Pb(CH_3COO)_2$, fue expuesto a los rayos X 70 minutos por marco y difractó a 2.79 Å. La derivada de ácido *p*-hidroximercuribenzoico difractó a 3.18 Å de resolución y fue con un cristal remojado 20 minutos en 1 mM del compuesto metálico ($C_7H_5HgO_3Na$) y expuesto a los rayos X 90

minutos por marco.

La derivada de hexacloro platinato fue la que difractó a mayor resolución de todos los datos colectados, llegó hasta 1.78 Å de resolución. Se utilizó uno de los cristales que creció en la solución 32 del *kit Cryo*, que era considerablemente mayor que los cristales obtenidos posteriormente. Este cristal fue remojado 9 horas en 1 mM de K_2PtCl_6 y expuesto a la radiación 20 minutos por marco. De las derivadas de yodo que se probaron, la que difractó fue aquella donde el cristal se remojó en 300 mM de NaI por 10 segundos, fue expuesto a los rayos X 40 minutos por marco y difractó a 2.99 Å de resolución. Además de estas derivadas, se colectó un nuevo conjunto de datos nativos con un cristal que creció en la solución 32 del *kit Cryo*.

El primer paso en el procesamiento de los datos consiste en concatenar o unir estos conjuntos de datos en un mismo archivo, para que puedan ser *escalados*. Puesto de que los datos del cristal nativo y de cada derivada son colectados independientemente y que el tamaño del cristal y el tiempo de exposición son diferentes para cada uno de ellos, la intensidad de los puntos de difracción correspondientes entre cada conjunto de datos es diferente, de tal forma que, para comparar los datos de un cristal con otro es necesario hacer un *escalamiento*.

Posteriormente se hace un mapa de *Patterson* de diferencia entre un derivado y el nativo. Esta comparación nos da *picos iniciales*, que en teoría corresponden con los átomos pesados, a partir de los cuales se pueden buscar más átomos pesados. Una vez determinada la ubicación de estos átomos y sus fases calculadas, se comienza a tratar de calcular las fases para la proteína. Existen varios criterios numéricos que permiten evaluar la certeza de los cálculos en cada uno de estos pasos.

El análisis del escalamiento de los datos mostró que, de existir átomos pesados incorporados, estos son muy pocos en cada derivada y con ocupaciones más o menos altas, o en su defecto, son varios pero que tienen baja ocupación, en este caso, la búsqueda de átomos pesados es más complicada. Los mapas de *Patterson* obtenidos para cada derivada mostraron, en su mayoría, picos en *posiciones especiales*, es decir, picos en los ejes de simetría del grupo espacial. Es posible que los átomos pesados puedan incorporarse en los ejes de simetría de los grupos espaciales, sin embargo, la mayoría de las veces estos picos son sólo ruido y en un principio no deben tomarse en cuenta. La búsqueda adicional de átomos pesados, basada a partir de aquellos que no estaban en *posiciones especiales*, dio como resultado varios átomos por derivada pero con baja ocupación. Debido a esto, es

difícil determinar cuales de estos son reales y cuales son resultado del ruido.

El trabajo con estos datos ha mostrado que, aun con las derivadas combinadas, las posiciones de los átomos pesados encontradas y sus fases calculadas no nos permiten tener “*poder de fase*” suficiente para determinar las fases de la proteína y construir un modelo a partir de una densidad electrónica. Es necesario obtener una derivada que tenga pocos sitios con muy alta ocupación.

El análisis de estos datos, a pesar de que no nos dan señal suficiente para resolver las fases de la proteína, permiten diseñar una estrategia de remojado de cristales con la intención de mejorar las ocupaciones de los átomos pesados.

Se colectaron datos de difracción de cristales remojados en: 2 mM de platino durante 20 minutos, 1.5 mM de platino durante 3 horas, 2 mM de holmio durante 6 horas, 1 mM de mercurio durante 6 horas, 1 mM de plomo durante 30 minutos, 0.75 mM de plomo durante 10 hrs y 1mM de plomo durante 10 horas y un nuevo cristal nativo con un cristal que creció en 1.7 M de fosfatos.

Cada uno de los derivados que tienen calidad suficiente para ser usados en MIR fueron procesados en el grupo espacial $P4_22_12$ como datos anómalos y no anómalos. Se calcularon los mapas de *Patterson* y se hicieron búsquedas de átomos pesados para cada caso. El análisis del procesamiento de datos muestra, en primer lugar, que la señal anómala es muy pobre y se encuentra debajo del borde de lo útil, por lo tanto, intentar usarla para localizar átomos pesados incrementa el ruido en el procesamiento. La calidad e integridad de los datos procesados como no anómalos es mejor.

Además de estos datos, se colectaron otras tres derivadas de platino, una de holmio y una de plomo, que eran las que mostraban diferencias con los datos nativos, sugiriendo que tenían átomos pesados. Estos datos se colectaron en una fuente de energía de Sincrotrón que permitió tener una mejor calidad e integridad de datos que los colectados en el difractor de ánodo rotatorio. También se colectaron un conjunto de datos de un cristal no remojado en solución de átomos pesados con la intención de usar una técnica llamada *Sulfur-SAD*, en la que se usa la señal anómala del azufre de la proteína para resolver las fases. Esta señal anómala es más débil que la de los átomos pesados, sin embargo, ha sido útil en algunos casos ya que, si se conoce la secuencia, se tiene la ventaja de saber cuantos átomos de azufre hay por proteína, que en este caso son 7, y además sus ocupaciones

son totales.

El procesamiento de estos datos no dio una respuesta satisfactoria para el problema de fases, pues ninguna derivada, incluyendo la técnica de *sulfur-SAD*, por sí misma tienen poder de fase suficiente que permita generar mapas de densidad electrónica y construir un modelo, por lo que es necesario seguir buscando combinatorias adecuadas. Es probable que los datos colectados hasta ahora permitan resolver el problema de fase, sin embargo, no es posible estimar el tiempo necesario para ello.

8. CONCLUSIONES

La purificación de las proteínas ThiDN_Tm, ThiDN_Pf y ThiN_Ss fue adecuada para hacer pruebas de cristalización y obtener cristales con características necesarias para realizar estudios de difracción de rayos X. Sin embargo, en las digestiones hechas con trombina para retirar la marca de histidinas, la proteína se precipitaba y daba un rendimiento menor de proteína útil para pruebas de cristalización. Lo mismo ocurrió con la proteína purificada por medio de calor.

Se obtuvieron cristales de las tres proteínas. Estos cristales fueron de tamaño adecuado para hacer estudios de difracción de rayos X. No obstante, los cristales de las proteínas ThiDN_Tm y ThiDN_Pf difractaron a baja resolución, independientemente del tamaño del cristal y de la fuente de rayos X usada. Ambos presentaron un valor de B de Wilson mayor de 60, indicando parámetros de movilidad atómica altos. Un aspecto común en estas proteínas es que tienen dos dominios, y es probable que exista cierta movilidad entre ellos, dando como resultado diferencias entre las proteínas dentro del cristal que se refleja en un valor de B de Wilson alto y como consecuencia una baja resolución intrínseca de los cristales.

A pesar de la baja resolución de los datos se logró resolver parcialmente el Reemplazo molecular en los cristales de ThiDN_Tm y ThiDN_Pf ubicando la posición relativa de los monómeros dentro del cristal, pero no fue posible construir un modelo para los átomos individuales ni ubicar al dominio N.

Los experimentos de SAD/MAD y SIR/MIR hechos con los cristales de ThiDN_Tm y ThiDN_Pf para resolver el problema de fase no permitieron determinar la posición de los átomos pesados dentro del cristal, muy probablemente debido a la movilidad que presentan los monómeros, generando una deslocalización de los átomos pesados.

Los experimentos de SIR/MIR hechos con los cristales de ThiN_Ss no permitieron determinar la posición de los átomos pesados dentro del cristal debido a la baja señal que se detectaba. A pesar de los múltiples intentos por conseguir poder de

fase que permitiera construir un modelo de esta proteína no se encontró una combinación adecuada de derivados metálicos que permitiera resolver el problema de fase.

Se cumplieron los objetivos de obtener cristales de las proteínas ThiDN_Tm, ThiDN_Pf y ThiN_Ss y hacer estudios de difracción de rayos X. Aún así, los datos obtenidos no permitieron determinar su estructura.

9. PERSPECTIVAS

Como perspectivas para este trabajo podemos mencionar las siguientes:

Realizar más pruebas de remojado de cristales de la proteína ThiN_Ss con átomos pesados. Para ello se necesitaría preparar una purificación de proteína a partir de varios litros de cultivo bacteriano para colocar el mayor número posible de condiciones de cristalización, de tal manera que se tengan suficientes cristales disponibles para realizar las pruebas de supervivencia de cristales y la colecta de datos necesarios.

De la misma manera, se puede hacer una preparación de proteína ThiDN_Pf expresada con seleniometioninas, buscar condiciones de cristalización en donde se formen cristales deseables para experimentos de difracción y con ello colectar datos de calidad suficiente que permitan resolver la estructura.

Por otro lado, también es factible separar los dominios *D* y *N* en las proteínas ThiDN_Tm y ThiDN_Pf para realizar pruebas de cristalización solamente del dominio que no tiene homólogos con estructura conocida. Es probable que estos dominios separados cristalicen en formas que permitan la expresión con seleniometioninas o la incorporación de átomos pesados que faciliten la determinación de la estructura. Sin embargo, sería necesario hacer estudios posteriores que permitan construir un modelo de la enzima híbrida completa para estudiar su funcionamiento.

Otro camino a tomar es intentar cristalizar alguna proteína de este grupo perteneciente a otro microorganismo. Determinar su estructura permitiría, con los datos colectados resolver por reemplazo molecular el problema de fase de las demás proteínas, teniendo de esta manera, al menos cuatro estructuras diferentes que permitirían hacer un estudio estructural comparativo más extenso sobre estas enzimas análogas.

9. REFERENCIAS

1. **Asherie, N.** 2004. "*Protein crystallization and phase diagrams*". *Methods* 34: 266-272
2. **Barynin, V.**, Whittaker, M., Antonyuk, S., Lamzin, V., Harrison, P., Artymiuk, P., Whittaker, J. 2001. "*Crystal Structure of Manganese Catalase from Lactobacillus plantarum*". *Structure* Vol 9: 725-738.
3. **Begley, T. P.**, Downs, M. D., Ealick, S. E., McLafferty, F. W. 1999. "*Thiamin biosynthesis in prokaryotes*". *Arch Microbiol.* Vol 171: 293-300.
4. **Bergfors, T.** 1999. "*Proteín Crystallization Techniques, Strategies and Tips*". Ed. IUL Biotechnology Series.
5. **Boggon, T.**, Shapiro, L. 2000, "*Screening for phasing atoms in protein crystallography*". *Structure* Vol 8: 143-149.
6. **Branden, C.**, Tooze, J. 1999. "*Introduction to Protein Structure*". Ed. Garland.
7. **Brunger, A. T.**, Adams, P. D., Clore, G. M., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., Warren, G. L. 1998. "*Crystallography & NMR System (CNS), A new software suite for macromolecular structure determination*". *Acta Cryst.* D54: 905-921.
8. **Campobasso, N.**, Mathews, I. I., Begley, T. P., Ealick, S. E. 2000. "*Crystal Structure of 4-Methyl-5- β -hydroxyethylthiazole Kinase from Basillus subtilis at 1.5 Å Resolution*". *Biochemistry.* Vol 39: 7868-7877.
9. **Collaborative Computational Project.** 1994. "*The CCP4 suite: Programs for protein crystallography*". *Acta Crystallogr.* D50: 760-763.
10. **Cheng, G.**, Bennett, E., Begley, T., Ealick E. 2002, "*Crystal Structure of 4-Amino-5hydroxymethyl 2-Methylpyrimidine Phosphate Kinase from Salmonella typhimurium at 2.3 Å resolution*". *Structure* Vol 10: 225-235.
11. **Chiu, H-J.**, Reddick, J J., Begley, T. P., Ealick, S. E. 1999. "*Crystal Structure of Thiamin Phosphate Synthase from Basillus subtilis at 1.25 Å Resolution*". *Biochemistry.* Vol 38: 6460-6470.
12. **Dauter, Z.**, Dauter, M., Rajashankar, KR. 2000. "*Novel approach to phasing proteins: derivatization by short cryo-soaking with halides*". *Acta Crystallogr* D56: 232-237.
13. **Doublié, S.** 1997. "*Preparation of Selenomethionyl for Phase Determination*". *Methods in Enzymology.* Vol 276: 523-530.

14. **Drenth, J.** 1999. "*Principles of protein x-ray crystallography*". Ed. Springer
15. **Galperin, M. Y.,** Walker, D R. Koonin, E V. 1998. "*Analogous Enzyme: Independent Inventions in Enzyme Evolution*". Genome Research. Vol 8: 779-790.
16. **Garman, E.** 1999. "Leaving no element of doubt: analysis of proteins using microPIXE". Biophysical Soc. Vol 7: 291-299.
17. **Hedstrom, L.** 2002. "*Serine Protease Mechanism and Specificity*". Chem. Rev. Vol 101: 4501-4523.
18. **Hegy, H.,** Gerstein, H. 1999. "*The Relationship between Protein Structure and Function: a Comprehensive Survey with Application to the Yeast Genome*". J. Mol. Biol. Vol 288: 147-164.
19. **Horjales, E.,** Altamirano, M., Calcagno, M., Garrat, R., Oliva, G. 1999. "*The allosteric transition of glucosamine-6-phosphate deaminase: the structure of the T state at 2.3 Å resolution*". Structure Vol 7 No5: 527-537.
20. **Janssen, S.,** Schäfer, G., Anemüller, S., Moll, R. 1997. "*A Succinate Dehydrogenase with Novel Structure and Properties from the Hyperthermophilic Archaeon Sulfolobus acidocaldarius: Genetic and Biophysical Characterization*" J. Bacteriology. Vol 197 No17: 5560-5569.
21. **Jeffery, C.** 2003. "*Moonlighting proteins: old proteins learning new tricks*". Trends in Genetics. Vol 19 No.8: 415-417.
22. **Jude, M.,** Banerjee, A., Haldar, M., Manokaran, S., Roy, B., Mallik, S., Srivastava, D., Christianson, D. 2006. "*Ultrahigh Resolution Crystal Structures of Human Carbonic Anhydrases I and II Complexed with "Two-Prong" Inhibitors Reveal the Molecular Basis of High Affinity*". J. Am. Chem. Soc. Vol 128(9): 3011-3018.
23. **Jurgenson, C.,** Begley, T., Ealick, S. 2009. "*The Structural and Biochemical Foundations of Thiamin Biosynthesis*". Annu. Rev. Biochem. Vol 78: 569-603.
24. **Kabsh, W.** 1993. "*Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants*". J. Appl. Cryst. Vol 26: 795-800.
25. **Karp, P.,** Krummenacker, M., Paley, S., Wagg, J. 1999. "Integrated pathway-genome databases and their role in drug discovery". Trends Biotechnology. Vol 17: 275-281.
26. **Kissinger, C.,** Gohlhaar, D., Fogel, D. 1999. "Rapid automated molecular replacement by evolutionary search". Acta Crystallographica, D55: 484-491.
27. **Koonin, E. V.,** Mushegian, A R. Bork, P. 1996. "*Non-orthologous gene displacement*". Trends Genetics. Vol 12: 334-336.

28. **Krishna, S.**, Grishin, N. 2004. "Structurally Analogous proteins do exist!". *Structure*. Vol 12: 1125-1127.
29. **Lehninger**. 1997. "Biochemistry". Ed. Omega.
30. **Lane, T.**, Saito, M., George, G., Pickering, I., Prince, R., Morel, F. 2005. "Isolation and Preliminary Characterization of a Cadmium Carbonic Anhydrase from a Marine Diatom". *Nature* Vol 435: 42.
31. **Leslie, A.** 2006. "The integration of macromolecular diffraction data". *Acta Cryst. D62*: 48-57.
32. **Liljas, A.**, Laurberg, M. 2000. "A wheel invented three times". *EMBO Reports* Vol 1: 16-17.
33. **Lim, S.**, Springstead, J., Yu, M., Bartkowski, W., Schröder, I., Monbouquette, H. 2009. "Characterization of a key trifunctional enzyme for aromatic amino acid biosynthesis in *Archaeoglobus fulgidus*". *Extremophiles* Vol 13: 191-198.
34. **Morett, E.**, Korb J, Rajan E, Saab-Rincon G, Olvera L, Olvera M, Schmidt S, Snel B, Bork P. 2003. "Systematic discovery of analogous enzymes in thiamin biosynthesis". *Nature biotechnology*. Vol 21: 790-795.
35. **Morett, E.**, Saab-Rincón, G., Olvera, L., Olvera, M., Flores, H., Grande, R. 2008. "Sensitive Genome-Wide Screen for Low Secondary Enzymatic Activities: The YjbQ Family Shows Thiamin Phosphate Synthase Activity". *J. Mol. Biol.* Vol 376: 839-853.
36. **Mount, D.** 2001. "Bioinformatics, sequence and genome analysis". Cold Spring Harbor Laboratory Press.
37. **Navaza, J.** 1994. "AMoRe: an automated package for molecular replacement" *Acta Cryst.* A50: 157-163.
38. **Nelson, K.**, et al, 1999. "Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*". *Nature* 399: 323-329
39. **Oyedotun, K.**, Lemire, B. 2004. "The Quaternary Structure of the *Saccharomyces cerevisiae* Succinate Dehydrogenase" *J. of Biol. Chem.* Vol 279 No 10: 9424-9431.
40. **Reddick, J.**, Nicewonger, R., Begley, T. 2001. "Mechanistic Studies on Thiamin Phosphate Synthase: Evidence for a Dissociative Mechanism". *Biochemistry*. Vol. 40: 10095-10102.
41. **Saridakis, E.**, Chayen, N. 2003. "Systematic Improvement of Protein Crystals by Determining the Supersolubility Curves of Phase Diagrams". *Biophysical Journal*. Vol 84: 1218-1222.
42. **Settembre, E.**, Begley, T., Ealick, S. 2003. "Structural biology of enzymes of the thiamin biosynthesis pathway". *Current Opinion in Structural Biology* Vol 13: 739-747.

43. **Snel, B.**, Lehmann, G., Bork, P., Huynen, M. A.2000.”*STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene*”. Nucleic Acid Research. Vol 28: 3442-3444.
44. **Sternberg, M.** 1996.”*Protein Structure Prediction*”. The Practical Approach Series.
45. **Supuran, C.** 2008. “*Carbonic Anhydrases - An Overview*”. Current Pharmaceutical Design Vol 14: 603-614.
46. **Todd, A.**, Orengo, C., Thomson, J. 2002. “*Plasticity of enzyme active sites*”. Trends Biochemical sciences Vol 27 No 8: 419-126.
47. **Temperini, C.**, Scozzafava, A., Supuran, C. 2008. “*Carbonic Anhydrase Activation and the Drug Design*”. Current Pharmaceutical Design. Vol 14: 708-715.
48. **Tripp, B.**, Smith, K., Ferry, J. 2001. “*Carbonic Anhydrase: New Insights for an Ancient Enzyme*”. The Journal of Biological Chemistry. Vol. 276, No. 52: 48615-48618.
49. Vagin, A., Teplyakov, A. 1997. “*MOLREP: an Automated Program for Molecular Replacement*” J. Appl. Cryst. Vol 30: 1022-1025
50. **Xu, Y.**, Feng, L., Jeffrey, P., Shi, Y., Morel F. 2008. “*Structure and metal exchange in the cadmium carbonic anhydrase of marine diatoms*”. Nature. Vol 452: 56-62.
51. **Yankovskaya, V.**, Horsefield, R., Törnroth, S., Luna-Chavez, C., Hiyoshi, H., Léger, B., Cecchini, G., Iwata, S. 2003. “*Architecture of Succinate Dehydrogenase and Reactive Oxygen Species Generation*”. Science Vol 299: 700-704.

REFERENCIAS DE INTERNET

Ref. 1.- http://www.ncbi.nlm.nih.gov/Complete_Genomes/AnalEnzymes.html

Ref. 1.- <http://www.hamptonresearch.com/>