



**UNIVERSIDAD NACIONAL  
AUTÓNOMA DE MÉXICO**

---

**PROGRAMA DE MAESTRÍA Y  
DOCTORADO EN CIENCIAS  
BIOQUÍMICAS  
INSTITUTO DE BIOTECNOLOGÍA  
DEPARTAMENTO DE INGENIERÍA CELULAR Y  
BIOCATÁLISIS**

**“ANÁLISIS DE POSICIONES  
ESTADÍSTICAMENTE ACOPLADAS EN  
FAMILIAS DE PROTEÍNAS ORTÓLOGAS CON  
PLEGAMIENTO DE BARRIL TIM”**

**T E S I S**  
QUE PARA OBTENER EL GRADO ACADÉMICO  
DE  
**MAESTRO EN CIENCIAS**

**P R E S E N T A :**

**Dagoberto Armenta Medina**

**DIRECTOR DE TESIS:** Dr. Lorenzo Patrick Segovia  
Forcella



**CUERNAVACA, MAYO, 2009**



Universidad Nacional  
Autónoma de México



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

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

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

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

## **AGRADECIMIENTOS:**

A CONACYT por la beca que me fue otorgada durante la realización del posgrado en Maestría en Ciencias.

A la unidad de docencia por los trámites realizados durante mi estancia en el Instituto de Biotecnología.

A mi comité tutorial por contribuir en mi formación y con ello en el desarrollo crítico de mi trabajo.

A los miembros de mi jurado de examen por las observaciones constructivas realizadas al presente trabajo.

A todos los miembros y ex miembros del grupo de Lorenzo que he tenido la suerte de conocer. Por su contribución en mi aprendizaje y su amistad.

A todos mis amigos los cuales han sido fuente inagotable de energía y diversidad de sentimientos, especialmente a Emilia.

Por sus opiniones y libertades delegadas en la realización de mi trabajo al Dr. Lorenzo Segovia Forcella.

Al Dr. Ernesto Pérez Rueda por su apoyo en los diferentes aspectos técnicos de mi trabajo.

A toda mi familia especialmente a mis padres y hermanos por su apoyo incondicional en todos los aspectos de mi vida.

El presente trabajo se realizó bajo la dirección del Dr. Lorenzo Segovia Forcella, en el Instituto de Biotecnología de la Universidad Nacional Autónoma de México. Durante la realización del mismo se recibió una beca del Consejo Nacional de Ciencia y Tecnología (CONACYT).

COMITÉ TUTORAL:

Dr. Lorenzo Segovia Forcella (IBT-UNAM)  
Dr. Joel Osuna Quintero (IBT-UNAM)  
Dr. Alejandro Fernández Velasco (FM-UNAM)

JURADO DE EXAMEN:

Dr. Xavier Soberón Mainero (IBT-UNAM)  
    Presidente  
Dr. Enrique Rudiño Piñera (IBT-UNAM)  
    Secretario  
Dr. Arturo Becerra Bracho (FC-UNAM)  
    Vocal  
Dr. León Martínez Castilla (FQ-UNAM)  
    Suplente  
Dr. Lorenzo Segovia Forcella (IBT-UNAM)  
    Suplente

# Índice

RESUMEN.....	2
1- INTRODUCCIÓN .....	3
1.1-Ingeniería de Proteínas.....	3
1.2-Relación secuencia, estructura, dinámica y función de proteínas.....	5
1.2.1-El Nivel de la secuencia.....	6
1.2.2-El Nivel de la estructura y dinámica.....	7
1.2.3-El Nivel de la Función.....	9
1.3- Los Potenciales en base al conocimiento de secuencia, estructura, dinámica y función de proteínas.....	10
1.3.1- Los Potenciales en base al conocimiento de estructuras y dinámica de proteínas.....	11
1.3.2- Los Potenciales en base al conocimiento de la función de proteínas.....	11
1.3.3- Los Potenciales en base al conocimiento de secuencias.....	13
2-ANTECEDENTES.....	15
2.1-Diseño computacional de proteínas.....	15
2.2-Características del Barril ( $\beta/\alpha$ ) <sub>8</sub> .....	18
2.3-Enzima modelo PRAI (Fosforibosil antranilato isomerasa).....	19
3-HIPÓTESIS Y OBJETIVOS.....	21
4-DESARROLLO Y RESULTADOS DEL ANÁLISIS DE ACOPLAMIENTO ESTADÍSTICO .....	22
4.1-Funcionamiento general del SCA.....	22
4.2-Obtención de las secuencias de la familia de PRAI.....	23
4.3-Evaluación de la diversidad y tamaño del MSA para el SCA.....	25
4.4-Validación de perturbaciones dentro del MSA.....	28
4.5-Resultados del SCA con el valor de corte óptimo.....	29
4.6-Agrupamiento y clasificación de sitios acoplados en PRAI.....	32
4.7-Análisis de la familia ortóloga de proteínas (PRAI).....	33
4.7.1-GRUPO GAP .....	35
4.7.2-GRUPO GBP .....	37
4.7.3-Sitios acoplados de gran relevancia en la dinámica de plegamiento y estabilidad del barril.....	42
4.7.4-GRUPO CON (altamente conservados).....	45
4.8-Análisis de una segunda familia ortóloga de proteínas, Triosafosfato Fosfato Isomerasa (TIM) con plegamiento de barril TIM ( $\beta/\alpha$ ) <sub>8</sub> .....	47
4.8.1-Grupo G2.....	50
4.8.2-Grupo G1.....	51
4.8.3-Grupo G3.....	53
4.8.4-Grupo G4.....	54
4.8.5- Red de sitios acoplados y conservados .....	58
5-CONCLUSIONES: .....	62
6-PERSPECTIVAS .....	64
7-REFERENCIAS.....	65

## RESUMEN

El presente trabajo, pretende lograr un mayor entendimiento sobre la información contenida en las secuencias proteicas, que determinan el plegamiento del barril TIM ( $\beta/\alpha$ )<sub>8</sub>. Partiendo de nuestras familias modelo, Fosforibosil Antranilato Isomerasa (PRAI) y Triosafofato Isomerasa (TIM), se capturaron las secuencias ortólogas a partir de las cuales se generó un alineamiento múltiple (MSA) para cada una de ellas. Posteriormente los MSA's fueron analizados con el *software* basado en el método de acoplamiento estadístico (*statistical coupling análisis*, en inglés), mediante el cual fue posible establecer los valores de acoplamiento para cada posición de los alineamientos, descubriendo patrones globales de co-evolución para cada familia.

Esto se fundamenta en que el patrón de co-evolución de interacciones funcionales es disperso, por lo tanto un pequeño grupo de posiciones co-evolucionan, entre la mayoría que están desacopladas. Los residuos co-evolucionados fuertemente, están espacialmente organizados dentro de redes físicamente conectadas uniendo sitios funcionales, de gran relevancia para el proceso de plegamiento y función de las proteínas.

Los análisis de acoplamiento estadístico para las dos familias de barriles (PRAI y TIM), revelan sitios de gran importancia involucrados en los procesos de estabilidad, unión a sustrato y tentativamente en redes que gobiernan la dinámica necesaria para la catálisis. Estos sitios identificados podrían estar co-evolucionando para el mantenimiento del sitio activo y la adecuada relación estabilidad y actividad en las enzimas. Las posiciones con altos valores de acoplamiento estadístico, son mencionadas como críticas para el mantenimiento del plegamiento y actividad del barril, esto de acuerdo a estudios de mutagénesis, dinámicas moleculares, cristalográficos y termodinámicos. Finalmente nuestros resultados aportan información valiosa para diseños posteriores de barriles así como los sitios potenciales a ser mutados y su posterior estudio.

# **“Análisis de posiciones estadísticamente acopladas en familias de proteínas ortólogas con plegamiento de barril TIM”**

## **1- INTRODUCCIÓN**

En las últimas décadas el campo de estudio de la bioquímica ha dado lugar al surgimiento de la disciplina denominada ingeniería de proteínas, una creciente y prometedora herramienta de la biotecnología. Esta herramienta dirigirá en gran medida al desarrollo de la biotecnología moderna mediante su contribución en el diseño de proteínas, principalmente enzimas, con una amplia variedad de aplicaciones en las industrias farmacéutica, química y alimentaría <sup>(1, 2)</sup>.

Estas proteínas diseñadas por el hombre, son creadas con la intención de exhibir estabilidad, selectividad a nuevos sustratos y actividades catalíticas mejoradas, inclusive persiguiendo metas tan ambiciosas como la generación de actividades catalíticas *de novo* <sup>(2-4)</sup>.

### **1.1-Ingeniería de Proteínas**

Existen dos formas en las que se realiza la ingeniería de proteínas, la primera es el diseño racional y consiste en generar cambios específicos en regiones determinadas de la proteína, esta estrategia requiere de conocimientos previos acerca de la estructura y propiedades funcionales de nuestra secuencia a modificar <sup>(5)</sup>. La segunda trata de la evolución dirigida que mimetiza el proceso de evolución natural de los organismos vivos, el cual se basa en mutación-selección, y se le considera dirigida por que lo que se selecciona son las propiedades requeridas para la aplicación deseada <sup>(6)</sup>.

Entre los resultados más impactantes obtenidos de la evolución dirigida podemos mencionar los de estabilidad térmica <sup>(7)</sup> y resistencia a solventes <sup>(8)</sup>. En cuanto al diseño racional existen un gran número de ejemplos en la literatura los cuales describen modificaciones mínimas en los sitios activos, logrando con esto, variar el repertorio catalítico de las enzimas <sup>(9)</sup>.

Otra estrategia empleada, es la generación de anticuerpos catalíticos seleccionados por afinidad a un análogo de transición, aquí también el enfoque semi-razonal ha posibilitado la introducción de actividades presentes en la naturaleza y *de novo* <sup>(10)</sup>. De cualquier forma las actividades logradas por estas vías en la mayoría de los casos están muy lejos de ser comparables con las de enzimas presentes en la naturaleza y más aun han necesitado de un despliegue impresionante de recursos económicos y humanos.

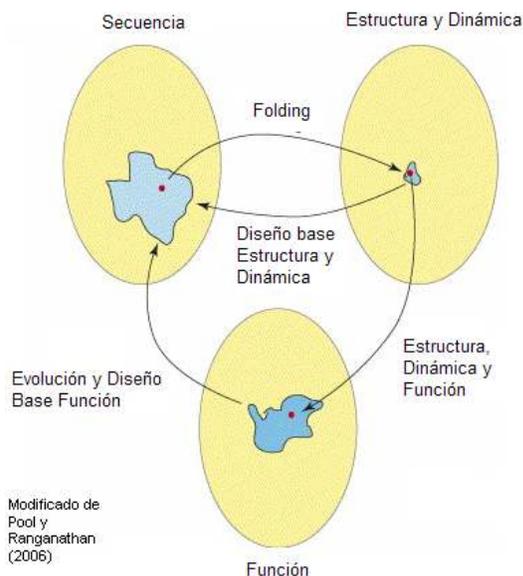
En el contexto mencionado anteriormente se denota la importancia de contribuir al mejoramiento en el desarrollo del diseño racional mediante enfoques que nos ayuden a entender cómo la naturaleza logra generar secuencias de proteínas que mantienen un conjunto de plegamientos y funciones estables. A partir de este entendimiento será posible la creación de reglas más generalizadas y eficientes en el diseño de proteínas.

El conocimiento de las reglas o patrones que determinan las características mencionadas anteriormente en las proteínas, es catalogado como un "santo grial" de la bioquímica, por la gran relevancia que tendría el hecho de poder manipular y construir proteínas con las características deseadas <sup>(11)</sup>. Por lo tanto, el primer paso para la creación de ciertas reglas o patrones de diseño generalizado en la ingeniería de proteínas, es responder a preguntas clave, que contribuyan a una mayor comprensión de la relación existente entre la secuencia, estructura, dinámica y función de proteínas <sup>(12, 13)</sup>.

El estudio de la relación secuencia, estructura, dinámica y función, es fundamental en el diseño y selección de nuevas variantes enzimáticas con características específicas que nos permitan acercarnos cada vez más a la meta final de la ingeniería de proteínas, el diseño *de novo* de enzimas con la función y propiedades que sean requeridas para cualquier tipo de proceso en una escala de tiempo y costo factible.

## 1.2-Relación secuencia, estructura, dinámica y función de proteínas

A un nivel primario, tenemos que la relación secuencia, estructura, dinámica y función es determinante en el problema del plegamiento de proteínas (*Protein Folding Problem*), donde sólo la información correcta contenida en la secuencia, decretará el éxito de las conformaciones predeterminadas dentro un vasto número de posibles conformaciones <sup>(14)</sup>. A su vez, estas conformaciones predeterminadas, son las determinantes de la función <sup>(15, 16)</sup>.



**Fig1.-** En el esquema de la figura 1, se pueden observar óvalos en amarillo que simbolizan los 4 niveles: secuencia, estructura, dinámica y función de las proteínas, las flechas denotan la interrelación. Una secuencia natural presente en el nivel de las secuencias corresponde a un punto rojo que pertenece a una estructura con su dinámica perteneciente a un conjunto de estados nativos estructurales, los cuales, finalmente, están interrelacionados para determinar la función.

En la figura 1, los letreros mostrados entre estos distintos niveles representan varios problemas fundamentales en el análisis y diseño de proteínas, los cuales dependen en última instancia del entendimiento de patrones globales de fuerzas físicas entre los átomos <sup>(17)</sup>.

El principal problema al querer resolver las preguntas de ¿cuál es la relación existente entre secuencia, estructura, dinámica y función de proteínas? y ¿cómo utilizar esta información en la ingeniería de proteínas?, si lo vemos desde el punto de vista de los aminoácidos que son los bloques estructurales de las proteínas, es la vasta complejidad de interacciones cooperativas entre residuos dentro de su estructura, proceso en el cual la energía libre de un residuo depende de los demás residuos del conjunto al cual pertenece <sup>(18)</sup>.

Gracias a la tecnología, desarrollada en los últimos años, se han detonado grandes avances en los 4 niveles mostrados en el esquema de la figura 1. Enseguida trataré de comentar de manera general la situación particular actual de cada uno de estos 4 niveles: secuencia, estructura, dinámica y función.

### **1.2.1-El Nivel de la secuencia.**

En el nivel de las secuencias el proceso evolutivo a través de los años ha generado una enorme variedad de estas, que en la actualidad han sido resueltas por métodos automatizados de secuenciación. Estos métodos cada vez más eficientes han permitido la secuenciación genómica de organismos completos, estos hechos han arrojado cerca de 462764 entradas de secuencias de proteínas para el 14 de abril del año 2009, las cuales se encuentran disponibles para su análisis en bases de datos públicas <sup>(19)</sup>.

La importancia del estudio de este gran número de secuencias radica en que en las mismas se encuentra depositada la información que dicta el plegamiento y función de las proteínas. Haciendo la analogía de las proteínas con una máquina, la información de las secuencias sería lo equivalente a tener los comandos que

crean la máquina y dirigen la funcionalidad de esta, sin conocer el lenguaje en el que estos comandos fueron escritos.

El concepto de ingeniería inversa el cual es aplicado en el mundo de las máquinas se describe como el proceso de descubrir los principios tecnológicos de un dispositivo, objeto o sistema, a través de razonamiento abductivo de su estructura, función y operación. Actualmente este concepto es aplicado para el entendimiento y diseño de proteínas.

Gracias a la información generada en otros niveles por medio de ingeniería inversa, conociendo la estructura y la función de proteínas, ha sido posible inferir medianamente el significado de estos comandos implícitos en la secuencia. Algunos ejemplos de la aplicación del proceso de ingeniería inversa son la predicción de estructura y función a partir de secuencia de proteínas.

### **1.2.2-El Nivel de la estructura y dinámica.**

Actualmente un acercamiento en el entendimiento del mundo de las estructuras, se ha dado debido a los importantes avances obtenidos en las técnicas de difracción de Rayos X y Resonancia Magnética Nuclear (NMR por sus siglas en ingles), lo cual ha hecho posible conocer el plegamiento de varias familias de proteínas.

Estas estructuras se encuentran depositadas en la base de datos del *Protein Data Bank* (PDB) <sup>(20)</sup>, la cual para Septiembre 4 del 2007 contenía 45632 entradas, sin embargo, generalmente estas sólo representan un número limitado de una población total de conformaciones o sub-estados funcionales presentes en las proteínas <sup>(21)</sup>. Además, no sabemos que proporción de los plegamientos existentes en el universo está representada en el PDB.

Aunque, como mencionamos en el párrafo anterior, las estructuras de mínima energía pueden ser determinadas experimentalmente, el entendimiento de su dinámica de plegamiento, es decir la comprensión de la serie de pasos necesario para llevar una secuencia lineal de aminoácidos a su conformación nativa <sup>(14)</sup> y el conocimiento del como las fuerzas moleculares dirigen las transiciones entre los diferentes sub-estados, permanece poco explorado <sup>(22-25)</sup>.

Si hacemos la analogía, esto es equivalente a tomar una foto a alguna máquina, la foto probablemente revelara sólo un instante o algunos instantes, en el caso de contar con estructuras en distintas conformación, de todo el proceso dinámico realizado por la máquina y por consecuencia se podría conocer algunas piezas que la componen.

Por ejemplo; analizando estas fotos, se podría observar que la máquina esta constituida de engranes y palancas, entre otros componentes, lo cual es de gran relevancia. Sin embargo, este conocimiento no es suficiente para entender el funcionamiento colectivo de las piezas, ni mucho menos como se llevó a cabo la creación o construcción de dicha máquina. De igual forma que en la foto de la máquina, las estructuras determinadas mediante difracción de Rayos X o NMR revelan detalles de los componentes estructurales de la proteína, sin embargo, esta información por sí sola no es suficiente para entender como actúan de manera dinámica estos componentes estructurales que le confieren la funcionalidad a la proteína, ni mucho menos el proceso en el cual secuencia lineal de proteínas llega a adoptar tal conformación tridimensional.

Las técnicas experimentales tradicionales, como la microcalorimetría y el dicroísmo circular, entre otras, han sido utilizadas ampliamente para el estudio del entendimiento de la dinámica de plegamiento de proteínas. Este estudio, hoy en día, se ha estado complementado con el uso de técnicas modernas principalmente basadas en NMR <sup>(26)</sup>.

Por otro lado, ha emergido un especial interés en los últimos años sobre la dinámica intrínseca y extrínseca de las proteínas, derivado de estudios que sugieren una íntima relación entre la movilidad molecular y la catálisis <sup>(15, 27-29)</sup>. De igual forma, las técnicas modernas basadas principalmente en NMR y transferencia de energía por resonancia de fluorescencia (FRET), han permitido estudios más detallados sobre este fenómeno <sup>(22, 25, 26)</sup>.

En ambos casos, las técnicas computacionales como simulaciones de Dinámicas Moleculares (MD por sus siglas en inglés), Monte Carlo (MC), Análisis de Modos Normales (NMA por sus siglas en inglés), han contribuido en complementar los huecos dejados por la imposibilidad experimental derivado del gran reto que implica el análisis de estos fenómenos de movilidad, aportando en varios casos información valiosa <sup>(30-32)</sup>.

### **1.2.3-El Nivel de la Función**

Por último tenemos el nivel de la función, el entendimiento de esta área viene de la mano con la caracterización funcional de proteínas, principalmente mediante la ayuda de herramientas bioquímicas y de biología molecular <sup>(33)</sup>. El proceso llevado a cabo para este tipo de caracterización funcional ha implicado gran esfuerzo, dinero y tiempo, por lo que últimamente, con ayuda de herramientas computacionales, se ha tratado de predecir la función a partir de secuencias y el reconocimiento de motivos estructurales funcionales <sup>(34, 35)</sup>.

Generalmente, la mayor parte de lo que se conoce de la función en las enzimas es el tipo de reacción que lleva a cabo, debido a los productos generados a partir de determinados sustratos. La biología molecular, en algunos casos gracias a estudios de mutagénesis, ha revelado sitios indispensables para la catálisis y/o unión a sustratos, con lo que se ha posibilitado el conocimiento de la química de las reacciones <sup>(36)</sup>; sin embargo, en la mayoría de los casos se desconoce de manera integral la relación estructura, dinámica y función de las proteínas, es decir,

el como estas se adecuan en espacio y tiempo para llevar a cabo determinada función.

Retomando la analogía comentada en el mundo de las estructuras contando sólo con esta información sería como conocer la función de la máquina pero sin saber que partes la componen, por lo tanto tampoco se podría conocer como funcionan entre ellas, ni mucho menos con que información fue creada.

A pesar del gran avance en cada uno de los 4 niveles mencionados anteriormente, es necesario utilizar y desarrollar metodologías, que permitan extraer información útil de cada nivel en particular. De esta manera podrá ser aplicada ya sea en el diseño o comprensión profunda del funcionamiento de las proteínas. También, es importante que estas metodologías sean capaces de complementarse entre sí, para ser evaluadas desde una perspectiva holística.

### **1.3- Los Potenciales en base al conocimiento de secuencia, estructura, dinámica y función de proteínas.**

Un concepto importante que surge del intento de manejar de manera eficiente la gran cantidad de información aportada por los niveles de la secuencia, estructura, dinámica y función, son los potenciales en base a conocimientos (KBPs, por sus siglas en inglés). Los KBPs, representan parámetros estadísticos extraídos de las bases de datos de los distintos niveles de las proteínas, capturando de manera empírica, aspectos físico-químicos de las proteínas<sup>(17)</sup>.

Estos potenciales, de manera individual como en conjunto, pueden aportar información importante que contribuya eficientemente en la selección de proteínas correctamente plegadas y posiblemente secuencias artificiales funcionales. Adicionalmente la información presente en uno de los niveles puede utilizarse para predecir aspectos de los demás niveles.

### **1.3.1- Los Potenciales en base al conocimiento de estructuras y dinámica de proteínas.**

El primero de estos, está basado en conocimientos derivados de librerías estructurales extraídas del PDB, incluidas pero no limitados a los siguientes: potenciales de solvatación, librería de rótameros, librería de fragmentos cortos observados en el PDB, potenciales de enlace de hidrógeno y electrostáticos <sup>(17)</sup>. Algunos de estos potenciales, son combinados, logrando algoritmos exitosos de predicción en el diseño e identificación de plegamientos.

En el nivel de las dinámicas, tenemos que hoy en día es factible extraer información de los datos experimentales y simulaciones con ayuda de herramientas computacionales, primordialmente análisis de co-varianzas o componentes principales, sobre la trayectoria de la dinámica representativa de las cuales se puede obtener gran número de datos <sup>(37)</sup>. Actualmente, existe un esfuerzo por mantener bases de datos y servidores donde sea posible extraer información de este tipo <sup>(38-40)</sup>. De igual forma, se han complementado algunos algoritmos principalmente de *docking*, con información de la dinámica de las proteínas; además, algunos autores comentan la necesidad de incluir de manera recurrente este tipo de información en los algoritmos de diseño <sup>(41-43)</sup>.

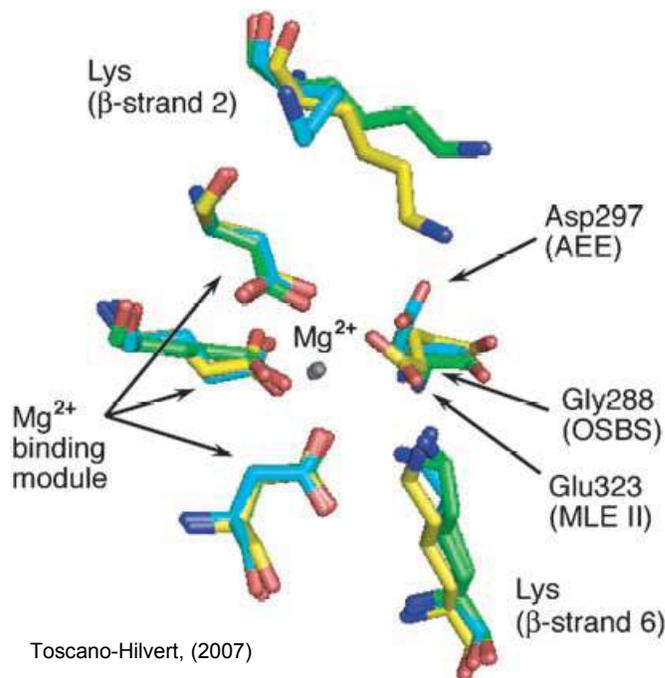
### **1.3.2- Los Potenciales en base al conocimiento de la función de proteínas.**

Bases de datos como UniProtKB <sup>(44)</sup> y Brenda <sup>(45)</sup> cuentan con anotaciones funcionales de un gran número de familias de proteínas. Otra base de datos más especializada al respecto, es *The Catalytic Site Atlas (CSA)* <sup>(46)</sup>, el cual provee anotaciones de los residuos catalíticos para proteínas presentes en el PDB. La base de datos esta constituida de dos tipos de sitios anotados: un conjunto anotado a mano, que contiene información extraída primordialmente de literatura usando un criterio definido para la asignación de residuos catalíticos y adicionalmente un

conjunto de homólogos, que contiene información inferida por PSI-BLAST <sup>(47)</sup> y alineamiento de secuencias a una del conjunto original.

Los conocimientos de potenciales, derivados de estas bases de datos funcionales, se centran principalmente en correlacionar variaciones funcionales con variaciones tanto en secuencia como estructura. Lo anterior, puede ser de gran ayuda en el diseño, mediante la introducción de cambios en una secuencia blanco, los cuales especifiquen determinantes funcionales en la proteína.

Un ejemplo de la utilización de este conocimiento en el diseño, se observa en el trabajo realizado por el grupo de Gerlt, en base al conocimiento de la función y los sitios funcionales de la enzima orto-succinilbenzoato sintasa (OSBS), fueron capaces de introducir la actividad en un par de miembros de la superfamilia, específicamente a la L-Ala-D/L-Glu epimerasa (AEE) y la muconato lactonizante enzima II (MLE II) mediante diferentes métodos: mutagénesis sitio dirigida en AEE y mutagénesis al azar en MLE II <sup>(9)</sup>.



**Fig.2-** Sobreposición estructural de AEE (en azul claro), OSBS (verde) y finalmente MLE II en (amarillo); oxígeno en rojo, azul nitrógeno y finalmente magnesio en gris (9).

Comparaciones de las estructuras de AEE y la silvestre OSBS se detallan en la figura 2. Aunque los residuos catalíticos y el cofactor  $Mg^{2+}$  se observan en la misma posición en ambos sitios activos, la cadena lateral del Asp297 en AEE, se sobrepone con el lado succinil del producto orto-succinilbenzoato (OSB) en el sitio activo de OSBS.

Para remover este impedimento estérico con el sustrato, la variante Asp297Gli de AEE fue diseñada mediante mutación dirigida, exhibiendo actividades con el número de recambio ( $K_{cat} = 2.5 \times 10^{-3} s^{-1}$ ), en paralelo, mutaciones al azar en MLE II permitieron seleccionar una mutación Glu323Gli la cual le confiere más de  $10^6$  veces mayor actividad OSBS ( $K_{cat} = 1.5 s^{-1}$ ).

Estos resultados demuestran como diferentes estrategias pueden converger en resultados positivos en la ingeniería de proteínas. Estas estrategias también se usan complementariamente, donde por medio de mutaciones sitio dirigidas los ingenieros de proteínas varían el repertorio catalítico de las enzimas, y posteriormente utilizan evolución dirigida para aumentar las actividades logradas.

### **1.3.3- Los Potenciales en base al conocimiento de secuencias**

Finalmente, los potenciales derivados de bases de datos de secuencias, es una realidad. A partir de alineamientos múltiples de secuencias (MSA, por sus siglas en inglés) de enzimas homólogas, ha sido posible entender relaciones evolutivas como la conservación y la co-evolución de los caracteres de estado de las posiciones dentro del MSA. De estas relaciones se obtiene información útil que nos permite inferir aspectos tanto de la estructura como de la función de las proteínas (48, 49).

Generalmente los valores de identidad se utilizan para agrupar las secuencias en familias de proteínas homólogas; también como se mencionó previamente, con los alineamientos, es posible identificar sitios o motivos conservados generalmente

pertenecientes a residuos relevantes en las familias de proteínas. El conocimiento de estos sitios y/o motivos de secuencias caracterizadas funcionalmente hacen posible predecir la función de secuencias no caracterizadas.

Por otro lado, en los últimos años, la información generada a partir de los MSA de familias de proteínas, ha sido explotada fuertemente mediante algoritmos de análisis estadístico. Un ejemplo es el algoritmo desarrollado por el grupo de Ranganathan, el cual basándose en la identificación de residuos que están acoplados evolutivamente, han logrado el diseño exitoso de proteínas con características similares a las presentes en la naturaleza <sup>(18, 50)</sup>. En la siguiente sección del escrito comentaremos a profundidad sobre la relevancia de este método.

## 2-ANTECEDENTES

### 2.1-Diseño computacional de proteínas

El diseño de proteínas ha evolucionado de la mano con nuestro entendimiento de la secuencia, la dinámica, la función y la estructura bio-macromolecular de proteínas. Este crecimiento se ha obtenido principalmente gracias al avance de las metodologías computacionales, la biología molecular y la bioquímica, así como de conocimientos basados en los niveles de secuencia, estructura, dinámica y función de proteínas. Gracias a esto, cada vez tenemos mejores modelos que predicen los parámetros termodinámicos de las proteínas con gran calidad, los cuales a su vez están siendo utilizados de manera integradora <sup>(17)</sup>.

Varios grupos de investigación científica, han aplicado y experimentado evaluaciones sistemáticas a métodos cuantitativos para el diseño de proteínas, con la meta de desarrollar algoritmos de diseño generales <sup>(51)</sup>. El primer diseño de una proteína con una topología en su estructura novedosa, inexistente en la naturaleza, fue realizada por el grupo de David Baker. Se diseñó la proteína Top 7 compuesta por 93 residuos, la cual experimentalmente se demostró estar plegada y ser muy estable termodinámicamente. Lo anterior fue posible mediante la integración de algoritmos de predicción de estructura *ab initio*, refinamiento de energía a nivel atómico y diseño de secuencia con el programa *Rosetta* <sup>(52)</sup>.

El grupo de Baker, formando parte de un equipo interdisciplinario nuevamente nos sorprendió con la hazaña mostrada en dos artículos recientes, donde describen un método para diseñar enzimas que catalizan reacciones no encontradas en la naturaleza y demuestran su uso para dos transformaciones químicas diferentes. Estos ejemplos son los más sorprendentes en cuanto al diseño de actividades de novo <sup>(42, 53)</sup>.

La idea detrás de estos trabajos fue simple pero llevarlo a la práctica necesitó de una gran cantidad de recursos humanos y económicos. La idea y ejecución se basan en modelar el estado de transición de cada una de las reacciones y estabilizarlas mediante cambios cuidadosos de grupos químicos alrededor del sitio activo. Se implanta el sitio activo resultante en una proteína existente y posteriormente, se altera su secuencia de aminoácidos para acomodar los cambios. Para lograr estos objetivos, el grupo de Baker se apoyó en algoritmos computacionales basados en técnicas de *hashing* geométrico, previamente aplicados en la búsqueda de sitios funcionales para un gran número de estructuras, combinándolo principalmente con algoritmos de diseño de proteínas <sup>(54)</sup>.

Los trabajos comentados en el diseño de actividades *de novo*, son ejemplo del éxito de combinar principalmente los potenciales basados en el conocimiento de la secuencia, la estructura y la función de proteínas, por otro lado mencionan que actualmente están trabajando en complementar sus algoritmos con información que les permita introducir flexibilidad a sus estructuras debido a la importancia de esta en el proceso de diseño para lograr un mejoramiento en la capacidad de catálisis.

El grupo de Mayo también ha logrado la creación de proteínas *de novo*, usando valores (scores) de funciones, encargadas en evaluar si la manera de aplicar algunas propiedades termodinámicas es exitosa para el diseño de proteínas. Estas funciones se basaron principalmente en el conocimiento de secuencias y estructuras, mediante las cuales fue posible obtener un péptido con menos de un 40% de identidad con cualquier secuencia conocida, sin contener ningún sitio de unión a metales o cisteínas. El péptido obtenido fue el pda8d con la propiedad de alta solubilidad <sup>(55)</sup>.

Otro de los resultados más impactantes es el obtenido por el grupo de Ranganathan, el cual mediante el conocimiento de secuencias, utiliza información contenida en el MSA de dominios WW (dominios de unión a proteínas con motivos

particulares de prolinas AP]-P-P-[AP]-Y, y/o motivos conteniendo fosfoserina-fosfotreonina), para generar sus contrapartes artificiales con el programa Statistical Coupling Analysis (SCA, por sus siglas en inglés). Los resultados mostraron alrededor de un 25% de efectividad en el diseño ya que de 43 secuencias artificiales generadas 12 lograron plegamientos nativos con propiedades termodinámicas similares validados mediante análisis de NMR <sup>(18)</sup>. En un segundo artículo demuestran que sus secuencias artificiales conservan la funcionalidad de las secuencias naturales<sup>(50)</sup>.

Con estos antecedentes, partiendo de nuestras enzimas modelo: Fosforibosil Antranilato Isomerasa (PRAI) y Triosafosfato Isomerasa (TIM), se capturaron las secuencias homólogas contenidas en bases de datos públicas y posteriormente fueron analizadas con software basado en el método SCA, facilitado por el grupo de Ranganathan. Este método establece los valores de conservación en cada posición de las secuencias alineadas. Estas a su vez, están dadas en forma de energía estadística, parámetro que mide la desviación de la distribución de los valores observados de los aminoácidos de los valores medios encontrados en todas las proteínas <sup>(18)</sup>. El método se fundamenta, en que el patrón global de co-evolución de interacciones funcionales es disperso, por lo tanto un pequeño grupo de posiciones co-evolucionan, entre la mayoría que están desacopladas y los residuos co-evolucionados fuertemente están espacialmente organizados dentro de redes físicamente conectadas de gran relevancia para las proteínas <sup>(50)</sup>.

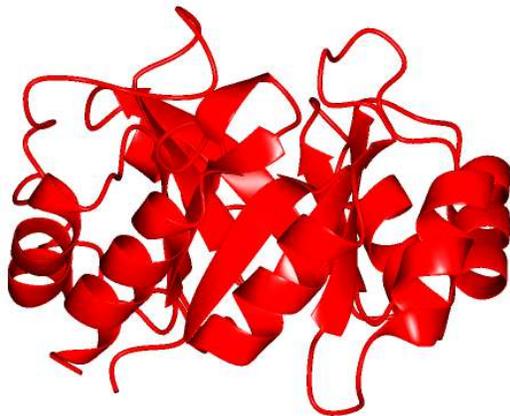
La determinación de estos sitios con alto grado de conservación y acoplamiento estadístico será de gran utilidad en la comprensión profunda de la relación secuencia, estructura, dinámica y función y a la hora de aplicar las estrategias tanto de evolución dirigida como racional dentro del conjunto de proteínas con plegamiento barril TIM, permitiendo una estrategia que se limite a hacer cambios discriminando sobre los sitios que hipotéticamente poseen mayor peso funcional de acuerdo a los altos valores de conservación y acoplamiento estadístico, de aquellos que poseen valores bajos y por lo tanto menor peso funcional.

El conocimiento de sitios acoplados durante la evolución hace posible reconocer que cambios deben de realizarse concertadamente evitando alterar el patrón de co-evolución de las proteínas.

## 2.2- Características del Barril ( $\beta/\alpha$ )<sub>8</sub>

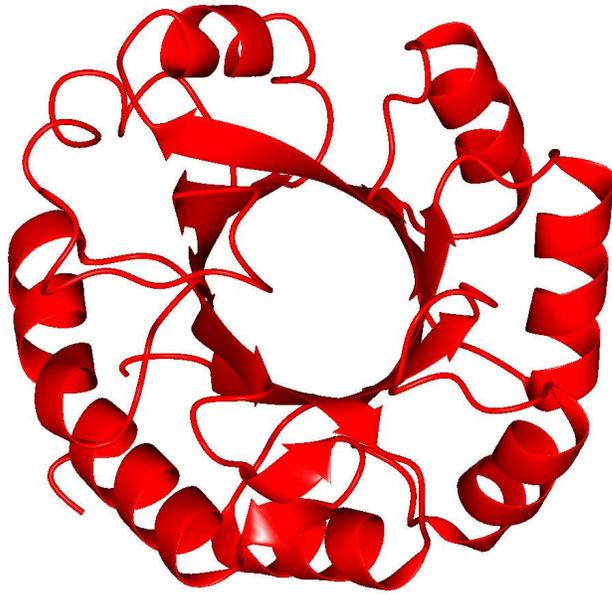
El barril TIM es el plegamiento más versátil que conocemos hasta el momento al presentar cinco de los seis tipos de reacciones definidos por la comisión de enzimas (EC). Debido a esto es un buen candidato para la evolución *in vitro* y racional hacia nuevas actividades <sup>(56)</sup>.

Este plegamiento es el más abundante en la base de datos estructurales, representando el 10% de las estructuras conocidas depositadas en el *protein data bank* (PDB) <sup>(57)</sup>. El barril TIM consiste en ocho repeticiones Beta-Alfa ( $\beta-\alpha$ ), ocho hojas  $\beta$  plegadas forman el centro del barril y cada una está cubierta por una  $\alpha$  hélice en el exterior.



**Fig.3-** Vista del plegamiento de barril TIM ( $\beta/\alpha$ )<sub>8</sub>.

Las asas que conectan las hojas  $\beta$  hacia las hélices  $\alpha$  son denominadas  $\beta-\alpha$  y las asas que conectan las hélices  $\alpha$  hacia las hojas  $\beta$  son denominadas  $\alpha-\beta$ . Estas asas son de longitudes variables con la tendencia de ser mayores las  $\beta-\alpha$ .



**Fig.4-** Vista de la cara catalítica del plegamiento de barril TIM ( $\beta/\alpha$ )<sub>8</sub> .

Los sitios activos de todos los barriles TIM están localizados en el extremo carboxilo terminal de las hojas  $\beta$  localizados en la parte superior del barril, de esta manera el sitio activo esta formado la mayoría de las veces por los residuos de las ocho asas  $\beta$ - $\alpha$ , mientras que las asas  $\alpha$ - $\beta$  son asociadas con la estabilidad del barril <sup>(58)</sup>.

### **2.3-Enzima modelo PRAI (Fosforibosil antranilato isomerasa)**

La Fosforibosil Antranilato Isomerasa o PRAI es una enzima que presenta el plegamiento clásico de barril TIM( $\beta/\alpha$ )<sub>8</sub>, en *Escherichia coli* se encuentra formando una enzima bi-funcional que consta de dos dominios implicados en la vía de síntesis de triptófano. Uno de los dominios (PRAI), cataliza una reacción intramolecular redox, conocida también como re-arreglo tipo amadori <sup>(59)</sup>.

Se ha postulado varias hipótesis que explican la fusión de estos dos dominios. Mediante técnicas de ingeniería recombinante han sido separados en enzimas

mono-funcionales con la misma actividad que su contraparte bi-funcional. Este hecho no es sorprendente ya que en un gran número de organismos, esta proteína no está fusionada con ningún otro dominio <sup>(60, 61)</sup>.

Actualmente en el laboratorio, esta enzima modelo ha sido utilizada con el objetivo de evaluar la flexibilidad estructural y evolutiva del barril TIM, para soportar sustituciones de gran diversidad y tamaño. Se pretende obtener complementaciones exitosas de dominios estructurales con características diversas. La metodología para lograr lo anterior ha sido mediante el intercambio de fragmentos de DNA genómico no homólogos en diferentes regiones del barril en la familia de PRAI. El presente estudio, complementa los resultados mediante el entendimiento de las determinantes estructurales que presenta, así como el descubrimiento de regiones potenciales a sustituir dentro del barril. Finalmente, contamos con una gran diversidad natural de secuencias, ya que existen un gran número reportadas en bases de datos de familias de proteínas como Pfam, en donde se encuentran representados los diferentes taxa.

### **3-HIPÓTESIS Y OBJETIVOS.**

#### **Hipótesis:**

La información contenida en las secuencias de las familias de barriles TIM( $\beta/\alpha$ )<sub>8</sub> es suficiente para determinar los sitios con acoplamientos funcionales.

#### **Objetivo General:**

Identificar los determinantes estructurales y funcionales del plegamiento mediante el conocimiento de las posiciones altamente acopladas dentro de las familias de barriles TIM ( $\beta/\alpha$ )<sub>8</sub>.

#### **Objetivos específicos:**

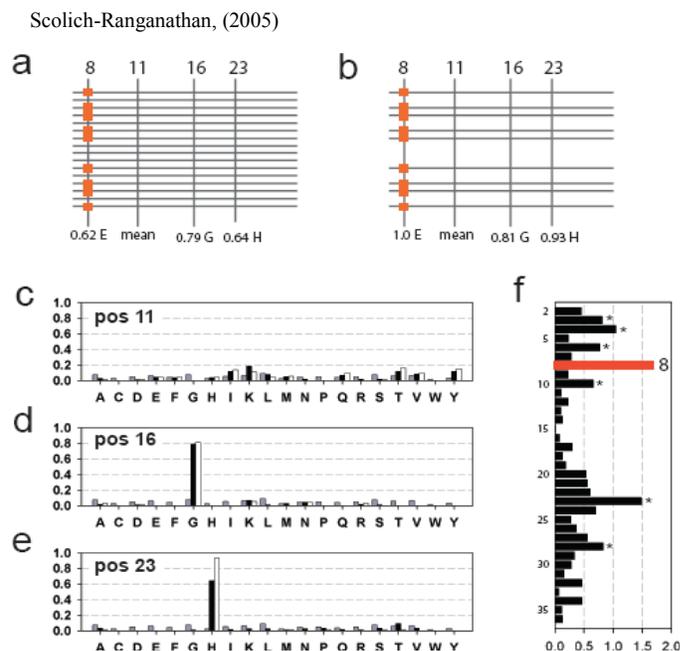
- \* Determinar los sitios con acoplamientos funcionales conservados evolutivamente a partir de secuencias homólogas de la familia funcional PRAI.
- \* Determinar los sitios con acoplamientos funcionales conservados evolutivamente a partir de secuencias homólogas de familia funcional de TIM.
- \* Comparar los resultados obtenidos del análisis de acoplamiento estadístico realizado a las dos familias funcionales PRAI y TIM pertenecientes a la familia estructural de barriles TIM( $\beta\alpha$ )<sup>8</sup>.

## 4-DESARROLLO Y RESULTADOS DEL ANÁLISIS DE ACOPLAMIENTO ESTADÍSTICO

### 4.1-Funcionamiento general del SCA.

El análisis realizado con el programa SCA, cuantitativamente mide el cambio de la distribución de un aminoácido en una posición  $j$  en el alineamiento múltiple de secuencia (MSA), dado una perturbación en otra posición  $i$  en forma de energía estadística de acoplamiento ( $\Delta\Delta G_{ij}^{stat}$ ) entre los dos sitios <sup>(62)</sup>.

Un claro ejemplo de esto se observa en la siguiente figura:



**Fig.5-** Funcionamiento General Del SCA <sup>(18)</sup>

En la figura 5, en primera instancia tenemos el sitio 8 el cual contiene al ácido glutámico (E) en un 62% del MSA, posteriormente es perturbado (se puede observar partiendo de la figura 5a) a 5b) es decir (E) se lleva al 100 % dentro del MSA mediante la eliminación de las secuencias en las que no se encuentra. La perturbación ocurrida altera el grado de conservación de los demás sitios (11,16,

23). Si algunos de ellos tiene un aumento significativo en su conservación con respecto al que tenía en el MSA original como en el caso del sitio 23, que tiene histidina (H) en 64% y con la perturbación cambia a 93%, se puede asumir que el sitio 23 esta altamente acoplado al sitio 8; por otro lado el sitio 16 que contenía un 79% de glicina (G) en el MSA original y cambió al 81%, está débilmente acoplado.

Nótese que el programa no considera aquellos sitios muy conservados que se presentan en alrededor del total dentro del MSA, ya que a la hora de ser perturbados, no ocasionaran cambios en la distribución de los demás sitios y de igual forma la perturbación en otros sitios, no cambiará la distribución de estos sitios con alto grado de conservación. Esto se puede explicar con el ejemplo anterior, si como el sitio 8, que esta conservado al 62% hubiese estado al 100%, los sitios 16 y 23 no hubiesen cambiado en conservación, ya que una perturbación en el sitio 8 tendría el mismo número de secuencias del MSA original. Lo anterior no demerita el método ya que los residuos altamente conservados por si mismos son de gran relevancia para las familias de proteínas.

## **4.2-Obtención de las secuencias de la familia de PRAI**

El primer paso para realizar el análisis de acoplamiento estadístico es encontrar las secuencias homólogas de la familia de PRAI. Para lo anterior, se procedió a hacer una búsqueda de ortólogos en el servidor GenContext <sup>(63)</sup>, actualmente alojado en la página del Instituto de Biotecnología (IBT) UNAM ([http://www.ibt.unam.mx/server/PRG.base?alterno:0,clase:binf,edi:d,tit:GeConT\\_An%E1lisis\\_contextual\\_de\\_Genomas,tra:GeConT\\_Gene\\_Context\\_Tool,final:no,tipo:doc,dir:PRG.gecont,par:F,pre:binf,final:no](http://www.ibt.unam.mx/server/PRG.base?alterno:0,clase:binf,edi:d,tit:GeConT_An%E1lisis_contextual_de_Genomas,tra:GeConT_Gene_Context_Tool,final:no,tipo:doc,dir:PRG.gecont,par:F,pre:binf,final:no)). Estas secuencias, fueron comparadas con los resultados obtenidos por otro servidor de nombre JPRED, el cual recupera secuencias para la predicción de estructuras secundarias <sup>(64)</sup>.

En la utilización de ambos servidores, la secuencia de búsqueda fue PRAI de *Thermotoga maritima* (tPRAI), para el análisis de la familia de la Triosafosfato Isomerasa se utilizó la secuencia de humano (hTIM). Los resultados obtenidos por ambos servidores eran muy similares en cuanto al tipo de secuencias que arrojaron, aunque el servidor de JPRED, recuperó mayor número de estas.

Por lo mencionado anteriormente, se procedió a tomar todas las secuencias arrojadas por el servidor JPRED, las cuales fueron inspeccionadas para los sitios catalíticos previamente reportados. Aquellas secuencias que mantenían el patrón de identidad y similitud en estas posiciones, fueron las que se utilizaron para el análisis, posteriormente fueron filtradas para evitar la redundancia de secuencias, con la intención de mantener un conjunto de secuencias representativas del proceso de evolución substancial de la familia. Finalmente, fueron eliminadas todas las secuencias que estuviesen alineadas en menos del 75% de las posiciones respecto a la secuencia pregunta, para evitar posibles secuencias incompletas.

Una vez filtradas estas secuencias, el siguiente paso fue realinearlas con el programa de alineamiento múltiple MUSCLE<sup>(65)</sup>, con los parámetros estándar. Por último el alineamiento fue refinado manualmente tratando de corregir posiciones recorridas incorrectamente por el programa. A continuación, se observa el alineamiento múltiple de secuencias (MSA), que consiste en 272 secuencias de la familia de PRAI.

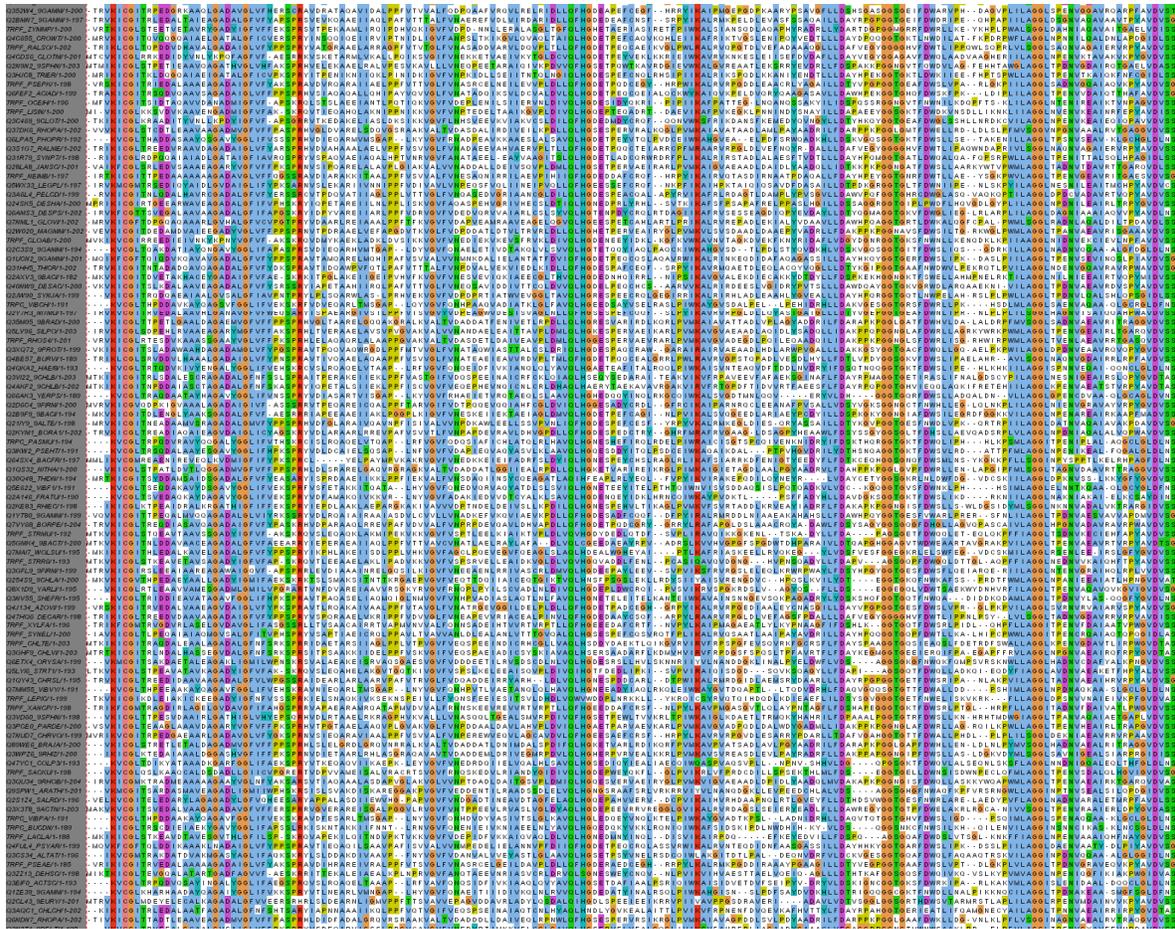


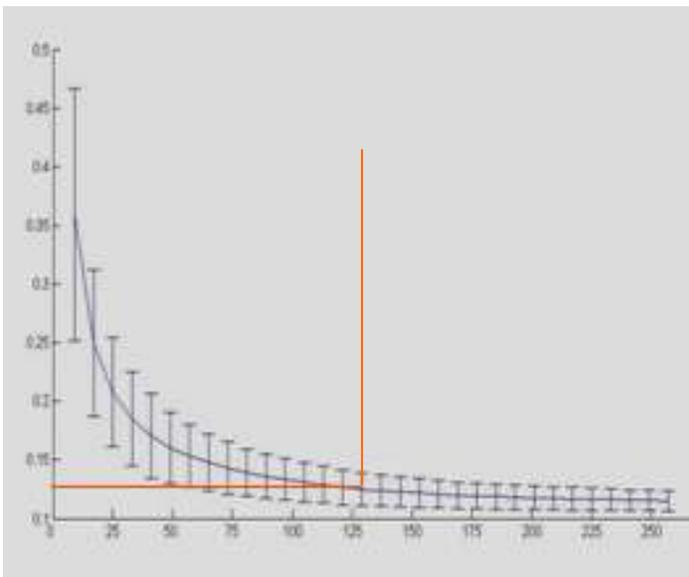
Fig.6.- Alineamiento múltiple de secuencia de la familia PRAI.

### 4.3-Evaluación de la diversidad y tamaño del MSA para el SCA

El MSA es evaluado en cuanto a su diversidad y tamaño con el programa SCA. El algoritmo, se basa en tomar los 5 sitios menos conservados con al menos el 85% de ocupación dentro del alineamiento, posteriormente se van eliminando al azar fracciones del MSA (diferentes números de secuencias de secuencias del alineamiento), se hacen 100 rondas de extracción de secuencias de manera aleatoria dentro del MSA para cada fracción, enseguida se promedian las 100 rondas y se calcula cómo cambió el grado de conservación para estos 5 sitios (18).

Se espera que estos sitios sigan manteniéndose no conservados, ya que en el alineamiento original presentan una distribución de la media de los aminoácidos presentes en todas las proteínas (como el sitio 11 de la figura 5a) y 5c)), de lo cual también se concluye que el MSA ha experimentado una evolución substancial y que la distribución de los aminoácidos en los demás sitios son efectivamente reflejo de los compromisos funcionales de la familia de proteína.

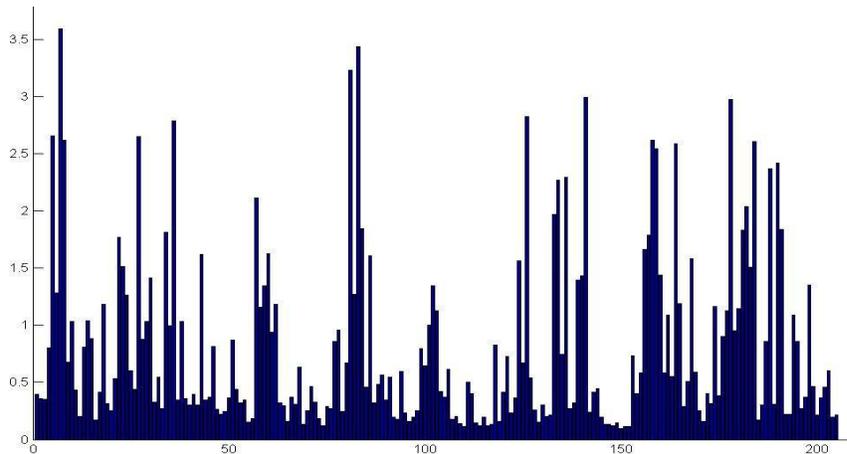
Cuando estos 5 sitios empiezan a presentar mayor conservación al ir extrayendo mayor número de secuencias es señal de que la muestra de secuencias no es representativa de la familia. Inversamente, si el patrón de conservación de estos 5 sitios se mantiene constante, con el valor de conservación similar al inicial del MSA, es un indicativo de que el alineamiento se encuentra en equilibrio estadístico. Las siguientes graficas describen estos comportamientos.



**Fig.7-** DEstat (Conservación en valores de energía estadística) VS No. de secuencias para los 5 sitios menos conservados del MSA. La intersección de las líneas anaranjadas denota el punto de corte de 130 secuencias.

En la figura 7, se puede observar claramente como a partir de 130 secuencias, se empiezan estabilizar los 5 sitios con bajos valores de conservación, esta

estabilidad refleja que el perfil de conservación en todos los sitios se comportaran de manera muy similar, ya sea que se utilicen 130 secuencias o más hacia el total del número original de secuencias del MSA, que son 272 y como con menos de 125 secuencias, los valores de conservación empiezan a subir alejándose de los valores originales del MSA.



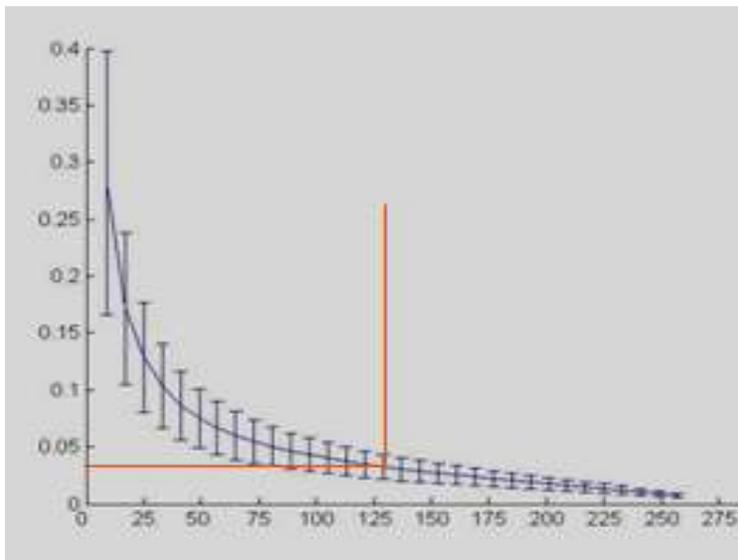
**Fig.8-** Perfil de conservación en el MSA.

La figura 8, muestra el perfil de conservación del MSA, hecho con las 272 secuencias para la familia de PRAI. Si tomamos un grupo mayor de 130 secuencias al azar del total de 272 secuencias (Ej. 135, 149, 180), el perfil de conservación global del MSA, seria muy similar al de la figura 8, ya que a partir de esas 130 secuencias el MSA ha alcanzado una evolución substancial. Por otro lado, si se agregaran más secuencias de la familia a esas 272, uno esperaría que el perfil de conservación permaneciera también muy similar, de acuerdo a la constante mostrada a partir de las 130 secuencias. De manera contraria, si se seleccionaran al azar grupos menores de 130 secuencias (Ej. 40, 80, 120) los perfiles de conservación en el MSA resultarían muy variables y no serían los representativos para la familia de PRAI.

#### 4.4-Validación de perturbaciones dentro del MSA

Esta validación sigue un algoritmo similar al mostrado anteriormente. En primer instancia, de igual forma que el primer algoritmo, el programa toma los 5 sitios menos conservados con al menos el 85% de ocupación. Enseguida, se hacen 100 rondas de extracción de secuencias de manera aleatoria dentro del MSA, para cada fracción se realiza el calculo de acoplamiento entre los 5 sitios no conservados y se determina el grado de acoplamiento entre los mismos <sup>(18)</sup>.

Si estos 5 sitios empiezan a mostrar un aumento en los valores de acoplamiento entre sí, es reflejo de que el número de secuencias dentro del alineamiento no es representativa para realizar el análisis de acoplamiento, ya que estos sitios no conservados deben presentar una distribución de la media de los aminoácidos presentes en todas las proteínas que indique que el alineamiento se encuentre en equilibrio estadístico. Por lo tanto al usar perturbaciones con esos números de secuencias del MSA el programa nos arrojaría información no confiable.



**Fig.9-** DDEstat (acoplamiento en unidades de energía estadística) VS No. de secuencias para los 5 sitios menos conservados del MSA. Intersección de líneas anaranjadas denotan el nivel de acoplamiento estadístico al valor de corte 130.

En la figura 9, se observa que a partir de 125 secuencias, el valor entre los sitios menos conservados se mantiene igual o por debajo de 0.05 de acoplamiento entre ellos. Podemos decir, que este número de secuencias caen dentro del rango validado de diversidad y tamaño del MSA que es plausible en la figura 7, donde se mantiene constante la grafica (de 130 secuencias en adelante). Por lo tanto, el valor de corte factible para las perturbaciones de la familia de proteínas PRAI, estará dado entre la región de 130 secuencias en adelante. Esto corresponde a la fracción  $130/272 \Rightarrow$  mínimo de secuencias a perturbar (MsPer), donde la formula es MsPer/total de secuencias del MSA.

El valor de corte donde se establecen las perturbaciones es arbitrario para cada familia de proteínas y hasta ahora en todos los trabajos publicados resultan intuitivos. Debemos resaltar, que el mecanismo más viable para asignar un valor de corte a las perturbaciones, es a través de combinar los resultado de los 2 métodos de validación representados en las graficas 7 y 9, siempre teniendo en cuenta que el MSA, deberá de permanecer con los valores más bajos de acoplamiento entre los 5 sitios no conservados y a su vez permitir que el MSA conserve su diversidad y suficiente número de secuencias, para ser representativo de la familia.

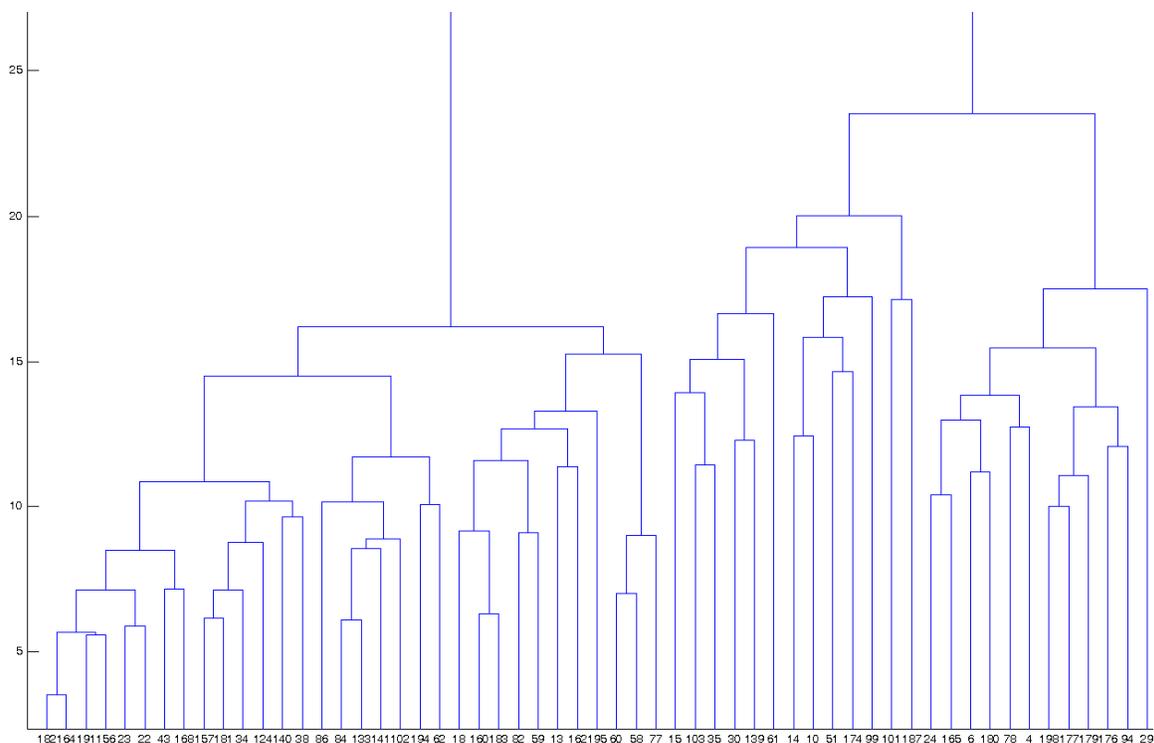
Finalmente, dada la metodología de selección, se encontró que el valor de corte optimo resultaba en 130 secuencias es decir  $130/272 = 0.47$  fracción del MSA. A continuación se explicará el proceso de cálculo del programa SCA al valor de corte seleccionado 0.47.

#### **4.5-Resultados del SCA con el valor de corte óptimo**

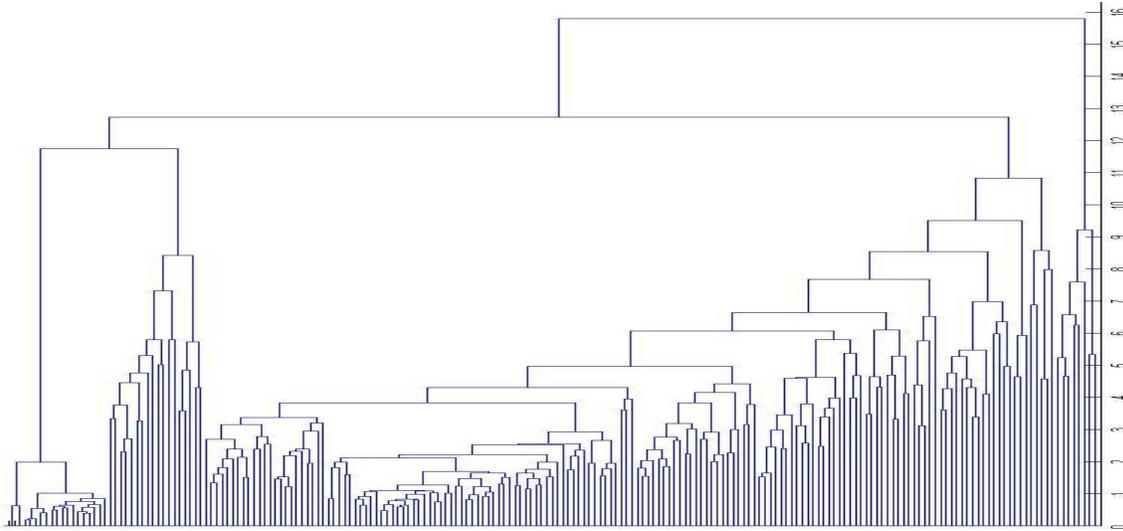
La numeración de los sitios se hace con referencia a (tPRAI), del 1 al 205, representando el total de sitios evaluados dentro del MSA. Los procesos de salida de los resultados ocurren de la manera que se describe enseguida. El primer dato que arroja el programa SCA es el número de perturbaciones posibles para el valor

de corte de 0.47. Arrojó 74 perturbaciones, es decir 74 sitios a perturbar que cumplían con al menos 0.47 de conservación. De los 74 sitios a perturbar, sólo 57 eran informativos por lo cual 17 se omiten ya que poseen valores de conservación muy altos, cercanos al 100%, y la estadística del programa no los considera. En la figura 10, se muestran los 57 sitios a perturbar, que se utilizaron para determinar la matriz de co-evolución de la familia de PRAI. El mismo dendograma también refleja la co-evolución entre los mismos 57 sitios a perturbar.

Una vez obtenidos estos sitios a perturbar se obtienen los valores de acoplamiento para todas las posiciones del N a C terminal de la familia de proteínas, (figura 11), con el algoritmo explicado anteriormente. De estos valores se obtiene la matriz global de co-evolución para PRAI (figura 12) en la cual se observan agrupamientos que reflejan los sitios mayormente acoplados entre sí y que pueden estar asociados a algún tipo de función.



**Fig.10-** Sitios validados para la realización de perturbaciones correspondientes para el análisis en SCA, este dendograma se observa en el eje de las x de la figura 12.



**Fig.11-** Patrón global de co-evolución de los 205 sitios del MSA: Posiciones del alineamiento vs Valores de acoplamiento en unidades de energía estadística (ddEstat). Este dendograma se observa en el eje y de la figura 12.

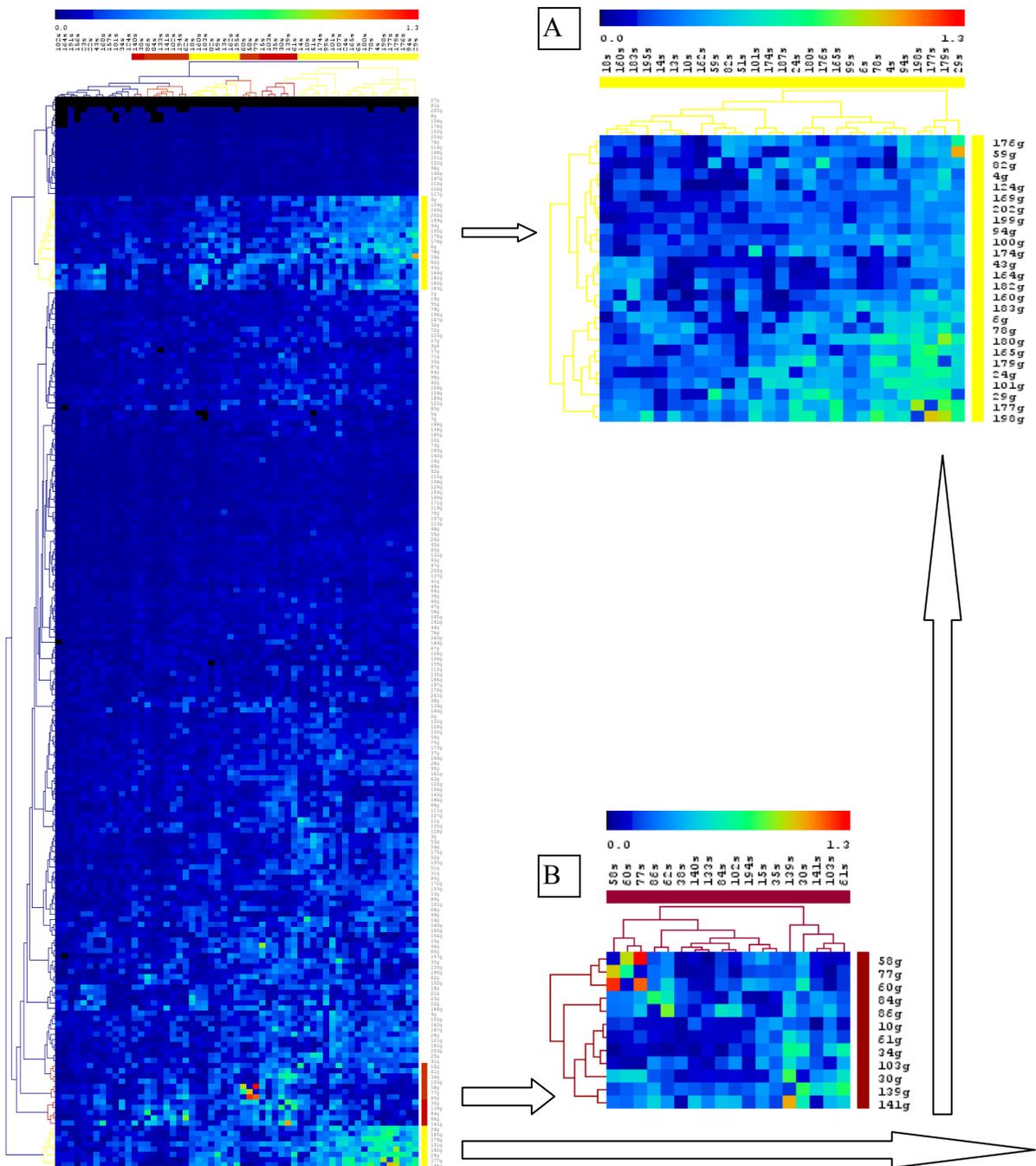
## 4.6-Agrupamiento y clasificación de sitios acoplados en PRAI.

La matriz de acoplamiento estadístico para cada familia de proteínas representa todas las perturbaciones seleccionadas con los criterios mencionados anteriormente. En el caso particular de la familia de PRAI se obtuvieron 57 posiciones. Esta matriz está compuesta por sitios que van desde N a C terminal en filas (Eje de las Y en el mapa de colores) y las perturbaciones de N a C terminal en columnas (Eje de las X en el mapa de colores). El algoritmo de agrupamiento en dos dimensiones (en inglés llamado *Two Way Clustering*) es una metodología inicialmente desarrollada para el análisis de micro arreglos de DNA y ahora utilizada para el análisis de proteínas.

La idea general es generar rondas secuenciales de agrupamiento en dos dimensiones, cada vez se extrae la sub-matriz que contiene todas las posiciones y perturbaciones, que forman grupos en iteraciones previas y que poseen valores sobresalientes de acoplamiento. Por lo tanto, cada paso iterativo es un intento de refinar la asignación de grupos, enfocándose alrededor de regiones de las posiciones  $i$  y perturbaciones  $j$ , que muestran valores significativos de acoplamiento<sup>(66)</sup> y que finalmente observamos en el mapa de colores de la figura 12. El agrupamiento se llevo a cabo con el software libre MeV4.0 microarray<sup>(67)</sup>.

El paso inicial para empezar la clasificación se da determinando los grupos con mayor grado de acoplamiento que se observan en el mapa de colores en la parte de abajo.

## 4.7-Análisis de la familia ortóloga de proteínas (PRAI).



**Fig12-** Matriz de co-evolución global de PRAI. Al lado izquierdo de la figura se encuentra la matriz madre donde se extraen los grupos de las figuras A) y B), el eje de las X representa las perturbaciones y el eje de las Y todas las posiciones del alineamiento. En las secciones B y A de la figura se observan los grupos más sobresalientes de sitios acoplados.

En el análisis de la familia PRAI se lograron obtener dos grupos sobresalientes que a continuación se muestran como resultado del proceso final de agrupamiento. De los grupos finales de la figura 12A y 12B mostrados anteriormente, se formó un grupo para cada uno de ellos conteniendo los sitios altamente acoplados. Adicionalmente se formó un grupo más con los sitios altamente conservados, ya que como hemos mencionado previamente, el algoritmo del SCA no evalúa estos sitios de gran relevancia. La intención principal de dividir los sitios en los distintos grupos es para dar una mejor explicación de los mecanismos en los que estos podrían estar involucrados.

**Tabla 1-** Grupo de sitios altamente acoplados y altamente conservados (en rojo).

GRUPO	SITIOS INTEGRANTES	TOTAL
GAP	4, 6, 13, 14, 18, 24, 29, 43, 51, 59, 78, 82, 94, 99, 100, 101, 124, 160, 162, 164, 165, 169, 174, 176, 177, 179, 180, 182, 183, 195, 198, 199, 202	33
GBP	10, 15, 30, 34, 35, 38, 58, 60, 61, 62, 77, 84, 86, 102, 103, 133, 139, 140, 141, 194, 187	21
CON	5, 7, 8, 81, 126, 178, 8, 27, 157, 158, 159, 181, 188, 190, 36, 83, 134, 136, 184	19

Recortando la matriz de acoplamiento, obtenemos los grupos con valores de acoplamientos más significativos, los cuales se encuentran en su mayoría dentro del 5 por ciento de pares correlacionados más significativos del matriz global. Los sitios se muestran en la parte superior de la tabla 1.

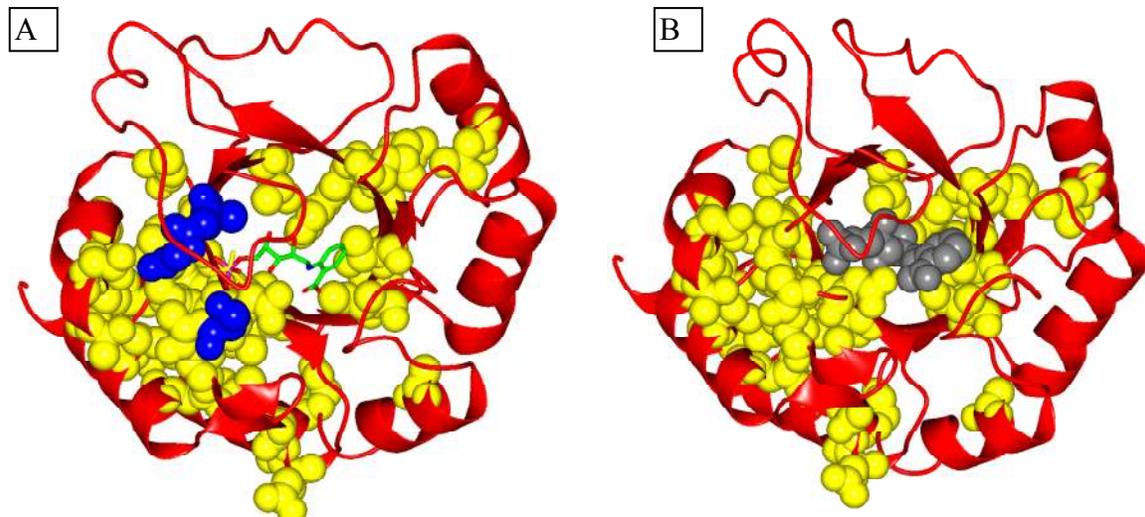
Una vez obtenidos los grupos más sobresalientes, surgen las siguientes preguntas, ¿cómo discriminar, en los grupos con altos valores de acoplamiento, los sitios que son indispensables para el plegamiento de los que son importantes para otro tipo de función?, o ¿existen sitios que estén comprometidos tanto en el plegamiento como la función biológica?

Asignar la función que representa cada grupo, es una tarea difícil y minuciosa, ya que se necesita observar sitio por sitio y comparar datos bibliográficos si es que se cuenta con ellos. Los datos provenientes de análisis termodinámicos, dinámicas moleculares y cristalográficos, son los más relevantes para hacer la comparación con la matriz de co-evolución global. Otro aspecto importante que se evalúa, es la distribución de estos sitios acoplados en la estructura de la proteína, de lo cual se proponen las posibles contribuciones funcionales de estos sitios en la familia.

#### **4.7.1-GRUPO GAP**

El grupo GAP está compuesto por todos los sitios presentes en el grupo que se observa en la figura 12A. Adicionalmente, algunos sitios altamente conservados se encuentran en regiones estructuralmente cercanas al grupo, lo que sugiere un probable nexo entre estos sitios y los altamente acoplados. Este grupo de sitios, están muy relacionado con la formación de la bolsa hidrofóbica y la hélice de unión a fosfato, de acuerdo a datos estructurales de miembros de la familia de la termófila (tPRAI) y la mesófila (ePRAI) <sup>(61, 68)</sup>.

El grupo fosfato y el ácido antranílico se encuentran a los extremos terminales de los substratos Fosboribosil Antranilato (PRA). El sitio activo de PRAI está delimitado por el sitio de unión a fosfato en un extremo y por la bolsa de unión hidrofóbica asociada al ácido antranílico de PRA en el otro extremo. En este dominio, la bolsa hidrofóbica está formada principalmente por residuos de las asas  $\beta\alpha 2$  y  $\beta\alpha 3$  y se asume que tiene la función de unión al grupo aromático del ácido antranílico <sup>(68)</sup>, figura 13.



**Fig.13-** Estructura tridimensional de *E. coli* PRAI. En esferas amarillas con radios de vandewaals se muestran los sitios pertenecientes al grupo G1P. Se observa en azul algunos sitios conservados del lado de unión al grupo fosfato del sustrato. El sustrato se muestra en palillos de color verde en la figura 13A y en esferas grises en el 13B.

En el otro extremo del sustrato, los sitios altamente acoplados 179, 180, 182, 183 del grupo GAP y algunos altamente conservados 157, 158, 159, 181 en esferas azules, forman el sitio de unión al grupo fosfato, figura 13A. En la estructura de ePRAI, dos átomos de oxígeno del ion fosfato mantienen enlaces de hidrógeno con el grupo peptidil NH de los sitios 159 y 181 (de acuerdo a la numeración correspondiente del alineamiento con la secuencia de tPRAI), así como en el grupo hidroxilo del sitio 181.

Los sitios del alineamiento 182, 183, 178, 158, 180 y 181 forman enlaces de hidrógeno, mediados por agua con un átomo de oxígeno del ion fosfato, necesario para la actividad en ePRAI. En el caso de tPRAI, estos sitios desempeñan funciones similares a su contraparte mesófila, de acuerdo con observaciones cristalográficas<sup>(61, 68)</sup>.

Actualmente, este sitio de unión a fosfato es considerado uno de los motivos más conservados estructuralmente entre las enzimas con plegamiento de barril TIM ( $\beta/\alpha$ )<sub>8</sub>, ya que aproximadamente 2/3 de las familias enzimáticas con este plegamiento utilizan sustratos o cofactores que contienen al menos un grupo

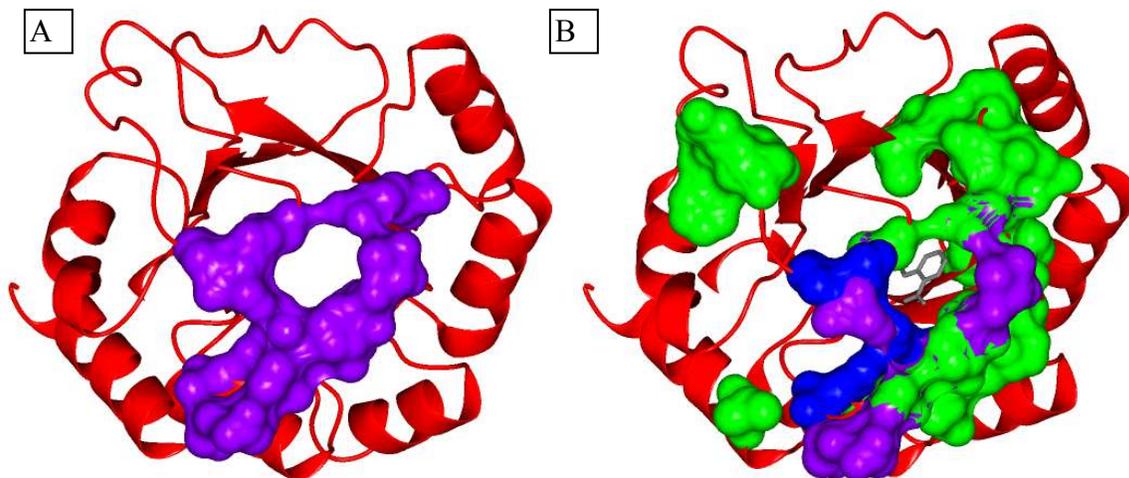
fosfato <sup>(58)</sup>. La conservación, en términos de la presencia del sitio de unión a fosfatos en los barriles, no sólo describe su importancia, sino también su conservación estructural en regiones determinadas del plegamiento. Los barriles de la subunidad  $\alpha$  de la Triptofano sintasa, PRAI e IGPS mantienen contactos con grupos fosfatos en regiones estructuralmente equivalentes. La sobreposición de estas tres estructuras en la posición de los átomos de fósforo tiene un RMSD menor a 1 amstrong ( $1\text{\AA}$ ) <sup>(68)</sup>.

Los datos descritos para este grupo, sugieren que los sitios que se encuentran en las regiones cercanas a la unión del sustrato, mantienen una co-evolución substancial. Debido a que la unión del sustrato, es determinante para tener una adecuada catálisis, es posible que estos sitios co-evolucionen de manera que la enzima mantenga una relación adecuada de actividad y estabilidad en su plegamiento. Esto podría estar ocurriendo, mediante cambios en los sitios que co-evolucionan, los cuales sacrifican la estabilidad del plegamiento por la unión del sustrato y viceversa. Esta co-evolución podría depender de la presión de selección en el contexto de la enzima o simplemente soluciones equivalente que mantienen relaciones similares de actividad y estabilidad.

#### **4.7.2-GRUPO GBP**

Un canal cerca del pocket hidrofóbico en la estructura de ePRAI conecta el sitio activo al solvente con un corte transversal de cerca de  $7 \times 5 \text{\AA}$ . Está formado por los residuos 33 al 36 y los sitios 61, 133, 134, 135, 136 y 184 al 187 <sup>(68)</sup>. La mayoría de estos sitios, conforman el grupo GBP, por lo cual parece que este grupo ha co-evolucionado mayormente para mantener la formación del canal.

El canal que conecta el sitio activo con el solvente está predominantemente formado por residuos de las asas  $\beta\alpha 2$  y  $\beta\alpha 6$ , figura 14A.

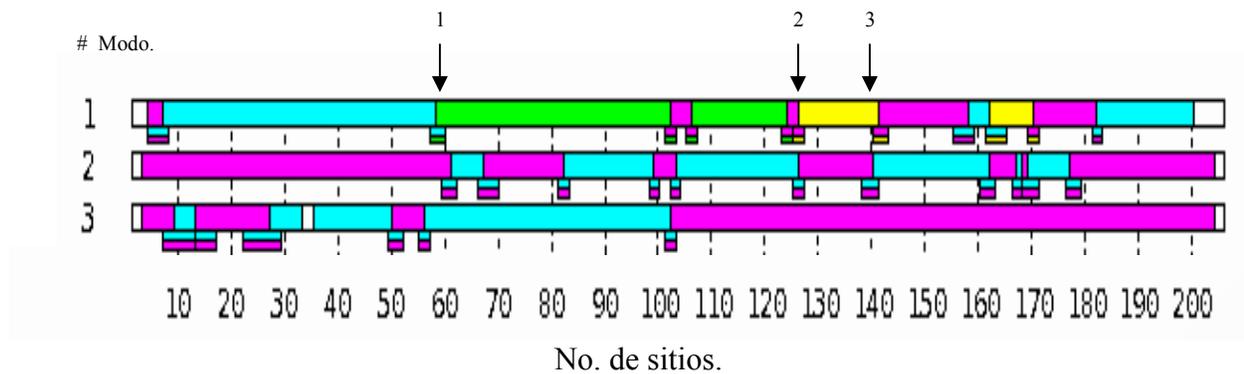


**Fig.14-** Estructura tridimensional de *E. coli* PRAI. En la superficie de color violeta se observa un canal que conecta el sitio activo con el solvente descrito bibliográficamente. En la parte B de la figura el grupo G2P es representado en color verde conformando en mayor medida este canal en conjunto con algunos sitios conservados mostrados en azul.

La alta movilidad del asa  $\beta\alpha_6$  en PRAI, se observa experimentalmente en los factores-B de las estructuras de los miembros de la familia, simulaciones de dinámicas moleculares y NMA. En conjunto, estas metodologías muestran que se lleva a cabo un considerable re-arreglo conformacional, en esta región de la proteína. El asa  $\beta\alpha_6$  equivalente mantiene gran flexibilidad en los barriles de las estructuras sin ligando en la unidad  $\alpha$  de la triptófano sintasa, la ribulosa-1,5-bifosfatasa carboxilasa/oxigenasa, triosafosfato isomerasa, indol glicerol fosfato sintasa y se encuentra formando mayoritariamente la tapa del sitio activo <sup>(68-70)</sup>.

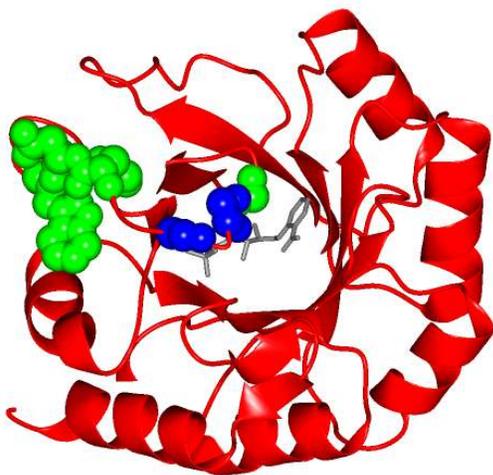
Algunos trabajos mencionan la relevancia de regiones bisagras dentro de las proteínas las cuales facilitan la dinámica de dominios móviles <sup>(24, 71)</sup>. La predicción de estas bisagras ha sido posible utilizando metodologías como NMA y algoritmos basados en información estructural como el utilizado por el servidor Dyndom <sup>(72)</sup>. En esta dirección realizamos un análisis de modos normales a nuestra proteína modelo con el número de PDB 1NSJ. Previamente se han descrito los primeros modos normales como los más significativos en la descripción de la movilidad de las proteínas. Estos se han utilizado para capturar la movilidad y direccionalidad representativa de la tapa de otros barriles <sup>(73)</sup>, por lo cual, decidimos analizar los 3

primero modos normales en el servidor Promode (<http://cube.socs.waseda.ac.jp/pages/jsp/index.jsp>). Este nos arrojó, con ayuda del algoritmo de dyndom, los resultados que se muestran en la siguiente figura.



**Fig.15-** Arriba se observan los 3 primeros modos normales. Las barras en colores describen los dominios móviles dentro de la proteína y en cada segmentación de las barras se encuentran fijados cuadrillos que representan las bisagras de los dominios. Las flechas indican los sitios acoplados presentes en las bisagras predichas.

La flecha 1 señala las bisagras donde se ubican los sitios acoplados 58, 60, 61 62, del asa  $\beta\alpha_3$ , la flecha 2 del sitio acoplado 124 y el conservado 126 en conjunto con los sitios 139, 140 y 141 presentes en la flecha 3 del asa  $\beta\alpha_6$ . Estos son de gran relevancia para la movilidad de la tapa del sitio activo. A continuación se observa en esferas verdes los sitios pertenecientes a las bisagras de la tapa, figura 16.



**Fig.16-** Estructura tridimensional de E. coli PRAI. Específicamente en esferas de color verde se observan los sitios que conforman la bisagra de la tapa y en azul los sitios altamente conservados del asa flexible.

Las comparaciones de las estructuras con y sin ligando de la ribulosa-1,5-bifosfatasa carboxilasa/oxigenasa y la triosafosfato isomerasa, demuestran que el asa  $\beta\alpha 6$  o tapa representada en la figura 16, se estabiliza en su estado cerrado con la unión de sus substratos y análogos <sup>(68)</sup>.

Existen dos visiones principalmente que explican el origen en la movilidad de la tapa. Una de ellas es la visión clásica de alosterismo, donde los cambios conformacionales ocurren entre estructuras estáticas, las cuales son inducidas al movimiento por la unión del ligando. Los cambios alostéricos de este tipo pueden ser considerados como asociados a ligando o movilidad extrínseca. En contraste, el entendimiento actual del alosterismo, asume que las estructuras tridimensionales de proteínas no son estáticas, sino un conjunto de conformaciones donde el ligando cambia el balance preexistente de sub-estados. Desde esta perspectiva, la proteína está constantemente ínter-convirtiéndose en múltiples conformaciones y el cambio conformacional puede ser considerado intrínseco <sup>(23)</sup>.

Recientemente Kern Dorothee y asociados, han encontrado mediante información experimental y simulaciones, que el proceso dinámico para la apertura de la tapa de unión a nucleótido de la proteína adenilato cinasa se da en gran medida debido a su movilidad intrínseca y es un paso limitante en la catálisis de la enzima adenilato cinasa tanto mesófila como termófila, de acuerdo a medidas de relajación en NMR y dinámicas moleculares donde también demuestran notablemente que la tasa de apertura de la tapa coincide cuantitativamente con el número de recambio en el estado estacionario de la actividad de la enzima <sup>(74)</sup>.

En un segundo artículo demuestran que las diferencias en la actividad y por lo tanto en las tasas de apertura y cerradura de la tapa, se deben a regiones denominadas bisagras relacionadas con fluctuaciones locales, también deducen que la movilidad de esta tapa se da en una escala lenta de microsegundos-milisegundos, debido al requerimiento de movilidad colectiva de muchos átomos dentro de cada bisagra y al mismo tiempo de la correlación de movimiento de las

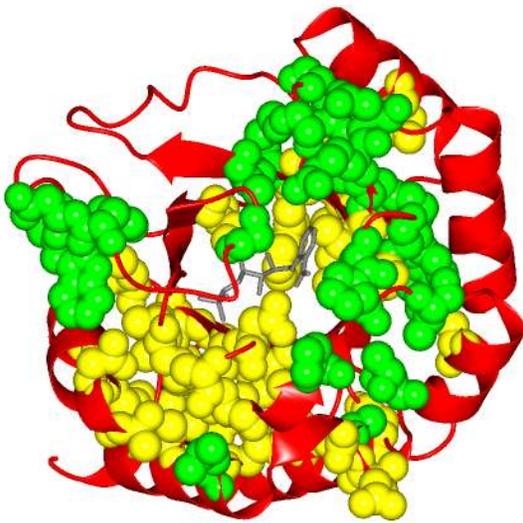
bisagras entre sí <sup>(24)</sup>. Este fenómeno, también se ha observado en la familia de barril de la triosafosfato isomerasa, donde se ha determinado que el movimiento del asa flexible  $\beta\alpha6$  perteneciente a la tapa, se da mayormente de manera intrínseca y no por la unión del ligando. De igual forma, varios estudios realizados en NMR en estado sólido y líquido demuestran que la química de la reacción en dirección dihidroxiacetona fosfato a D-glicer-aldehído-3-fosfato (GaP) puede estar limitada parcialmente por la movilidad de la tapa <sup>(75-77)</sup>. Posteriormente discutiremos a profundidad los resultados obtenidos de esta familia de proteínas.

Estos trabajos en conjunto sugieren que residuos dístales al sitio activo, involucrados en la dinámica de la proteína, pueden tener gran repercusión en la actividad enzimática, adicionalmente, algunos estudios recientes en base información experimental y computacional de enzimas no relacionadas han revelado mutaciones de residuos dístales (residuos no implicados directamente en la catálisis) tener efectos en la actividad catalítica, sin tener interacción alguna con los sustratos. Algunos ejemplos de estas familias son: la Dihidrofolato reductasa (DHFR), la Triosafosfato isomerasa (TIM), la alcohol dihidrogenasa (LADH), la DNA metil transferasa HhaI y la Beta-lactamasa (TEM-1), las cuales no están estructural y funcionalmente relacionadas <sup>(27, 29)</sup>. Específicamente, las enzimas dihidrofolato reductasa y alcohol deshidrogenasa las cuales catalizan transferencia de hidruro, presentan evidencia de redes de movilidad acoplada dentro del plegamiento de la proteína las cuales facilitan la catálisis <sup>(15, 28)</sup>.

Los datos obtenidos de distintas familias estructurales en conjunto con los resultados de nuestro análisis particularmente del grupo GBP, sugieren que muy probablemente los sitios que están involucrados en la formación del canal y las bisagras de la tapa, han estado co-evolucionado para facilitar la entrada del sustrato y salida del producto en el sitio activo y a la vez proporcionar un mecanismo de exclusión del solvente y adecuación de la geometría del sitio activo proveyendo el ambiente adecuado para la catálisis.

### 4.7.3-Sitios acoplados de gran relevancia en la dinámica de plegamiento y estabilidad del barril.

En la tabla 1, se observan el grupo GAP y GBP con los valores más significativo de acoplamiento, puede notarse que la mayoría de los sitios pertenecientes se encuentran ubicados en mayor parte en las cadenas  $\beta$  y en menor medida en  $\alpha$ -hélices, figura 17.

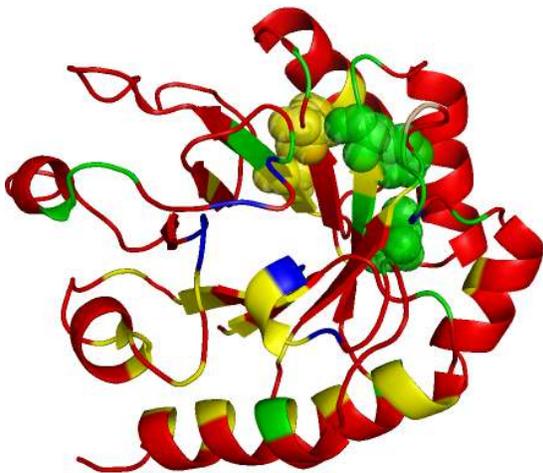


**Fig.17-** Estructura tridimensional de E. coli PRAI. En verde se representa el grupo G2P, amarillo el grupo G1P. De manera global se pueden observar núcleos de plegamiento entre las  $\alpha$  hélices y cadenas  $\beta$  además del núcleo central del barril.

Estudios tanto de dinámica molecular como termodinámicos, han calculado las estabilidades de los distintos núcleos formados por las estructuras secundarias, generando una idea de la ruta de plegamiento seguida por el barril. La mayoría de los datos recopilados acerca del barril TIM, provenientes de análisis cristalográficos, revelan que el barril es provisto de núcleos formados por residuos entre las hélices  $\alpha$  y cadenas  $\beta$ , así también como de un núcleo central formado por cadenas laterales hidrofóbicas que apuntan al interior exclusivos de cadenas  $\beta$ , demostrando que los núcleos del plegamiento del barril TIM son dominados por un gran número de cadenas laterales hidrofóbicas presentes en estos residuos <sup>(57)</sup>.

La dinámica molecular realizada por el grupo de Akihiko Yamagishi de la Universidad de Farmacia y Ciencia de la vida de Tokio, consistió en investigar la ruta de naturalización/desnaturalización del barril TIM, mediante una simulación de desnaturalización inducida por calor para ePRAI. La simulación fue hecha a 5 diferentes temperaturas y muestran una ruta secuencial de desnaturalización constituida de varios eventos. Los eventos tempranos de desnaturalización, involucran la ruptura de las últimas dos cadenas  $\beta$  y 3 hélices, mostrando intermediarios similares a los detectados experimentalmente. Posteriormente, la desnaturalización de las dos primeras unidades  $\beta\alpha$  y la separación de la sexta cadena  $\beta$  de la quinta, se da de manera independiente. El módulo central restante  $(\beta\alpha)_{3-4}-\beta_5$  es el más persistente durante todas las simulaciones sugiriendo un papel muy importante para el plegamiento global del barril <sup>(30)</sup>.

Estas simulaciones también predicen la presencia de un sitio de nucleación, en el cual varios residuos hidrofóbicos se condensan, iniciando la formación del módulo  $(\beta\alpha)_{3-4}-\beta_5$ . Estos sitios corresponde de acuerdo a nuestro MSA a los siguientes: 60, 77, 82 y 100, tabla 1. Los sitios 82 y 100 en GAP, con altos valores de acoplamiento, por otro lado el 60 y 77 se encuentran en el grupo GBP también con valores significativos de acoplamiento. Los cuatro sitios se muestran en esferas en la figura 18.



**Fig.18-** Estructura tridimensional de E. coli PRAI. En esferas se muestran los sitios acoplados descritos como de gran relevancia en la dinámica de plegamiento de la proteína. En color verde corresponden a sitios del grupo G2P y en amarillo al G1P.

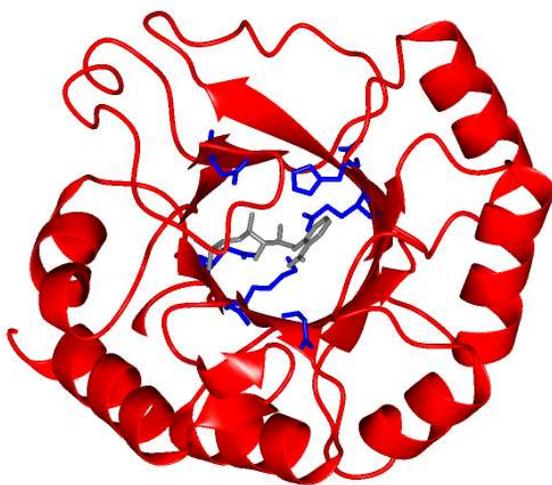
Los datos de acoplamiento presentan una clara congruencia con los antecedentes reportados de dinámica molecular, en cuanto a los sitios predichos para la nucleación del módulo  $(\beta\alpha)_{3-4}-\beta_5$ . A su vez, los antecedentes de dinámica molecular son apoyados por varios análisis estructurales de proteínas permutadas circularmente, donde muestran que si ePRAI es alterada con una incisión en su N terminal correspondiente a los módulos  $(\beta\alpha)_{1-2}$ , la cantidad de estructuras secundarias no es afectada, mientras que si la inserción ocurre en el módulo  $(\beta\alpha)_{3-4}-\beta_5$ , la proteína simplemente no se pliega <sup>(78)</sup>.

Los autores que han realizado estos estudios tanto computacionales como experimentales, han sugerido que la formación de las tres cadenas del barril generadas en el módulo  $(\beta\alpha)_{3-4}-\beta_5$ , puede ser uno de los importantes intermediarios para generar el plegamiento. También discuten que las interacciones encontradas en el fragmento  $(\beta\alpha)_{3-4}-\beta_5$  tal vez sean esenciales para los eventos tempranos en el plegamiento de PRAI. Otros estudios de fragmentación del barril han demostrado que el módulo  $(\beta\alpha)_{1-4}-\beta_5$  mantiene plegamiento autónomo y casi tan estable como el nativo. La adición covalente de la unidad  $(\beta\alpha)_6$  a  $(\beta\alpha)_{1-4}-\beta_5$  de PRAI ayuda a generar un incremento significativo en la estabilidad de la proteína <sup>(79)</sup>.

Los resultados previos que revelan que el módulo  $(\beta\alpha)_{3-4}-\beta_5$  es uno de los intermediarios importantes para generar y dirigir el proceso de plegamiento del barril TIM en conjunto con los datos de acoplamiento aquí mostrados, abren posibles blancos para posteriores diseños del barril TIM.

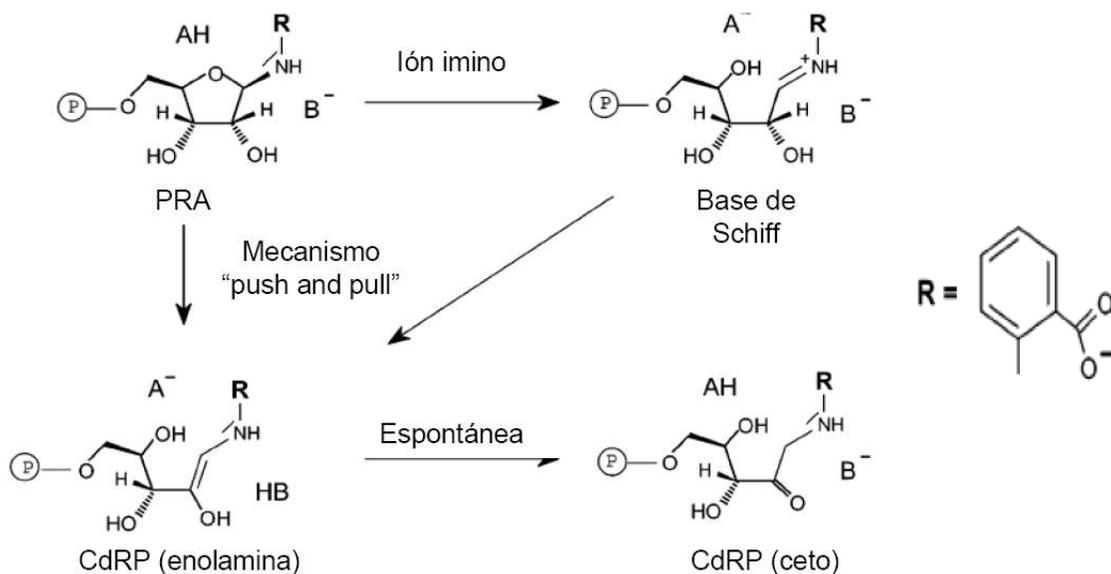
#### 4.7.4-GRUPO CON (altamente conservados).

Las cadenas laterales de los siguientes aminoácidos invariantes presentes en CON están orientados hacia la cavidad del sitio activo: 5, 7, 81, 83, 126 y 178, figura 19. El grupo amino del sitio 5 tiene un enlace de hidrógeno con el sitio 178 y 181. Debido a lo anterior se logra un empaquetamiento ajustado de los residuos hidrofílicos en el c-terminal del barril central  $\beta$  la cual funciona como fondo de la cavidad del sitio activo en PRAI. Las cadenas laterales de los demás sitios no mantiene enlaces de hidrógeno hacia otros grupos de la proteína <sup>(68)</sup>.



**Fig.19-** Estructura tridimensional de ePRAI. En azul se pueden observar los sitios altamente conservados que funcionan como cavidad del sitio activo en PRAI, en gris se observa el sustrato.

Las mutaciones y estructuras cristalográficas estudiadas por diversos grupos han sido útiles para entender la función de algunos de los residuos altamente conservados presentes en el sitio activo. En el mecanismo deducido de reacción de PRAI, un ácido general (AH) protona el átomo de oxígeno del anillo de furanosa del sustrato, iniciando el corte del enlace O-C1' y la formación del intermediario base de schiff. La base general (B-) posteriormente abstrae el protón del átomo C2' de la ribosa, generando la forma enolamina del producto, el cual espontáneamente isomeriza en la forma ceto <sup>(59, 80)</sup>.



Modificado de Henn y Sax (2002)

**Fig.20-** Posible mecanismo de reacción catalizado por PRAI. El ácido (AH), la base (B-) y el sustrato PRAI (R).

Tres enzimas mutantes C260V-7, E332V-81 y D379-126 (las numeraciones son en base a ePRAI-tPRAI), tienen valores de número de recambio ( $K_{cat}$ ) menores al 1% y constante de *Michaelis* ( $K_m$ ) del mismo orden con respecto a la enzima silvestre de *E. coli*, sugiriendo que estos sitios están involucrados en la catálisis de PRAI. La estructura cristalográfica con el ligando r-1-(*o*-carboxifenilamino)-l-deoxirribulosa 5-fosfato (rCdRP) sugiere que la cisteína 7, la cual está localizada al lado C-terminal de la cadena  $\beta_1$ , actúa como base general, y el aspartato 126, el cual está localizado en el C-terminal de la cadena  $\beta_6$  como ácido general<sup>(68)</sup>.

#### **4.8-Análisis de una segunda familia ortóloga de proteínas, Triosafosfato Fosfato Isomerasa (TIM) con plegamiento de barril TIM ( $\beta/\alpha$ )<sub>8</sub>.**

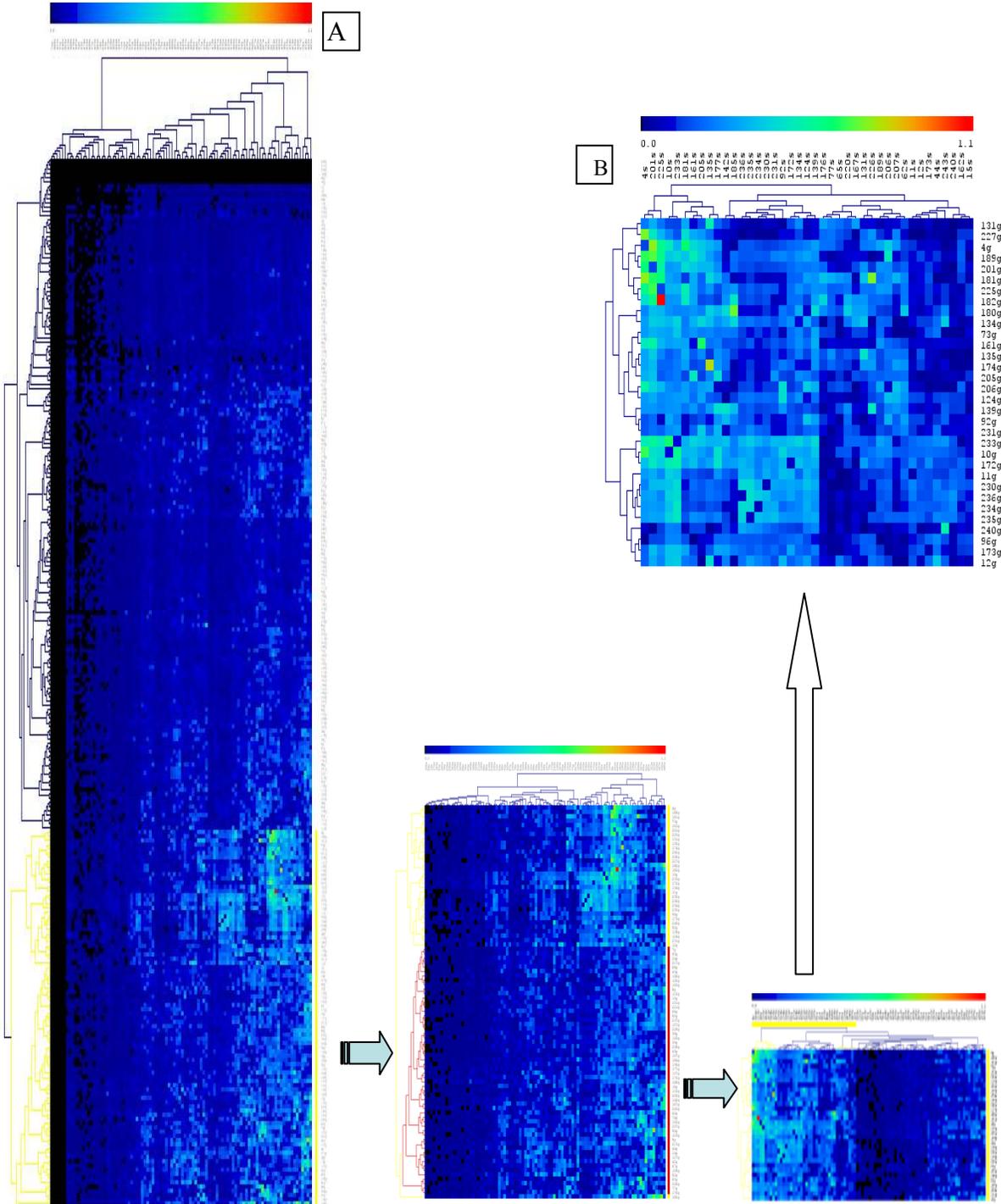
Con la intención de hacer una comparación, sobre la co-evolución de otra proteína con plegamiento de barril TIM ( $\beta/\alpha$ )<sub>8</sub>, decidimos enfocarnos a la familia de ortólogas de la triosafosfato isomerasa (TIM). Esta proteína participa en la ruta glucolítica donde cataliza la interconversión de D-glicer-aldehído-3-fosfato (GaP) y dihidroxiacetona fosfato <sup>(81)</sup>. Por otro lado, existe mucha información experimental que se puede contrastar y complementar con nuestro análisis, ya que miembros de la misma familia han sido extensamente estudiados. Así mismo, mutaciones en la enzima, han sido relacionadas con un gran número de desordenes genéticos de gran interés médico <sup>(82)</sup>.

El barril de la familia funcional TIM también ha sido seleccionado como blanco importante para el diseño de drogas en contra de parásitos que tienen la habilidad de sobrevivir en el torrente sanguíneo de mamíferos, ya que el azúcar es la principal fuente de energía de los mismos y es de esperarse que el bloqueo de la glucólisis conduzca a la eliminación de estos <sup>(83)</sup>. El conocimiento detallado de los mecanismos de interacciones de residuos, dentro de la TIM, facilitará el reconocimiento de sitios potenciales para el bloqueo en el diseño de drogas y en consecuencia la pérdida de funcionalidad de la proteína.

Por los intereses mencionados anteriormente, creemos que el análisis de acoplamiento estadístico, en conjunto con bibliografía sobre datos experimentales previos sobre la familia TIM, nos ayudará a entender en un contexto global el mecanismo de interacciones funcionales necesarias para actividad y estabilidad de la proteína así como la relación de la afectación de estas interacciones debido a mutaciones conocidas y ampliamente descritas en desordenes genéticos.

El proceso para análisis de acoplamiento estadístico, se llevo de manera idéntica a PRAI, el cual fue anteriormente descrito y en esta ocasión omito para ir

directamente al proceso de agrupamiento de la matriz global de acoplamiento, la cual se genero a partir de 298 secuencias con diversidad suficiente para el análisis. El agrupamiento se observa a continuación en la figura 21.



**Fig.21-** Proceso de agrupamientos de sitios acoplados de la familia de ortólogos TIM. En la figura 17B se observa el grupo más sobresaliente con los sitios altamente acoplados.

En la figura 21B podemos observar el grupo más sobresaliente de la familia TIM. Los sitios presentes en este grupo mantienen altos valores de acoplamiento estando aproximadamente dentro del 5 por ciento de los valores más altos en la matriz global de co-evolución. La confiabilidad de estos resultados es alta, ya que muchos sitios pertenecientes a este grupo, han sido descritos anteriormente en la bibliografía como de gran relevancia para la proteína.

A continuación, describo el papel de estos sitios mencionados en trabajos previos, para finalmente ubicarlos en un contexto global. Para una mejor comprensión de la relación existente entre los sitios presentes del grupo de sitios altamente acoplados, los dividiremos en cuatro grupos G1, G2, G3, G4. De manera adicional agregaremos a estos grupos los sitios altamente conservados mostrados en rojo en la tabla 2. Los grupos con sus sitios pertenecientes son:

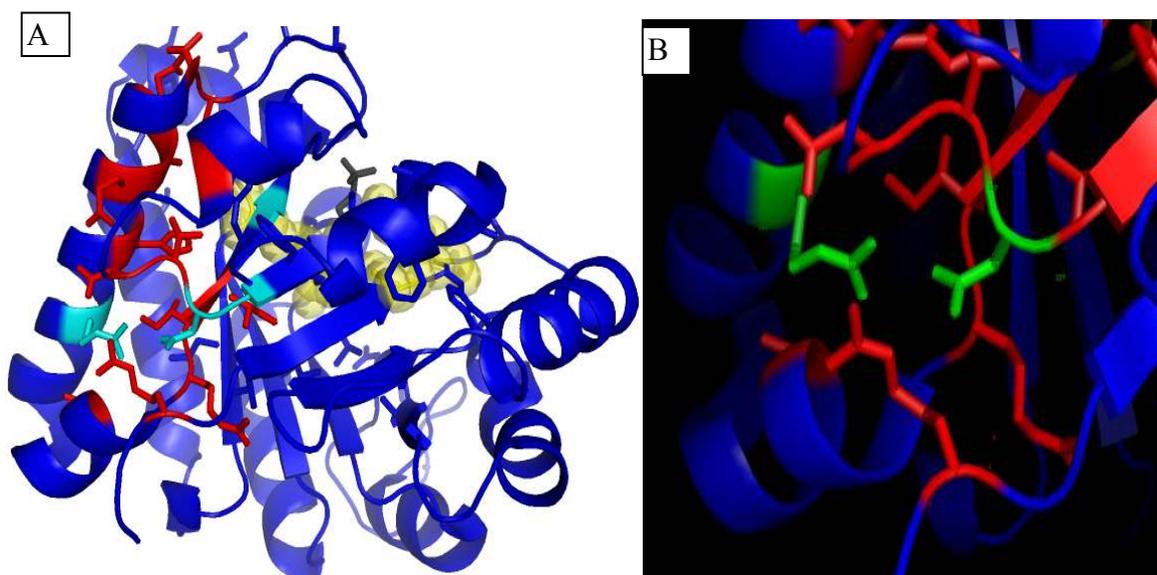
**Tabla 2-** Grupo de sitios altamente acoplados y altamente conservados (en rojo).

GRUPO	SITIOS INTEGRANTES	TOTAL
G1	10, 11, 12, 15, 44, 62, 230, 231, 233, 234, 235, 236, 240, 243	14
G2	4, 180, 181, 182, 185, 189, 201, 205, 206, 207, 220, 225, 226, 227, 176, 177.	16
G3	92, 96, 124, 131, 134, 135, 139, 142, 161, 162, 167, 172, 173, 174	14
G4	65, 73, 77	3
CON	13, 232, 208, 237, 209, 210, 211, 94, 95, 97, 98, 99, 126, 128, 129, 146, 163, 164, 165, 166, 168, 169, 170, 171, 64, 72, 75, 76, 104, 112, 228, 229	32

### 4.8.1-Grupo G2

Uno de los trabajos que soporta fuertemente los resultados de nuestro análisis, es el de ingeniería inversa realizado por Silverman <sup>(84)</sup>, el cual es considerado por algunos autores como una disección maestra del barril TIM  $(\beta/\alpha)_8$  con grandes implicaciones en el diseño de proteínas <sup>(85)</sup>.

En este trabajo, los sitios acoplados 189 y 227 en conjunto con los sitios altamente conservados por glicinas 209 y 228 presentes en G2, son considerados como inmutables ya que el cambio en estos sitios genera la incapacidad de complementar cepas mutantes de TIM *in vivo*. Es importante notar en la estructura tridimensional de TIM que todos los sitios presentes en G2 se encuentran en el contexto ubicado alrededor de estos 4 sitios inmutables <sup>(84)</sup>.



**Fig. 22-** Estructura tridimensional de TIM de humano. En rojo se observa el grupo G2 en azul cian los sitios 189, 227, 209, 228 y en esferas amarillo los sitios catalíticos presentes en la figura 22A, mientras que en la figura 22B en verde se encuentra el puente salino conformado por los residuos 189 y 227.

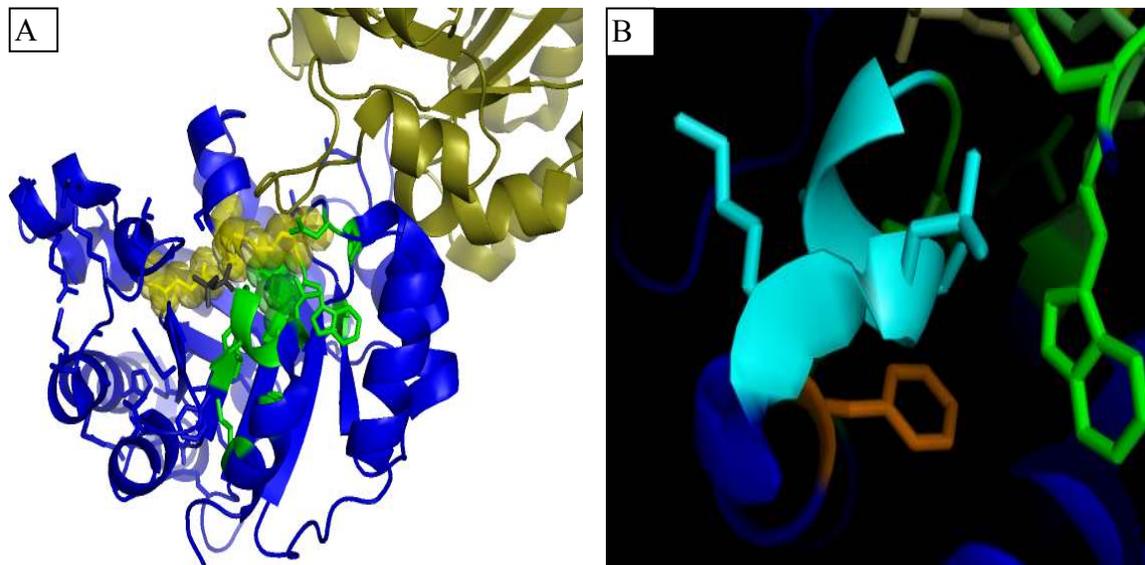
Los residuos 189 y 227 presentes en verde en la figura 22B, parecen formar un puente salino y están localizado en el lado opuesto del sitio activo del barril muy alejados del solvente. Esto no parece ser una casualidad ya que es sabido que

cadena lateral polares en regiones hidrofóbicas pueden jugar un papel de organización a expensas de un rápido plegamiento y estabilidad de la proteína <sup>(85)</sup>.

Mutaciones en la red de sitios cercanos alrededor a este puente salino, tiene un gran efecto perjudicial en la función de la proteína ya que muestran un considerable decremento en la constante de velocidad aparente; algunos de estos sitios también se encuentran presentes en G2 (205, 220, 225, 185, 201, 182) <sup>(84)</sup>.

#### 4.8.2-Grupo G1

El grupo de 17 sitios presentes en G1 se encuentran formando principalmente la hélice-3<sub>10</sub> (233-237) que contribuye sustancialmente a la unión a fosfato. Esta hélice apunta directamente al grupo fosfato del sustrato <sup>(86)</sup>, la cual puede observarse en azul cian dentro de la figura 23B, mientras que en la figura 23A en verde se muestran todos los sitios pertenecientes a G1.



**Fig.23-** En verde se observa el grupo G1 figura 23A. En azul cian correspondiente a la figura 23B se observa la hélice 3-10 de unión a fosfato y debajo de esta, en naranja el sitio perteneciente a la Phe240.

Otros sitios importantes en el grupo son el 240, el 11 en conjunto con el altamente conservado 13, particularmente este último ha demostrado ser esencial

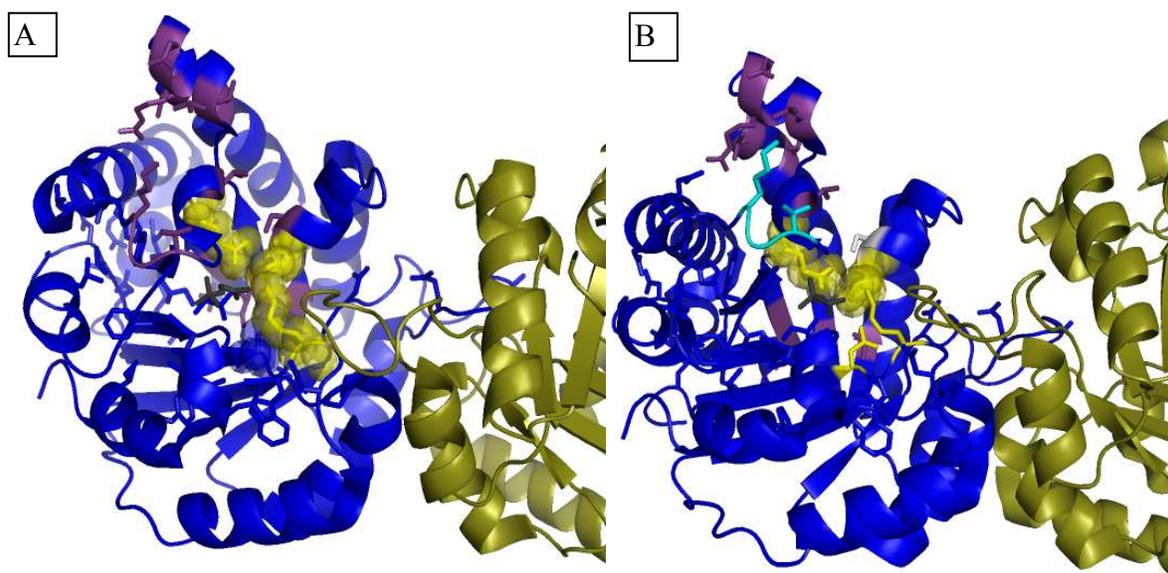
para la unión y catálisis ya que forma un enlace de hidrógeno directamente con el átomo oxígeno2 (O2) del ligando, estabilizando el desarrollo de la carga negativa en el mismo, mientras que el sitio 11 también proveería un potencial electrostático positivo, estabilizando una reacción intermedia o estado de transición además de un enlace de hidrógeno con el átomo oxígeno1 (O1) del ligando <sup>(87)</sup>.

El sitio 240 que se observa en la figura 23B no forma parte directamente del sitio activo pero se encuentra muy próximo. Es mencionado como factor crítico para el mantenimiento de la geometría del sitio activo, ya que forma varios contactos en el rango de 3.4 a 4 Å<sup>0</sup> con residuos de la cadena β1, así como de la hélice 3-10. La hélice 3-10 esta involucrada en la unión al grupo fosfato del sustrato <sup>(86)</sup>, por lo que cualquier alteración en su dirección, tendrá efectos substanciales en la unión al ligando y subsecuentemente en actividad catalítica. Los sitios acoplados 11 y 12 son seguidos de sitios que se encuentran en contacto con la cadena lateral del sitio 240. Debido a que el sitio 240 mantiene contactos con regiones de unión al ligando, es probable que este en la presencia de el ligando propague la señal a través de los sitios 11 y 12 de la cadena β1, para finalmente adecuar la geometría catalítica del sitio 13. Es importante señalar que mutaciones en humano en el sitio 240, en conjunto con mutaciones en otros sitios como el 189 y 104, son relacionados frecuentemente por diferentes autores con el desarrollo de anemia hemolítica y diversas neuropatías <sup>(82, 88)</sup>.

Finalmente, la co-evolución de sitios que conforman la hélice de unión a fosfato en la triosafosfato isomerasa también se presenta en otras familias de barriles, como PRAI e IGPS. Lo anterior sugiere que la co-evolución de esta región de unión a sustrato puede ser un patrón recurrente en el plegamiento de barril TIM.

### 4.8.3-Grupo G3

El grupo denominado G3 puede ser observado en la figura 24. Está conformado por 14 sitios, que en conjunto con sitios altamente conservados, constituyen el asa flexible  $\beta\alpha_6$  de la proteína (168-174) <sup>(86)</sup> y algunas estructuras secundarias formadas alrededor de esta asa. También de manera adicional se muestran sitios con carácter hidrofóbico en el interior de barril de frente al sitio activo.



**Fig.24-** En violeta dentro de la figura 24A se observa el grupo G3, en gris el sustrato y en esferas amarillos los sitios catalíticos 165, 95 y 13. En la figura 24B se puede observar el asa flexible de color azul cian y finalmente en blanco el sitio 96.

Particularmente existe un alto grado de conservación de los sitios 164, 165, 166, 168, 169, 170 y 171 del asa flexible. Nuestro análisis identificó los siguientes sitios acoplados 167, 172, 173 ubicados en medio del asa y dentro de la región denominada bisagra-carboxilo de la tapa a los sitios 174, 176, 177 <sup>(70)</sup>.

Como se comentó en la sección del análisis de la familia del barril PRAI, varios estudios realizados en NMR en estado sólido y líquido en la familia funcional TIM, sugieren que la química de la reacción en dirección dihidroxiacetona fosfato a D-glicer-aldehído-3-fosfato (GaP), es parcialmente limitada por la tasa de apertura del

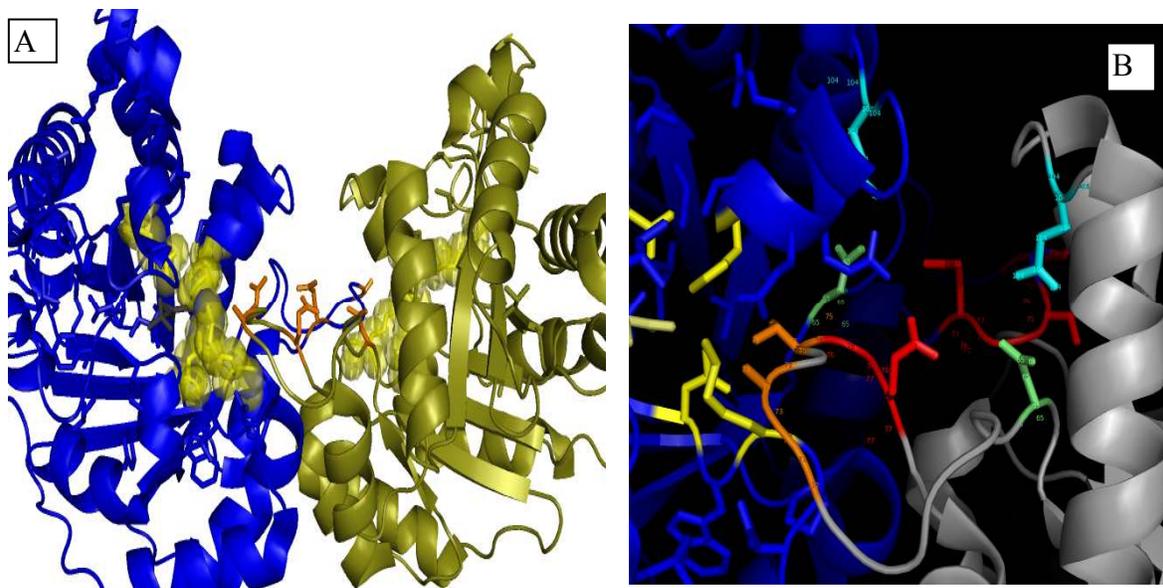
asa flexible  $\beta\alpha 6$ . Esto sugiere que en TIM al igual que en PRAI, los sitios en co-evolución dentro de la tapa, tiene un rol importante en su movilidad y por lo tanto en la velocidad de catálisis. Además de PRAI y TIM, en al menos otras 3 enzimas de de la familia estructural de barril TIM (Triptófano sintasa, ribulosa-1,5-bifosfatasa carboxilasa/oxigenasa, e indol glicerol fosfato sintasa) el asa  $\beta\alpha 6$  esta involucrada en diversos mecanismos, posiblemente jugando un papel preponderante en la exclusión del solvente en el sitio activo, manteniendo una posición competente del sustrato para la catálisis y realizando de manera efectiva la liberación del producto (68, 69).

Existen dos sitios catalíticos altamente conservados relacionados con el grupo G3, el 95 y 165, mostrados en gris en la figura 19. El primero de ellos tiene enlaces hidrógeno a ambos O1 y O2 del ligando y su rol en el transporte de protones aun no está totalmente entendido. El segundo es descrito como una base catalítica para la abstracción del protón del sustrato (87). La relación de estos sitios se da debido a que la cadena lateral del sitio 165 en la apo estructura del barril mantiene un enlace de hidrógeno con el sitio acoplado 96 presente en G3, mostrado en blanco en la figura 24B (*swing out*, lejos del sitio de unión a ligando, pero seguido del sitio 95). En la presencia de ligando (*swing in*, esta cadena lateral rota hacia el ligando interactuando directamente con el, lo cual puede verse en varias estructuras de complejos TIM-ligando (89), incluyendo la de humano). Las diferencias entre los estados *swing in* y *swing out* se muestra en la figura 28.

#### 4.8.4-Grupo G4

G4 está compuesto de 3 sitios el 65, 73 y 77 los cuales están ubicados principalmente en la interfaz del homodímero. Es sabido que la sustitución de aspartato por glutamato en el sitio altamente conservado 104 resulta en la pérdida de actividad debido a la inestabilidad de la enzima. De igual forma la mutaciones del sitio altamente conservado 98, el cual se encuentra localizado en la interfaz del dímero y cercano en la estructura al 104, también muestra una gran disminución en

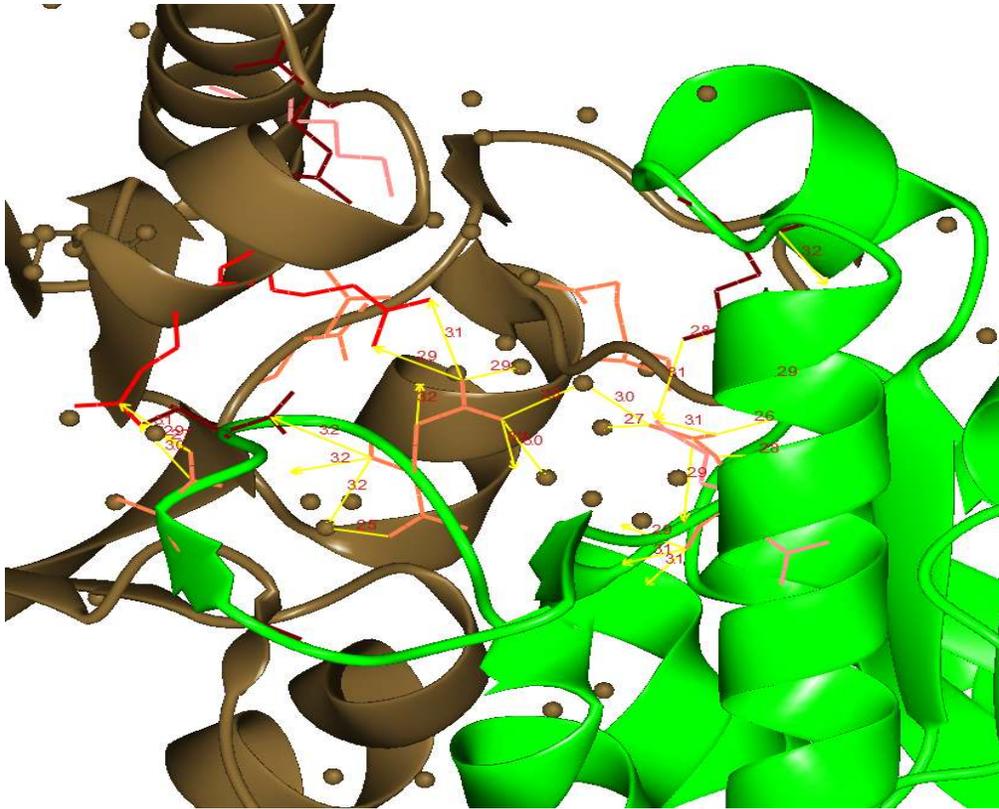
la actividad <sup>(82)</sup>. Este sitio 98 forma un puente salino intersubunidades con el sitio acoplado 77, lo cual es de gran relevancia para la formación de la interfaz.



**Fig.25-** En naranja de la figura 25A se observa el grupo G4 conteniendo los sitios 65, 73, 77. En la figura 25B se puede observar los sitios altamente conservado 72 y acoplado 73 en anaranjado alrededor del sitio activo. En cada uno de los monómeros los sitios 65 en azul cian y 104 en verde limón, se encuentran muy próximos sugiriendo algún tipo de interacción, finalmente el sitio 77 se aprecia en rojo.

El cambio en el sitio 65 de Lis-Asp genera la pérdida completa de actividad <sup>(82)</sup>. Es relevante mencionar que este sitio, al igual que el sitio altamente conservado 98, se encuentra en la interfaz del dímero con contactos medios por agua y el altamente conservado Lis112 hacia el sitio 104. Los sitios 72, 73 y 75 de uno de los monómeros (barril gris), están cercanos a la región del sitio activo del segundo monómero (barril azul) sugiriendo que los mismos le proveen un ambiente necesario para la catálisis (Figura 25B). Específicamente el 72 y el 73 se encuentran próximos al sitio catalítico 13 y el sitio 75 al sitio catalítico 11 respectivamente.

Interacciones de gran relevancia como las comentadas anteriormente, se observan en las diferentes estructuras del homodímero TIM, donde miembros del grupo G4 (en palitos coral) en cada uno de los monómeros interactúa con sitios conservados (en rojo) del monómero activo (monómero café), figura 26.

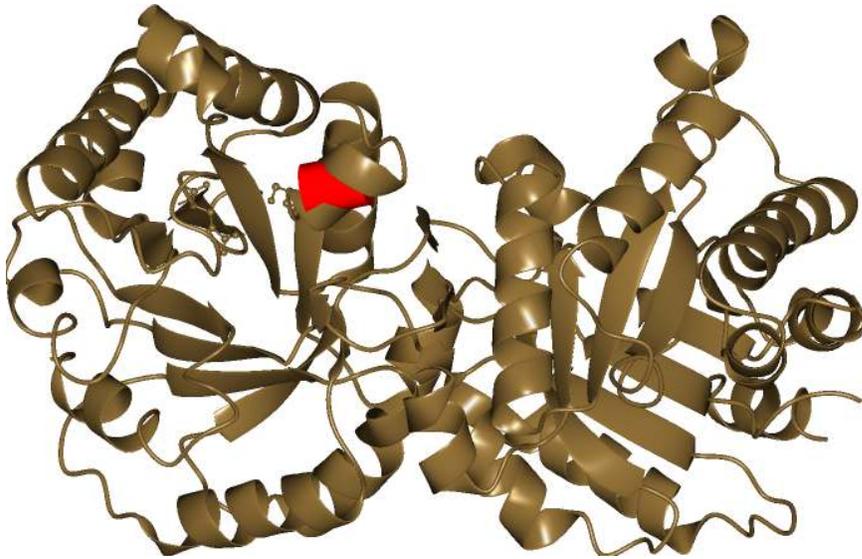


**Fig.26-** En rojo los sitios altamente conservados 98, 97, en rosa y marrón los sitios 112 y 104 respectivamente. Los sitios acoplados 73, 77, 65 se encuentran en coral en cada uno de los monómeros. Específicamente los contactos se representan en vectores amarillos para los sitios acoplados del monómero en verde principalmente con los sitios conservados en rojo del segundo monómero.

El grupo G4 del segundo monómero, a través de sus sitios 73 y 77, mantiene un número considerable de contactos cercanos a G3 del monómero activo, particularmente con los sitios altamente conservados 97 y 98 miembros de la hélice que va del sitio 95 al 102 denominada  $D_1$  (Figura 27).  $D_1$  uno de los factores que la enzima mantiene rígidamente fijos para la polarización de los oxígenos carbonilo del sustrato. Esto justifica la alta conservación de los sitios que conforman la hélice  $D_1$  dentro de la familia.

Los sitios que conforman la hélice comentada anteriormente, especialmente los conservados His95 y el sitio acoplado Lis12, generan un momento de macró-dipolo que induce un campo eléctrico positivo en el extremo terminal-amino. Esta

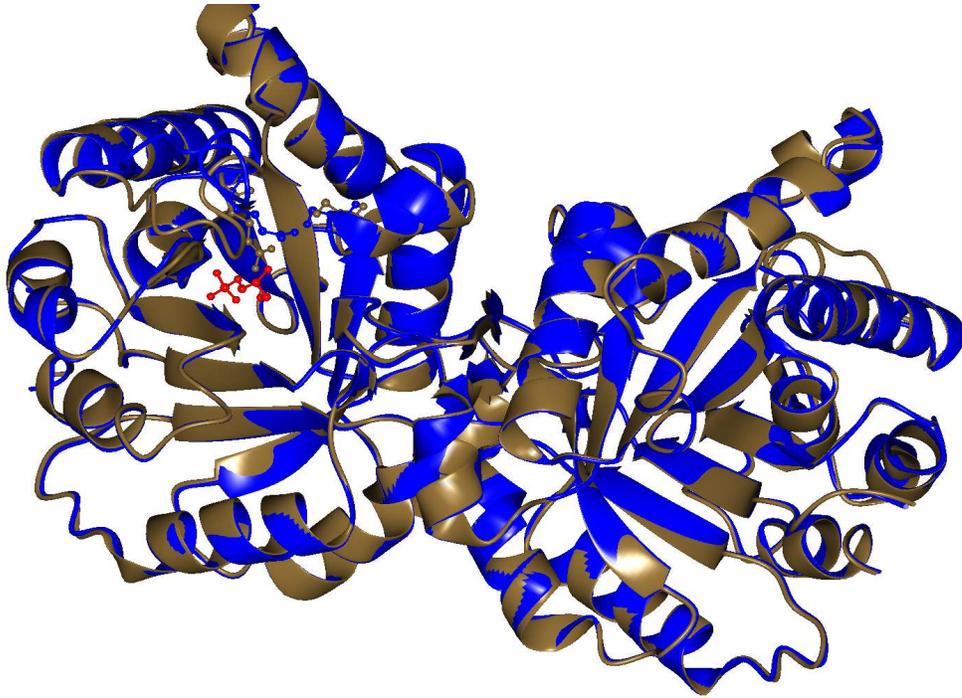
polarización es importante para debilitar los enlaces C-H del sustrato y facilitar la transferencia del protón de un carbón a otro <sup>(90)</sup>.



**Fig.27-** En cilindro rojo se representa la hélice que va del sitio 95 a 102 denominada D<sub>1</sub>.

Los estudios de NMR en solución de la TIM de levadura, demuestran que las perturbaciones más relevantes de intercambio químico son originadas mediante la unión del ligando, sobresaliendo la hélice de unión a fosfato y los sitios 95 y 100. Ambos sitios pertenecen a la hélice D<sub>1</sub> que se encuentra ubicada en el extremo terminal del barril formando parte del sitio activo <sup>(89)</sup>. El sitio 95 interactúa directamente con el extremo ribosa del ligando, lo cual explica su perturbación.

El elevado intercambio químico en estas regiones está asociada con su mayor movilidad, específicamente la unión del ligando provoca una movilidad en la hélice D<sub>1</sub>, donde se ubica el residuo 96 el cual mantiene un rol importante en la interconversión del estado activo (*swing in*) y el estado inactivo (*swing out*) de la proteína, figura 28. Estos resultados experimentales sugieren que a pesar de que se ha determinado que el movimiento de la tapa ocurre de manera intrínseca, también la unión del ligando contribuye sustancialmente en la adecuación de la geometría del sitio activo, lo cual encaja con el modelo de ajuste inducido.



**Fig.28**-Sobreposición de estructuras TIM (hapo y halo) de tripanosma brucei. En café se observa el dímero en estado halo (swing in), y en azul (swing out) en estado apo. En bolitas y palitos se observa el sitio catalítico 165 y 96 mencionados como proveedores del enlace de hidrógeno en el estado apo de la enzima, finalmente en rojo se observa el sustrato.

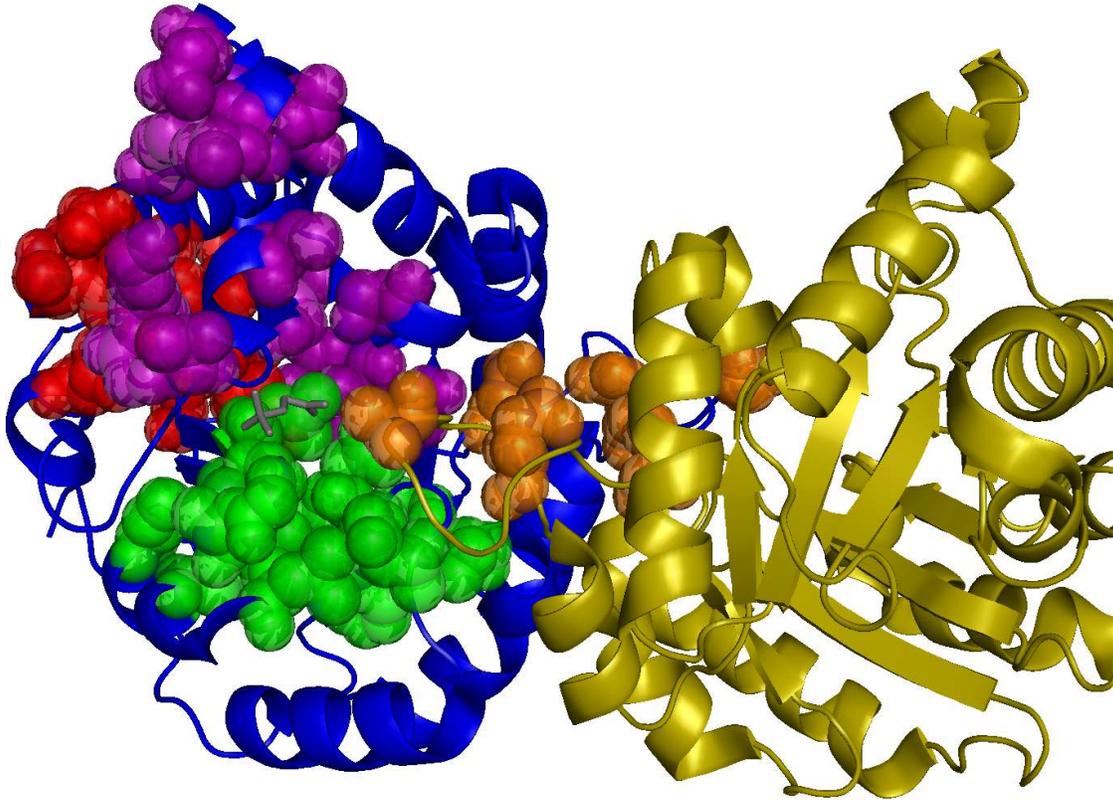
Los hechos mencionado anteriormente describe probablemente la relevancia de la forma homodimérica para la actividad y la explicación de la supresión de la actividad en su forma monómerica. Por otro lado, se observan sitios altamente conservados en el núcleo del barril los cuales pueden ser de gran relevancia para el colapso hidrofóbico de la proteína, figura 29.

#### **4.8.5- Red de sitios acoplados y conservados**

Como mencionamos en la sección del análisis de PRAI, existe un gran cuerpo de evidencia experimental, que describe residuos fuera del sitio activo que intervienen en la catálisis. Estos sitios posiblemente estén involucrados en dirigir la dinámica intrínseca y extrínseca del plegamiento necesaria para la catálisis.

Los sitios pertenecientes a los 4 grupos forman una red global de interacciones de (G1-G4), ya que de manera sugerente los sitios que conforman el grupo de

sitios acoplados de TIM se encuentran circundando el sitio activo, manteniendo un gran número de contactos de vanderwaals entre ellos. Los 4 grupos se observan en la figura 29.



**Fig.29-** En rojo el grupo G2, verde G1, lila G3 y finalmente en anaranjado G4 (este último en los dos monómeros).

En la visión tradicional de movimientos extrínsecos mediados por ligando, algunos autores proponían que posterior a la unión del ligando en la zona de sitios pertenecientes a G1, ocurría un cambio conformacional en el asa flexible perteneciente al grupo G3, donde se ha demostrado que la punta del asa se mueve alrededor de 7Å formando una tapa sobre el ligando enlazado <sup>(27)</sup>. Estudios posteriores han demostrado que tal movimiento de la tapa se da aun en ausencia del ligando <sup>(75)</sup>, apoyando con esto la visión moderna de alosterismo. Tal mecanismo es apoyado por el modelo de intercambio poblacional de conformaciones (en inglés, *Selection/Population Shift Model*). A pesar que el

movimiento intrínseco describe el movimiento global de la tapa, existe evidencia experimental de que la unión del ligando, propicia cambios conformacionales que intervienen directamente en adecuar la geometría del sitio activo apoyando con esto el modelo de ajuste inducido (*Induced Fit Model*, en inglés).

Holyoak y Sullivan proponen en un artículo publicado recientemente, que el mecanismo para llegar al estado activo de llave-cerradura, no puede ser explicado utilizando exclusivamente uno de los dos modelos comentados anteriormente, como en el caso de la trifosfato isomerasa y en otras enzimas donde el complejo llave-cerradura requiere del cerrado de la tapa del sitio activo, concluyendo que estos modelos, más que antagónicos, son complementarios <sup>(91)</sup>. La explicación de esto radica en que en estas enzimas se requiere la exclusión del solvente y sustrato durante la catálisis por lo cual si existe una conformación activa de manera poblacional pequeña de acuerdo al modelo poblacional de conformaciones, esta es la misma que el estado llave-cerradura que la enzima muestrea en solución. Esto significa que el sustrato no podría unirse directamente en este estado debido al impedimento estérico ocasionado por la oclusión del sitio activo. Esta explicación concilia los resultados experimentales que describen tanto movimientos intrínsecos como los originados por ligando en el barril; probablemente algunos sitios que co-evolucionan en conjunto con los conservados contribuyan en este aspecto dinámico de la enzima. Posiblemente, el mecanismo que dirige la movilidad intrínseca del asa flexible  $\beta\alpha 6$  funciona mediante una red de interacciones que se propaga de G1 a G3 conectados necesariamente por G2, es decir del color verde al lila, figura 29, para finalmente fijar el balance de sub-estados en dirección a la conformación en estado cerrado del asa flexible. De igual forma el solvente puede estar jugando un rol crítico en esta movilidad intrínseca como previamente se ha descrito <sup>(15)</sup>.

Básicamente, la conexión de G1 a G2 puede estar ocurriendo de manera intrínseca aunque como mencionamos se podría estar promoviendo con la unión del ligando desde la hélice 3-10 hacia a la cadena  $\beta 8$ , donde se da el cambio de

G1 a G2. El sitio 243 del grupo G1 también interaccionan con algunos sitios presentes en G2. Los sitios integrantes de G2 presentan un núcleo grande de interacciones de vanderwaals entre ellos y se encuentran detrás del sitio activo, principalmente los ubicados entre la  $\beta$  7 (201, 206) y hélice  $\alpha$  6 (180, 181, 182). El sitio 4 ubicado en el comienzo del N-terminal de la proteína también interactúa con esta hélice 6. El asa altamente conservada  $\beta\alpha$ 7 (209, 210, 211) presente en el contexto de G2 sufre un cambio conformacional igual que el asa flexible  $\beta\alpha$ 6 presente en G3 formando con este el pocket hidrofóbico <sup>(86)</sup>. G3 se encuentra conectado a G2, por medio de la  $\alpha$  hélice 6 y  $\alpha$  hélice 5 donde hay una interacción directa del sitio 180 en G2 con el sitio 134 de G3, esta interacción sugiere ser de gran relevancia en la dinámica del asa flexible. Los sitios presentes en G3 del monómero activo podrían estar relacionados con el grupo G4 de el segundo monómero, dirigidos principalmente por la unión del ligando. El ligando, al interactuar con el sitio catalítico 95, provoca una movilidad considerable en la hélice  $D_1$  que va del sitio (95-101) presente en G3, probablemente adecuando la geometría catalítica del sitio activo. Específicamente, promoviendo el *swing in* del residuo 165, ya que el enlace de hidrógeno que mantiene este con el sitio acoplado 96 perteneciente a  $D_1$  puede verse afectado por tal movilidad. Esta hélice a su vez mantiene contactos con sitios presentes en G4 del segundo monómero, los cuales son de gran relevancia, de acuerdo a estudios de mutagénesis que involucran a los sitios pertenecientes a este grupo.

Como mencionamos en la sección anterior, los estudios de NMR y dinámicas moleculares, indican un cambio en la movilidad de los sitios comentados previamente tras la interacción de la enzima con el sustrato y de manera intrínseca. En conjunto estos movimientos originados por ligando, como los intrínsecos que dirigen el movimiento de la tapa, juegan un papel indispensable para la adecuación de la geometría del sitio activo necesaria para la catálisis donde algunos sitios que co-evolucionan tendrían un rol importante.

## 5-CONCLUSIONES:

El análisis realizado a las dos familias funcionales PRAI y TIM, pertenecientes a misma familia estructural de plegamiento barril TIM ( $\beta/\alpha$ )<sub>8</sub>, revela sitios de gran importancia involucrados en los procesos de estabilidad, unión a sustrato y tentativamente en redes que gobiernan la dinámica necesaria para la catálisis. Estos sitios identificados podrían estar co-evolucionando para el mantenimiento del sitio activo y la adecuada relación estabilidad y actividad en las enzimas.

En las dos familias de barriles se encontraron sitios que co-evolucionan en regiones de los dos dominios hidrofóbicos del barril TIM los cuales han sido descritos de manera experimental y computacional como de gran relevancia en el proceso de plegamiento y estabilidad del barril.

Ambos barriles contienen redes de sitios acoplados que se encuentran en la cara catalítica del barril, principalmente en las asas  $\beta\alpha$ , destacando dentro de ellas la  $\beta\alpha 6$  que es denominada como la tapa del sitio activo. Esta tapa de igual forma es representada en diversas familias de proteínas donde se ha descrito tener un papel preponderante en la catálisis. El mecanismo de movilidad de la tapa del asa flexible  $\beta\alpha 6$  en el caso de la familia de PRAI podría ser uno de los pasos limitante en la catálisis. En la familia de la trifosfato isomerasa el mecanismo de catálisis es altamente eficiente. El hecho de que el barril necesite ser homodímero para ser funcional tal vez lleve implícito esta eficiencia del mecanismo. Algunos estudios recientes mencionan que la química de la reacción puede estar también parcialmente limitada por la apertura de la tapa perteneciente al asa flexible  $\beta\alpha 6$ . Adicionalmente, en PRAI se encontró una red de sitios que han co-evolucionado para formar un canal que parece tener un rol en la entrada y salida de sustrato al sitio activo. Esta red parece interactuar con la tapa del sitio activo, por lo que posiblemente la tapa y el canal trabajen concertadamente en la dinámica de la catálisis.

En otro trabajo publicado recientemente, a partir de el análisis de acoplamiento estadístico del barril de la familia funcional de IGPS, revelan de manera similar a lo encontrado por el análisis de nuestras dos familias de barriles  $(\beta/\alpha)_8$ , que los sitios acoplados estadísticamente se encuentran principalmente circundando el sitio activo <sup>(69)</sup>. La mayoría de estos residuos se encuentran presentes en las hojas  $\beta$  y en las asas  $\beta\alpha$ , sugiriendo su relevancia para el mantenimiento de una relación adecuada de estabilidad y actividad de las enzimas, probablemente interviniendo en adecuar la geometría funcional de sitio activo permitiendo una adecuada entrada del sustrato y liberación del producto.

Estos trabajos sugieren patrones de co-evolución general dentro de diferentes familias de barriles, lo cual indica que en la evolución, no sólo se seleccionan plegamientos para lograr las actividades necesarias por la naturaleza, sino que además se mantienen mecanismos generalizados para lograr una adecuada relación de actividad y estabilidad dentro de los mismos.

Finalmente, estudios previos tanto experimentales como computacionales, describen redes de interacciones indispensables para lograr el estado de transición de la enzima necesaria para la catálisis, estos sitios se han encontrado estar acoplados termodinámicamente como estadísticamente <sup>(15, 28, 92)</sup>. Esto nos da a pensar que la red de sitios estadísticamente acoplados encontrados por nuestro análisis juega un papel similar. Para corroborar nuestra hipótesis estudios más detallados de dinámica molecular como mutagénesis deberán ser llevados a cabo.

## 6-PERSPECTIVAS

Es importante utilizar nuestro análisis para complementar estudios previos de diseño de barriles como la disección realizada por Silverman. En esta dirección actualmente nuestro grupo está diseñando el reemplazo de las unidades  $(\beta\alpha)_3-\beta_4$  de PRAI con la intención de generar diversidad de núcleos de plegamiento.

La realización de estos diseños sigue un enfoque semi-racional, mediante la sustitución de  $(\beta\alpha)_3-\beta_4$  por fragmentos de DNA complementarios que mantengan el patrón de co-evolución de los sitios con mayor acoplamiento funcional y el diseño del resto de los sitios con sustituciones derivadas de resultados arrojados por el algoritmo de diseño de proteínas Rosetta design.

Otro aspecto relevante que ha emergido en los últimos años es el descubrimiento de redes de movilidad indispensables para la catálisis. El descubrimiento de estas redes representa un reto formidable y actualmente existen pocos métodos capaces de extraer de una manera práctica la información de posibles sitios implicados en estas dinámicas. En este momento en nuestro grupo estamos desarrollando una metodología predictiva, de bajo costo en recursos computacionales, tiempo y fácil manejo, con la intención de contribuir en el desarrollo del conocimiento de la personalidad dinámica de las proteínas.

## 7-REFERENCIAS

1. Rubin-Pitel, S.B., and H. Zhao. 2006. Recent advances in biocatalysis by directed enzyme evolution. *Combinatorial chemistry & high throughput screening* 9:247-257.
2. Vasserot, A.P., C.D. Dickinson, Y. Tang, W.D. Huse, K.S. Manchester, and J.D. Watkins. 2003. Optimization of protein therapeutics by directed evolution. *Drug discovery today* 8:118-126.
3. Cohen, N., S. Abramov, Y. Dror, and A. Freeman. 2001. In vitro enzyme evolution: the screening challenge of isolating the one in a million. *Trends in biotechnology* 19:507-510.
4. Ball, P. 2005. Synthetic biology for nanotechnology. *Nanotechnology* 16(1):R1-R8(1).
5. Jermutus, L., and J. Pelletier. 2001. Creating and evaluating protein diversity. *Curr Opin Biotechnol* 12:331-333.
6. Davis, B.K. 2002. Molecular evolution before the origin of species. *Prog Biophys Mol Biol* 79:77-133.
7. Giver, L., A. Gershenson, P.O. Freskgard, and F.H. Arnold. 1998. Directed evolution of a thermostable esterase. *Proceedings of the National Academy of Sciences of the United States of America* 95:12809-12813.
8. Moore, J.C., and F.H. Arnold. 1996. Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. *Nat Biotechnol* 14:458-467.
9. Toscano, M.D., K.J. Woycechowsky, and D. Hilvert. 2007. Minimalist active-site redesign: teaching old enzymes new tricks. *Angew Chem Int Ed Engl* 46:3212-3236.
10. Oda, M., N. Ito, T. Tsumuraya, K. Suzuki, M. Sakakura, and I. Fujii. 2007. Thermodynamic and structural basis for transition-state stabilization in antibody-catalyzed hydrolysis. *J Mol Biol* 369:198-209.
11. Pokala, N., and T.M. Handel. 2001. Review: protein design--where we were, where we are, where we're going. *J Struct Biol* 134:269-281.
12. Bahar, I., and A.J. Rader. 2005. Coarse-grained normal mode analysis in structural biology. *Curr Opin Struct Biol* 15:586-592.
13. Kinch, L.N., and N.V. Grishin. 2002. Evolution of protein structures and functions. *Curr Opin Struct Biol* 12:400-408.
14. Daggett, V., and A. Fersht. 2003. The present view of the mechanism of protein folding. *Nat Rev Mol Cell Biol* 4:497-502.
15. Agarwal, P.K. 2006. Enzymes: An integrated view of structure, dynamics and function. *Microb Cell Fact* 5:2.
16. Vendruscolo, M., and C.M. Dobson. 2006. Structural biology. Dynamic visions of enzymatic reactions. *Science (New York, N.Y)* 313:1586-1587.
17. Poole, A.M., and R. Ranganathan. 2006. Knowledge-based potentials in protein design. *Curr Opin Struct Biol* 16:508-513.
18. Socolich, M., S.W. Lockless, W.P. Russ, H. Lee, K.H. Gardner, and R. Ranganathan. 2005. Evolutionary information for specifying a protein fold. *Nature* 437:512-518.
19. Grant, A., D. Lee, and C. Orengo. 2004. Progress towards mapping the universe of protein folds. *Genome Biol* 5:107.

20. Berman, H.M., T. Battistuz, T.N. Bhat, W.F. Bluhm, P.E. Bourne, K. Burkhardt, Z. Feng, G.L. Gilliland, L. Iype, S. Jain, P. Fagan, J. Marvin, D. Padilla, V. Ravichandran, B. Schneider, N. Thanki, H. Weissig, J.D. Westbrook, and C. Zardecki. 2002. The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* 58:899-907.
21. Montiel Molina, H.M., C. Millan-Pacheco, N. Pastor, and G. del Rio. 2008. Computer-based screening of functional conformers of proteins. *PLoS Comput Biol* 4:e1000009.
22. Henzler-Wildman, K., and D. Kern. 2007. Dynamic personalities of proteins. *Nature* 450:964-972.
23. Whitford, P.C., S. Gosavi, and J.N. Onuchic. 2008. Conformational transitions in adenylate kinase. Allosteric communication reduces misligation. *The Journal of biological chemistry* 283:2042-2048.
24. Henzler-Wildman, K.A., M. Lei, V. Thai, S.J. Kerns, M. Karplus, and D. Kern. 2007. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* 450:913-916.
25. Henzler-Wildman, K.A., V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M.A. Wilson, G.A. Petsko, M. Karplus, C.G. Hubner, and D. Kern. 2007. Intrinsic motions along an enzymatic reaction trajectory. *Nature* 450:838-844.
26. Naganathan, A.N., U. Doshi, A. Fung, M. Sadqi, and V. Munoz. 2006. Dynamics, energetics, and structure in protein folding. *Biochemistry* 45:8466-8475.
27. Tousignant, A., and J.N. Pelletier. 2004. Protein motions promote catalysis. *Chem Biol* 11:1037-1042.
28. Hammes-Schiffer, S., and S.J. Benkovic. 2006. Relating protein motion to catalysis. *Annu Rev Biochem* 75:519-541.
29. Estabrook, R.A., J. Luo, M.M. Purdy, V. Sharma, P. Weakliem, T.C. Bruice, and N.O. Reich. 2005. Statistical coevolution analysis and molecular dynamics: identification of amino acid pairs essential for catalysis. *Proceedings of the National Academy of Sciences of the United States of America* 102:994-999.
30. Akanuma, S., H. Miyagawa, K. Kitamura, and A. Yamagishi. 2005. A detailed unfolding pathway of a (beta/alpha)<sub>8</sub>-barrel protein as studied by molecular dynamics simulations. *Proteins* 58:538-546.
31. Wako, H., M. Kato, and S. Endo. 2004. ProMode: a database of normal mode analyses on protein molecules with a full-atom model. *Bioinformatics* 20:2035-2043.
32. Dodson, G., and C.S. Verma. 2006. Protein flexibility: its role in structure and mechanism revealed by molecular simulations. *Cell Mol Life Sci* 63:207-219.
33. Fridman, E., and E. Pichersky. 2005. Metabolomics, genomics, proteomics, and the identification of enzymes and their substrates and products. *Curr Opin Plant Biol* 8:242-248.
34. Redfern, O.C., B. Dessailly, and C.A. Orengo. 2008. Exploring the structure and function paradigm. *Curr Opin Struct Biol* 18:394-402.
35. Jocker, A., F. Hoffmann, A. Groscurth, and H. Schoof. 2008. Protein function prediction and annotation in an integrated environment powered by web services (AFAWE). *Bioinformatics* 24:2393-2394.
36. Kawabata, T., M. Ota, and K. Nishikawa. 1999. The Protein Mutant Database. *Nucleic Acids Res* 27:355-357.

37. E. Lindahl, B.H.a.D.v.d.S. 2001. GROMACS 3.0: A package for molecular simulation and trajectory analysis. . *J. Mol. Mod.* 7:306-317
38. Gerstein, M., and W. Krebs. 1998. A database of macromolecular motions. *Nucleic Acids Res* 26:4280-4290.
39. Frank, M., P. Gutbrod, C. Hassayoun, and C.W. von Der Lieth. 2003. Dynamic molecules: molecular dynamics for everyone. An internet-based access to molecular dynamic simulations: basic concepts. *J Mol Model* 9:308-315.
40. Hollup, S.M., G. Salensminde, and N. Reuter. 2005. WEBnm@: a web application for normal mode analyses of proteins. *BMC Bioinformatics* 6:52.
41. Dobbins, S.E., V.I. Lesk, and M.J. Sternberg. 2008. Insights into protein flexibility: The relationship between normal modes and conformational change upon protein-protein docking. *Proceedings of the National Academy of Sciences of the United States of America* 105:10390-10395.
42. Rothlisberger, D., O. Khersonsky, A.M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J.L. Gallaher, E.A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K.N. Houk, D.S. Tawfik, and D. Baker. 2008. Kemp elimination catalysts by computational enzyme design. *Nature* 453:190-195.
43. Andrusier, N., E. Mashiach, R. Nussinov, and H.J. Wolfson. 2008. Principles of flexible protein-protein docking. *Proteins* 73:271-289.
44. Bairoch, A., R. Apweiler, C.H. Wu, W.C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M.J. Martin, D.A. Natale, C. O'Donovan, N. Redaschi, and L.S. Yeh. 2005. The Universal Protein Resource (UniProt). *Nucleic Acids Res* 33:D154-159.
45. Schomburg, I., A. Chang, and D. Schomburg. 2002. BRENDA, enzyme data and metabolic information. *Nucleic Acids Res* 30:47-49.
46. Porter, C.T., G.J. Bartlett, and J.M. Thornton. 2004. The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data. *Nucleic Acids Res* 32:D129-133.
47. Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389-3402.
48. Gerlt, J.A., and P.C. Babbitt. 2001. Barrels in pieces? *Nature structural biology* 8:5-7.
49. Holm, L., and C. Sander. 1996. Mapping the protein universe. *Science (New York, N.Y)* 273:595-603.
50. Russ, W.P., D.M. Lowery, P. Mishra, M.B. Yaffe, and R. Ranganathan. 2005. Natural-like function in artificial WW domains. *Nature* 437:579-583.
51. Dahiyat, B.I., C.A. Sarisky, and S.L. Mayo. 1997. De novo protein design: towards fully automated sequence selection. *J Mol Biol* 273:789-796.
52. Kuhlman, B., G. Dantas, G.C. Ireton, G. Varani, B.L. Stoddard, and D. Baker. 2003. Design of a novel globular protein fold with atomic-level accuracy. *Science (New York, N.Y)* 302:1364-1368.
53. Jiang, L., E.A. Althoff, F.R. Clemente, L. Doyle, D. Rothlisberger, A. Zanghellini, J.L. Gallaher, J.L. Betker, F. Tanaka, C.F. Barbas, 3rd, D. Hilvert, K.N. Houk, B.L. Stoddard, and D. Baker. 2008. De novo computational design of retro-aldol enzymes. *Science (New York, N.Y)* 319:1387-1391.

54. Zanghellini, A., L. Jiang, A.M. Wollacott, G. Cheng, J. Meiler, E.A. Althoff, D. Rothlisberger, and D. Baker. 2006. New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci* 15:2785-2794.
55. Dahiyat, B.I., and S.L. Mayo. 1997. De novo protein design: fully automated sequence selection. *Science (New York, N.Y)* 278:82-87.
56. Nagano, N., E.G. Hutchinson, and J.M. Thornton. 1999. Barrel structures in proteins: automatic identification and classification including a sequence analysis of TIM barrels. *Protein Sci* 8:2072-2084.
57. Nagano, N., C.A. Orengo, and J.M. Thornton. 2002. One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J Mol Biol* 321:741-765.
58. Vega, M.C., E. Lorentzen, A. Linden, and M. Wilmanns. 2003. Evolutionary markers in the (beta/alpha)<sub>8</sub>-barrel fold. *Curr Opin Chem Biol* 7:694-701.
59. Hommel, U., M. Eberhard, and K. Kirschner. 1995. Phosphoribosyl anthranilate isomerase catalyzes a reversible amidori reaction. *Biochemistry* 34:5429-5439.
60. Eberhard, M., M. Tsai-Pflugfelder, K. Bolewska, U. Hommel, and K. Kirschner. 1995. Indoleglycerol phosphate synthase-phosphoribosyl anthranilate isomerase: comparison of the bifunctional enzyme from *Escherichia coli* with engineered monofunctional domains. *Biochemistry* 34:5419-5428.
61. Hennig, M., R. Sterner, K. Kirschner, and J.N. Jansonius. 1997. Crystal structure at 2.0 Å resolution of phosphoribosyl anthranilate isomerase from the hyperthermophile *Thermotoga maritima*: possible determinants of protein stability. *Biochemistry* 36:6009-6016.
62. Lockless, S.W., and R. Ranganathan. 1999. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science (New York, N.Y)* 286:295-299.
63. Ciria, R., C. Abreu-Goodger, E. Morett, and E. Merino. 2004. GeConT: gene context analysis. *Bioinformatics* 20:2307-2308.
64. Cuff, J.A., M.E. Clamp, A.S. Siddiqui, M. Finlay, and G.J. Barton. 1998. JPred: a consensus secondary structure prediction server. *Bioinformatics* 14:892-893.
65. Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792-1797.
66. Getz, G., E. Levine, and E. Domany. 2000. Coupled two-way clustering analysis of gene microarray data. *Proceedings of the National Academy of Sciences of the United States of America* 97:12079-12084.
67. Saeed, A.I., V. Sharov, J. White, J. Li, W. Liang, N. Bhagabati, J. Braisted, M. Klapa, T. Currier, M. Thiagarajan, A. Sturn, M. Snuffin, A. Rezantsev, D. Popov, A. Ryltsov, E. Kostukovich, I. Borisovsky, Z. Liu, A. Vinsavich, V. Trush, and J. Quackenbush. 2003. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34:374-378.
68. Wilmanns, M., J.P. Priestle, T. Niermann, and J.N. Jansonius. 1992. Three-dimensional structure of the bifunctional enzyme phosphoribosylanthranilate isomerase: indoleglycerolphosphate synthase from *Escherichia coli* refined at 2.0 Å resolution. *J Mol Biol* 223:477-507.
69. Shen, H., F. Xu, H. Hu, F. Wang, Q. Wu, Q. Huang, and H. Wang. 2008. Coevolving residues of (beta/alpha)<sub>8</sub>-barrel proteins play roles in stabilizing active site architecture and coordinating protein dynamics. *J Struct Biol*

70. Kursula, I., M. Salin, J. Sun, B.V. Norledge, A.M. Haapalainen, N.S. Sampson, and R.K. Wierenga. 2004. Understanding protein lids: structural analysis of active hinge mutants in triosephosphate isomerase. *Protein Eng Des Sel* 17:375-382.
71. Hayward, S., and H.J. Berendsen. 1998. Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. *Proteins* 30:144-154.
72. Hayward, S., and R.A. Lee. 2002. Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *J Mol Graph Model* 21:181-183.
73. Kurkcuoglu, O., R.L. Jernigan, and P. Doruker. 2006. Loop motions of triosephosphate isomerase observed with elastic networks. *Biochemistry* 45:1173-1182.
74. Wolf-Watz, M., V. Thai, K. Henzler-Wildman, G. Hadjipavlou, E.Z. Eisenmesser, and D. Kern. 2004. Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nature structural & molecular biology* 11:945-949.
75. Williams, J.C., and A.E. McDermott. 1995. Dynamics of the flexible loop of triosephosphate isomerase: the loop motion is not ligand gated. *Biochemistry* 34:8309-8319.
76. Rozovsky, S., G. Jogl, L. Tong, and A.E. McDermott. 2001. Solution-state NMR investigations of triosephosphate isomerase active site loop motion: ligand release in relation to active site loop dynamics. *J Mol Biol* 310:271-280.
77. Rozovsky, S., and A.E. McDermott. 2001. The time scale of the catalytic loop motion in triosephosphate isomerase. *J Mol Biol* 310:259-270.
78. Akanuma, S., and A. Yamagishi. 2005. Identification and characterization of key substructures involved in the early folding events of a (beta/alpha)<sub>8</sub>-barrel protein as studied by experimental and computational methods. *J Mol Biol* 353:1161-1170.
79. Hocker, B., S. Beismann-Driemeyer, S. Hettwer, A. Lustig, and R. Sterner. 2001. Dissection of a (beta/alpha)<sub>8</sub>-barrel enzyme into two folded halves. *Nature structural biology* 8:32-36.
80. Henn-Sax, M., R. Thoma, S. Schmidt, M. Hennig, K. Kirschner, and R. Sterner. 2002. Two (beta/alpha)<sub>8</sub>-barrel enzymes of histidine and tryptophan biosynthesis have similar reaction mechanisms and common strategies for protecting their labile substrates. *Biochemistry* 41:12032-12042.
81. O'Donoghue, A.C., T.L. Amyes, and J.P. Richard. 2005. Hydron transfer catalyzed by triosephosphate isomerase. Products of isomerization of dihydroxyacetone phosphate in D<sub>2</sub>O. *Biochemistry* 44:2622-2631.
82. Orosz, F., J. Olah, and J. Ovadi. 2006. Triosephosphate isomerase deficiency: facts and doubts. *IUBMB Life* 58:703-715.
83. Espinoza-Fonseca, L.M., and J.G. Trujillo-Ferrara. 2004. Exploring the possible binding sites at the interface of triosephosphate isomerase dimer as a potential target for anti-tripanosomal drug design. *Bioorg Med Chem Lett* 14:3151-3154.
84. Silverman, J.A., R. Balakrishnan, and P.B. Harbury. 2001. Reverse engineering the (beta/alpha)<sub>8</sub> barrel fold. *Proceedings of the National Academy of Sciences of the United States of America* 98:3092-3097.
85. Kallenbach, N. 2001. Breaking open a protein barrel. *Proceedings of the National Academy of Sciences of the United States of America* 98:2958-2960.

86. Mande, S.C., V. Mainfroid, K.H. Kalk, K. Goraj, J.A. Martial, and W.G. Hol. 1994. Crystal structure of recombinant human triosephosphate isomerase at 2.8 Å resolution. Triosephosphate isomerase-related human genetic disorders and comparison with the trypanosomal enzyme. *Protein Sci* 3:810-821.
87. Kursula, I., S. Partanen, A.M. Lambeir, D.M. Antonov, K. Augustyns, and R.K. Wierenga. 2001. Structural determinants for ligand binding and catalysis of triosephosphate isomerase. *Eur J Biochem* 268:5189-5196.
88. Olah, J., F. Orosz, G.M. Keseru, Z. Kovari, J. Kovacs, S. Hollan, and J. Ovadi. 2002. Triosephosphate isomerase deficiency: a neurodegenerative misfolding disease. *Biochem Soc Trans* 30:30-38.
89. Massi, F., C. Wang, and A.G. Palmer, 3rd. 2006. Solution NMR and computer simulation studies of active site loop motion in triosephosphate isomerase. *Biochemistry* 45:10787-10794.
90. Lolis, E., and G.A. Petsko. 1990. Crystallographic analysis of the complex between triosephosphate isomerase and 2-phosphoglycolate at 2.5-Å resolution: implications for catalysis. *Biochemistry* 29:6619-6625.
91. Sullivan, S.M., and T. Holyoak. 2008. Enzymes with lid-gated active sites must operate by an induced fit mechanism instead of conformational selection. *Proceedings of the National Academy of Sciences of the United States of America* 105:13829-13834.
92. Chen, J., R.I. Dima, and D. Thirumalai. 2007. Allosteric communication in dihydrofolate reductase: signaling network and pathways for closed to occluded transition and back. *J Mol Biol* 374:250-266.