



**UNIVERSIDAD NACIONAL AUTONOMA  
DE MEXICO**

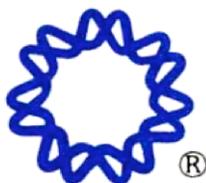
---

**PROGRAMA DE MAESTRIA Y DOCTORADO  
EN CIENCIAS BIOQUIMICAS  
INSTITUTO DE BIOTECNOLOGIA  
DEPARTAMENTO DE INGENIERIA CELULAR Y BIOCATALISIS**

**OBTENCION DE PROTEINAS PLEGADAS A TRAVES  
DE LA COMPLEMENTACION DE UNA UNIDAD  
( $\beta/\alpha$ )<sub>4</sub>  
CON SECUENCIAS GENOMICAS**

**T E S I S**  
QUE PARA OBTENER EL GRADO ACADEMICO DE  
**MAESTRO EN CIENCIAS**  
**P R E S E N T A :**  
**JOSE ARCADIO FARIAS RICO**

DIRECTOR DE TESIS:  
DR. LORENZO SEGOVIA FORCELLA



**CUERNAVACA, MORELOS 2008**



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.

The past is already past.  
Don't try to regain it.  
The present does not stay.  
Don't try to touch it.

From moment to moment  
The future has not come;  
Don't think about it  
Beforehand.

Whatever comes to the eye,  
Leave it be.  
There are no commandments  
To be kept;  
There's no filth to be cleansed.

With empty mind really  
Penetrated, the dharmas  
Have no life.

When you can be like this,  
You've completed  
The ultimate attainment.  
**Layman P'ang.**

## **Para**

Guadalupe, Juan José, León Fernando, Víctor Hugo y Satti, la fundación Farías Rico.

## **Agradecimientos**

A Beto, Eva, Evita, Rigoberto.

A todos los camaradas ibetlantes que me ayudaron de cualquier manera y también a los que por alguna razón se me olvidó mencionar aquí, ya ustedes saben. Lorenzo (el gran jefe), Ricardo (ni un homeless más), Viviana (el oráculo de Delfos), Daniela (hermanilla), Mónica, Adriana, Javier, Fernando, Jimena, Ettiene, Iliana, Mariana, Eugenio, Areli, Marel, Heaven, Dago, Alex, Sergio, Yadira, Maria, Fili, Joel, Flores, Adrian, Ernesto.

Las mil gracias a la bandera externa: Mike y Jan, la fundación sucursal Cuernavaca; Luis, Daniel y Omar: los puros cuates; Oscar y Lorna. Los defectuosos: Eduardo, Raúl, Diana, Constanza, Ale, Heide. Los rucos: Jorge, Hugo, Chucho, Irving, Yael.

Mariel (thundering silence, colorful darkness wanting to be free).

Quiero agradecer la correcta orientación proporcionada por mi comité tutorial:

Dra. Gloria Saab Rincón  
Dr. Arturo Carlos Il Becerra Bracho

De igual forma agradezco la participación de los miembros del jurado:

Presidente	Dr. Francisco Xavier Soberon Mainero
Secretario	Dr. Ricardo Canek Rodríguez de la Vega
Vocal	Dr. Alejandro Sosa Peinado
Suplente	Dr. Enrique Rudiño Piñera
Suplente	Dr. Lorenzo Segovia Forcella

Este trabajo pudo realizarse gracias al apoyo otorgado por el Consejo Nacional de Ciencia y Tecnología

Abreviaturas	
PRAI	Fosforibosil antranilato isomerasa (por sus siglas en inglés)
CAT	Cloramfenicol acetil transferasa
ADN	Ácido desoxirribonucleico
SCOP	Base de datos que consiste en una clasificación estructural de proteínas Clasificación estructural de proteínas (por sus siglas en inglés)
PDB	Base de datos de estructuras proteicas Banco de datos de proteínas (por sus siglas en inglés)
EMBL	Laboratorio Europeo de Biología Molecular (por su siglas en inglés)
TrEMBL	Base de datos del EMBL (por sus siglas en inglés) traducida
Pfam	Base de datos de dominios proteicos
TIM	Triosa Fosfato Isomerasa (por sus siglas en inglés)
ARN	Ácido ribonucleico
ORF	Marco abierto de lectura (por sus siglas en inglés)
NAMD	Paquete de simulación dinámica molecular
PSIPRED	Servidor de predicción de estructura secundaria de proteínas (por sus siglas en inglés)
TRIS	Hidroximetilaminometano (amortiguador)
DTT	Ditiotreitol
BSA	Albúmina bovina
EDTA	Ácido etilen diamino tetra-acético
TAE	Tris acetato EDTA (amortiguador)
SDS	Dodecil sulfato de sodio (por sus siglas en ingles)
TE	Tris EDTA (amortiguador)
PCR	Reacción en cadena de la polimerasa
dNTPs	Nucleótidos
H <sub>2</sub> O 4D	Agua tetradestilada
LGS	Solución de gel inferior (amortiguador)
UGS	Solución de gel superior (amortiguador)
TEMED	Tetrametiletilendiamina
APS	Persulfato amónico
LB	Medio de cultivo Luria Bertani
IPTG	Isopropil $\beta$ -D-1-Tiogalactopiranosido

Abreviaturas	
PBS	Amortiguador de fosfatos salino
PMSF	Fenilmetilsulfonil fluoruro
TBST	Amortiguador tris salino con tween 20 (detergente)
kDa	Kilo daltones
BLAST	Herramienta básica de alineamiento local (por sus siglas en inglés)

<b>Tabla de contenidos</b>	<b>Página</b>
<b>Resumen</b>	<b>7</b>
<b>Prefacio</b>	<b>8</b>
<b>Introducción</b>	<b>10</b>
<b>Antecedentes</b>	<b>17</b>
<b>Hipótesis</b>	<b>22</b>
<b>Objetivos</b>	<b>22</b>
<b>Metodología General</b>	<b>23</b>
<b>Métodos y Materiales</b>	<b>24</b>
<b>Resultados</b>	<b>36</b>
<b>Resultados Biología Molecular</b>	<b>37</b>
<b>Resultados Bioquímica</b>	<b>50</b>
<b>Resultados Bioinformática</b>	<b>56</b>
<b>Discusión y Perspectivas</b>	<b>62</b>
<b>Conclusiones</b>	<b>66</b>
<b>Apéndice</b>	<b>67</b>
<b>Referencias</b>	<b>71</b>

## Resumen

Debido a que las proteínas son los agentes funcionales principales de toda la vida sobre la tierra, el estudio de su origen y evolución constituye un objetivo central de la ciencia moderna. Estas biomoléculas adoptan diversas estructuras en el espacio, lo que les permite realizar múltiples funciones. La secuencia de aminoácidos de una proteína determina la estructura tridimensional o plegamiento que adopta; por tal motivo, en este trabajo complementamos la información estructural contenida en una unidad  $(\beta/\alpha)_4$ , la mitad de PRAI consenso [proteína sintética construida en un trabajo previo en este laboratorio y que adopta el plegamiento  $(\beta/\alpha)_8$ ] con secuencias al azar, obtenidas a partir de la fragmentación de ADN genómico de *Escherichia coli*. Generamos un banco de variantes clonando los fragmentos genómicos corriente arriba de la mitad carboxilo terminal de PRAI consenso [PRAI-C- $(\beta/\alpha)_4$ ]; esta librería de proteínas quiméricas fue tamizada mediante un sistema reportero de plegamiento. Este sistema involucra una fusión traduccional del banco de variantes con la cloramfenicol acetil transferasa (CAT). Si la variante proteica fusionada a CAT es capaz de plegarse correctamente, CAT puede plegarse también, confiriéndole resistencia al cloramfenicol a la bacteria que expresa la construcción. Con la utilización de este sistema logramos obtener 8 proteínas quiméricas que conformaron fusiones traduccionales solubles (plegadas) con CAT. Realizamos la caracterización bioinformática de los fragmentos que complementaron la unidad PRAI-C- $(\beta/\alpha)_4$ , identificando y modelando los elementos de estructura secundaria que fueron capaces de complementar nuestro templado. Frecuentemente la solubilidad que presenta una proteína, estimada en condiciones fisiológicas, se relaciona directamente con la estabilidad de su plegamiento; finalmente, evaluamos la solubilidad de las quimeras utilizando métodos bioquímicos. Estos análisis se realizaron con las quimeras fusionadas a CAT y separadas de la fusión. Mediante este trabajo de investigación experimental, aportamos información relevante acerca del origen y el camino evolutivo que pudo seguir el plegamiento  $(\beta/\alpha)_8$  para alcanzar su forma actual.

## Prefacio

La búsqueda del conocimiento ha constituido una parte muy importante del espíritu humano desde el inicio de la historia moderna. Es en esta búsqueda donde se origina la ciencia como la conocemos ahora, integrándose en una serie de metodologías y procedimientos que nos permiten generar teorías y modelos para describir el mundo natural. El quehacer científico motivado por las necesidades inherentemente humanas de descubrir cómo funciona la naturaleza, constituye una de las principales fuerzas motrices que impulsan nuestro desarrollo como civilización.

La arquitectura de la vida es impresionante; basta con observar la complejidad del ojo humano o las alas de una mariposa, para inmediatamente preguntarnos cómo ha surgido la composición del mundo que nos rodea. Este enigma ha sido objeto de reflexión de numerosos filósofos naturales. Podemos encontrar muchos antecedentes en la búsqueda del origen del orden natural: el mundo de las ideas de Platón, el transformacionismo limitado de Jean Lamarck, o la ciencia de los fósiles de Georges Cuvier.

La historia del conocimiento científico biológico se encuentra compuesta de tesis y antítesis que finalmente se unifican en una síntesis, capaz de modificar radicalmente nuestra visión del origen de las formas naturales: la teoría evolutiva descrita por Charles Darwin en *El origen de las especies*, que se constituyó como una de las más importantes explicaciones del origen del orden que caracteriza al mundo natural. Después de haber estado en contacto con la megadiversidad natural en el viaje que realizó por el mundo en el *Beagle*, y de haber leído por casualidad el *Ensayo sobre la población* de Malthus, Darwin fue capaz de explicar el origen de la vasta cantidad de especies que observamos en el mundo. La publicación de la obra maestra de este célebre científico en 1859 marca el inicio de una nueva etapa en el desarrollo del pensamiento científico.

En la mayoría de las especies podemos observar una marcada propensión a la multiplicación. Esta tendencia se aprecia muy claramente en la historia de vida de arácnidos o insectos, que tienen más crías de las que pueden sobrevivir. Estas crías van a formar en conjunto una población natural que muestra cierta variabilidad, que a su vez se traduce en adaptación cuando las poblaciones se ven sometidas a la selección natural. Solamente los individuos que se encuentren mejor adaptados serán capaces de sobrevivir a esta presión evolutiva.

Podemos ejemplificar de una manera muy clara lo que significa esta fuerza evolutiva al observar la selección artificial que el hombre ha hecho desde hace muchos siglos con las razas de perros o caballos. Los criadores cruzan caballos pura sangre para obtener una prole de campeones. La naturaleza actúa de manera similar al seleccionar al león que sea el predador más efectivo o la variedad de maíz que mejor se pueda defender de las plagas de insectos. Con el desarrollo de la teoría evolutiva de Darwin se podía explicar el origen de las especies.

Sin embargo, aún no quedaba claro cómo se heredaban las variaciones que surgían en las poblaciones naturales, ni mucho menos el porqué éstas ocurrían. Con el posterior avance del conocimiento científico somos testigos de cómo poco a poco se van descubriendo los mecanismos que Darwin no había clarificado en su teoría (Rostand 1985).

No tuvo que pasar mucho tiempo para que August Weismann iniciara los trabajos para definir la teoría cromosómica de la herencia. Un poco después, alrededor de 1900, se redescubren las memorias de Mendel, donde se establecen las bases de las leyes de hibridación. Y es finalmente, cuando al parejo del hallazgo del naturalista agustino, encontramos los primeros esfuerzos desarrollados por Hugo de Vries para definir la mutación, lo que ahora conocemos como el origen de las variaciones en las poblaciones naturales. De esta manera, hemos realizado una breve revisión de la historia del conocimiento científico-biológico, hasta antes del establecimiento de la Biología Molecular.

## Introducción

Al tratar de explicar los fenómenos vitales a partir de las macromoléculas que los constituyen, es cuando somos testigos del surgimiento de la biología molecular. Algunos de los descubrimientos más significativos que forman los cimientos de esta nueva ciencia son: la descripción de la estructura del ADN (la biomolécula de la herencia) por Watson y Crick, el reconocimiento de las proteínas como las macromoléculas que expresan la información contenida en el ADN y el desarrollo de los métodos de secuenciación de las mismas, desarrollados por Fred Sanger.

El origen de las variaciones en las poblaciones naturales se encuentra en los cambios que sufre la secuencia de ADN del genoma (repertorio genético de una especie determinada) de los organismos (Hughes 2005). Una manifestación importante de estos cambios es expresada en las proteínas que codifica; la biología molecular describe la estructura de estas proteínas y las funciones que pueden llevar a cabo en términos fisicoquímicos. El desarrollo de esta rama no puede existir sin el surgimiento de la evolución molecular, como una serie de metodologías que nos permiten entender y analizar la gran cantidad de información evolutiva que se encuentra contenida tanto en la estructura como en la secuencia de los ácidos nucleicos y las proteínas.

Es importante detenernos para hacer una diferenciación en el nivel de descripción de los estudios evolutivos: macroevolución y microevolución. Cuando hablamos de macroevolución nos referimos a los eventos que han transcurrido en el pasado distante, las relaciones que se establecen entre los distintos *phyla*, el origen de las categorías principales de organismos y las relaciones que observamos entre el medio ambiente y las especies (ya sea en su morfología o fisiología).

Los análisis microevolutivos involucran el estudio de procesos y mecanismos que ocurren en el presente, estos fenómenos incluyen la generación de diversidad genética y las distintas formas de selección. Las mutaciones en el material genético pueden ir desde mutaciones puntuales, hasta duplicaciones génicas (uno de los principales mecanismos descritos para el nacimiento de nuevas funciones proteicas) o rearrreglos cromosómicos. Para la selección existe un continuo que va desde la selección negativa (purificadora), pasando por las mutaciones neutrales hasta llegar a la selección positiva.

Nos damos cuenta que es gracias al desarrollo de la biología molecular y su complementación con estudios físicos, en la resolución de preguntas biológicas, que en las últimas décadas observamos un avance sin precedentes en los estudios microevolutivos. Sabemos que una de las metas principales de este tipo de aproximaciones es que seamos capaces de explicar la biología de las biomoléculas en términos físicos y químicos.

La evolución molecular es una ciencia relativamente nueva que nos permite contestar numerosas preguntas que hasta hace algunos años era muy difícil responder. Ahora no sólo podemos explicar la enorme cantidad de variación genética (con el desarrollo de la teoría neutral de evolución molecular) existente en las poblaciones naturales, sino también medirla.

Del mismo modo somos capaces de demostrar de una manera clara la inexistencia de un carácter determinista en los procesos evolutivos, así como el fuerte componente estocástico presente en ellos. Los estudios evolutivos a nivel molecular nos permiten explicar muchos fenómenos naturales, como puede ser el origen y posterior establecimiento tanto de funciones, como de estructuras proteicas.

El análisis de la evolución de la macromolécula de la herencia y la expresión de su producto proteico nos ha permitido establecer relaciones filogenéticas entre las distintas especies que componen la biosfera. La secuenciación de proteínas y su comparación entre distintos organismos es una de las herramientas más poderosas que complementan los estudios de evolución molecular.

### **Evolución de Proteínas**

Una de las demostraciones más importantes del proceso evolutivo se encuentra reflejada en la enorme diversidad de secuencias proteicas que se han descrito hasta el momento. La teoría neutral de evolución molecular, nos capacita para entender cómo han surgido a lo largo del tiempo como una consecuencia de la mutación espontánea. Estas mutaciones se han fijado en el complemento genético de todos los sistemas orgánicos que componen la biosfera como un efecto de la deriva génica y de la selección natural. La secuenciación del genoma completo de muchos organismos ha revelado cerca de un millón de secuencias de aminoácidos tan solo a principios del 2004 (Grant 2004), mientras que en uno de los proyectos de secuenciación masiva de muestras ambientales más grande realizado en la actualidad (Venter 2004) se lograron identificar cerca de 1.2 millones de genes previamente desconocidos.

Debido a que las proteínas son los agentes principales de toda la vida sobre la tierra, el estudio de su evolución es uno de los objetivos centrales de la ciencia moderna. A nivel celular juegan diversos papeles; entre otras cosas, son catalizadores de reacciones importantes en el metabolismo, se encuentran formando parte fundamental de la estructura de complejos ribo-núcleo-proteicos en la maquinaria de traducción, y también son efectores importantes en el reconocimiento de proteínas susceptibles de degradación.

Se ha propuesto que uno de los procesos de generación de nuevas funciones, es el de duplicación génica. Durante los mecanismos de replicación del material hereditario, a veces ocurren fallos en la maquinaria enzimática que derivan en genomas conteniendo dos copias del mismo gen. Ahora bien, es posible que uno de los genes sufra un proceso de mutagénesis acelerado, ya que no se encuentra sometido a presión selectiva porque la función del gen duplicado en el organismo, está siendo realizada por la otra copia. Posteriormente, cabe la posibilidad de que el genoma en el que se duplicaron estos genes se convierta en el más común dentro de la población natural a través de la deriva génica. El resultado del fenómeno antes descrito es la invención de una proteína que puede desempeñar una función que antes era inexistente. Este tipo de proteínas se denominan parálogos (proteínas duplicadas con diferente función presentes en el mismo organismo).

Hablando en términos filogenéticos (ancestría común), existe otro tipo de proteínas que también comparten un mismo ancestro (homólogas), que son denominadas proteínas ortólogas; estas proteínas se originan en procesos de especiación, por lo que se encuentran realizando la misma función en distintos organismos. Para que las proteínas sean capaces de realizar sus múltiples funciones deben adoptar una estructura tridimensional en el espacio que es determinada por su secuencia de aminoácidos.

Si analizamos las hipótesis descritas del posible origen de proteínas nuevas, podemos suponer que dos polipéptidos homólogos (ya sean parálogos u ortólogos) al provenir de un mismo precursor, también tendrían una secuencia de aminoácidos similar y por ende una estructura parecida. Aunque esto ha sido demostrado repetidas veces (Petsko 2004), la relación que se establece entre la secuencia de aminoácidos de una proteína, su plegamiento y la función que desempeña, aún representa uno de los problemas más confusos que existen. Una de las demostraciones de la complejidad que encierra la relación estructura-función en proteínas la encontramos en un trabajo de Alexander en el 2007, en el cual su grupo fue capaz de caracterizar dos proteínas con una identidad de secuencia de un 88%, pero presentando tanto una estructura como una función diferentes.

La duplicación génica ha jugado un papel muy importante en la historia de la vida sobre la tierra; inventar genes con funciones nuevas hacen posible que los organismos se adapten a distintos ambientes. Una de las pruebas más concluyentes de que la duplicación de genes es uno de los mecanismos más importantes que moldean la estructura de los genomas, es la existencia de familias génicas, particularmente en eucariontes. Es un hecho que la mayoría de las duplicaciones se dan a nivel de genes; sin embargo, se ha propuesto también que la duplicación de genomas completos es un mecanismo evolutivo sumamente importante. Wolfe en 1997 demuestra que *Saccharomyces cerevisiae* es un tetraploide degenerado que surgió como resultado de un proceso de duplicación del genoma después de la divergencia de *Saccharomyces* desde *Kluyveromyces* (Wolfe 1997).

El intenso estudio que se ha desarrollado alrededor de la evolución de proteínas nos ha permitido explorar la vida utilizando inferencias basadas en homología (ancestría común) en diversos sistemas modelo. Una muestra de los estudios rutinarios que se desarrollan en el laboratorio, lo ejemplificamos cuando hacemos búsquedas en bases de datos de proteínas (swiss-prot, TrEmBL, etc.) comparando secuencias de genes que no conocemos con sus homólogos caracterizados para tratar de identificar su función. Una de las interrogantes más importantes que ha arrojado el desarrollo de la evolución molecular es el problema de la relación estructura-función en proteínas. En la última versión del SCOP (Alexey 2007) se contaban mil plegamientos distintos para todas las proteínas con estructura determinada en el PDB. Se estima que el número de estructuras que todo el universo proteico podría adoptar no se acerca a 5000, sin embargo se han descrito muchas más funciones. De allí que encontramos que más de una secuencia de aminoácidos codifica para la misma estructura (Holm 1996). No es posible encontrar una correlación clara entre el número de estructuras que las proteínas pueden adoptar y las funciones que son capaces de realizar.

Valiéndonos de herramientas evolutivas, como la inferencia filogenética en el estudio de diversos sistemas modelo, podemos dar respuesta a varias preguntas que son clave en el problema de la relación estructura y función de las proteínas. Dado que la estructura tridimensional de una proteína se encuentra relacionada de manera íntima con su función (Rost 2002), el análisis de la misma juega un papel muy importante al tratar de entender el papel biológico de estas macromoléculas. En algunos casos es posible predecir ciertas características de las proteínas basándonos en las propiedades individuales de sus aminoácidos constituyentes, como pueden ser el punto isoeléctrico, el peso molecular y la hidrofobicidad, entre otras. Sin embargo, aún no es posible establecer modelos simplificados para hacer predicciones con buenos resultados en todos los casos.

Existen varias definiciones del término “plegamiento”; intuitivamente se refiere a un segmento polipeptídico de entre 30 y 700 aminoácidos de largo, que permite a los biólogos estructurales reconocer una estructura proteica particular (Coulson 2002). Se ha postulado que existen cerca de 5000 formas distintas de plegar la megadiversidad de secuencias acumuladas hasta el momento en las bases de datos. Estas formas o plegamientos se han subdividido en uni-plegamientos (han surgido una sola vez en la naturaleza), super-plegamientos (aparecen repetidas veces en varias superfamilias de proteínas) y meso-plegamientos (se encuentran ubicados en una frecuencia intermedia entre uni-plegamientos y super-plegamientos) (Coulson 2002). La distribución de las secuencias entre las estructuras obedece una ley de potencias, en la cual existe un número pequeño de entidades que aparecen con una alta frecuencia (super-plegamientos), mientras que existe un número moderado de instancias comunes (meso-plegamientos) y un gran número de instancias raras (uni-plegamientos).

Este tipo de distribuciones se ha observado en bastantes aspectos del ámbito biológico, como son las interacciones proteína-proteína o las redes reguladoras de vías biosintéticas (Koonin 2002). Existen dos acercamientos principales mediante los cuales se ha tratado de explicar la existencia de tan pocos plegamientos para tantas secuencias polipeptídicas. El primero de ellos (Rost 2002) plantea que los plegamientos con los que contamos actualmente son los que surgieron de manera temprana en el tiempo evolutivo. Posteriormente, estos plegamientos mediante procesos de duplicación y divergencia, se especializaron en diversas funciones y alcanzaron su distribución actual. La alternativa a esta explicación (Soding 2003) plantea que existen limitaciones termodinámicas que delimitan el espacio de plegamientos que una secuencia de aminoácidos puede poblar. Esta última aseveración implicaría que las formas denominadas super-plegamientos son las más exitosas en términos fisicoquímicos, aunque también podrían ser las más funcionales debido a sus características estructurales.

Un ejemplo claro es el presentado por el plegamiento de barril ( $\beta/\alpha$ )<sub>8</sub> en el cual encontramos que los sitios catalíticos de la mayoría de las enzimas que presentan esta forma tridimensional, se encuentran ubicados en las asas carboxilo terminal que unen las repeticiones de motivos ( $\beta/\alpha$ ), mientras que la estabilidad global está contenida en las estructuras secundarias que forman el núcleo de la proteína (Wierenga 2001). Esta compartimentalización de funciones hace de este plegamiento un sustrato único para la evolución de una gran diversidad de catálisis.

Al observar las conexiones evolutivas que existen entre las diversas proteínas es posible predecir algunos aspectos tanto de su estructura como de su función (Holm 1996). A partir de un alineamiento múltiple de enzimas homólogas, es posible hacer inferencias acerca de su forma o de su mecanismo catalítico. A pesar de que las estructuras proteicas podrían parecer superficialmente diferentes, existen componentes estructurales comunes a muchas formas distintas. La mayoría de los plegamientos que tenemos en la actualidad han derivado de un complemento básico de unidades que son capaces de estructurarse de manera autónoma, los cuales son llamados dominios estructurales.

Aunque la evolución se ve reflejada en la divergencia a nivel de secuencia de aminoácidos entre diversos grupos de proteínas, los dominios estructurales podrían ser considerados como los átomos de la evolución (Vogel 2004). Para dar forma al universo proteico mediante una clasificación jerárquica, primero es necesario dividir las proteínas en sus dominios constituyentes haciendo uso de las relaciones evolutivas que se detecten entre ellos. Estas unidades se reutilizan frecuentemente en la composición de muchas estructuras distintas, es por esto que se pueden realizar clasificaciones proteicas basándonos en ellas como es el caso CATH (*Class, Architecture Topology and Homology*) o Pfam (*protein families*). Es posible que los dominios estructurales no fueran los átomos de la evolución, aunque los análisis sistemáticos basados en ellos son más exactos que otros basados en la estructura completa de las proteínas. Dos polipéptidos pudieran no presentar una similitud a nivel de secuencia que los agrupara como homólogos; por otra parte, es muy posible que ambos se encuentren compartiendo dominios homólogos.

Aunque algunos autores (Lupas 2001) consideran que ciertas superestructuras secundarias más pequeñas (topologías  $\beta\alpha\beta$ ,  $\alpha\alpha$ , *hairpin*, *coiled coil*) son los ladrillos constructores de la evolución, la recurrencia de los dominios completos en las bases de datos los posicionan como unidades evolutivas primarias. La definición, tanto de fronteras estructurales como de los dominios evolutivos, es una de las tareas más difíciles en el estudio de la evolución de las macromoléculas. Al observar las bases de datos de proteínas como CATH es claro que algunos motivos estructurales son altamente reutilizados.

Esto implicaría que el universo de plegamientos debería ser visto más exactamente como un continuo en donde ocurren sobrelapamientos estructuralmente significativos en algunas regiones (Orengo 2008). Cuando tratamos de hacer simulaciones que exploran el universo de plegamientos en la naturaleza, nos podemos topar con aproximados fuera de la realidad debido a la naturaleza continua de este sistema. Es por esto que para resolver el problema de la relación estructura función, sería muy útil entender los mecanismos mediante los cuales se ensamblan los motivos comunes a muchas estructuras tridimensionales de las proteínas, en otras palabras, entender el origen de los dominios proteicos.

## Origen y evolución de los dominios

Una teoría recurrente en estudios de evolución molecular es que los dominios proteicos pudieron haber evolucionado mediante la fusión y recombinación de un grupo de péptidos ancestrales. En consecuencia, los dominios estructurales actuales contendrían ladrillos constructores homólogos. Sin embargo, las herramientas que nos permiten hacer inferencias filogenéticas no son capaces de detectar homología en fragmentos muy pequeños. Teichmann y sus colegas han sugerido que la evolución crea funciones nuevas al combinar los dominios existentes generando diversidad estructural (Vogel 2004).

La duplicación acompañada de fusión es también esencial en la generación de proteínas nuevas. Los mecanismos implicados en estos procesos comprenden recombinación homóloga y no homóloga, permutaciones circulares, y barajeo de dominios (Soding 2003). Este conjunto de evidencias demuestra que la duplicación y fusión de superestructuras secundarias es uno de los mecanismos principales que ha conducido la evolución de los dominios proteicos. En la búsqueda de una clasificación más natural de las proteínas, se han hecho esfuerzos para desarrollar un diccionario de elementos de estructura secundaria, encontrando muchas evidencias de que la topología que adoptan ciertos motivos estructurales, por debajo del nivel de dominio, aparece recurrentemente entre diversas familias proteicas (Lupas 2007).

Existen múltiples evidencias del posible origen de los dominios proteicos a partir de la repetición de motivos más pequeños. Por un lado, encontramos casos interesantes como la proteína fibrosa colágena, que está compuesta por tres hélices tipo II arregladas de forma paralela, cada una formada por cientos de repeticiones del motivo prolina-hidroxiprolina-Glicina-X-Y (Lupas 2003). De la misma manera que repeticiones de motivos forman proteínas fibrosas, también identificamos que estructuras globulares se estructuran a partir del uso recurrente de elementos de estructura secundaria ( $\beta\alpha\beta$ ) en barriles TIM y ( $\beta\beta$ ) en propelas beta (Fulop 1999). Ahora bien, si analizamos la complejidad que involucra la generación de *novo* de un dominio de 100 residuos de extensión a partir de 20 aminoácidos, encontramos que el número de secuencias que se podrían construir es inmenso, aunque muy pocas podrían adoptar una estructura estable en el espacio y mucho menos presentar alguna actividad biológica.

En contraste, el uso periódico de estructuras secundarias previamente optimizadas, resulta más factible bajo las capacidades de síntesis presentes de los sistemas biológicos. También identificamos un número cada vez mayor de fragmentos pertenecientes a proteínas no homólogas que son similares en secuencia y estructura, como es el caso del motivo de unión a RNA (KH). El análisis estructural permite descubrir que este motivo es homólogo en dos plegamientos topológicamente disimilares (Grishin 2001).

Con base en estas evidencias, es posible decir que la evolución de proteínas es un proceso que continuamente está generando nuevos polipéptidos. En situaciones muy raras encontramos que nuevas macromoléculas se generan a partir del ensamblaje de secuencias de ADN no codificantes. Generalmente nuevas proteínas surgen mediante los procesos de duplicación, mezcla y fusión de dominios o superestructuras secundarias, elementos que ya habían sido seleccionados para adoptar alguna estructura en el espacio. Es muy difícil suponer que los dominios por sí mismos sean el resultado de una invención de *novo*; es más probable que la mayor parte de ellos pertenecieran a un grupo tempranamente establecido que ha ido cambiando a lo largo de la evolución de las especies.

## Antecedentes

Al entender la historia evolutiva de las proteínas podremos acercarnos al diseño de polipéptidos con funciones biológicas nuevas (Gerlt 2001). La gran diversidad de estructuras tridimensionales que tenemos hoy en día, ha surgido mediante una combinación de mecanismos que frecuentemente incluyen duplicaciones génicas seguidas de mutación y selección (Hughes 2005), como se ha propuesto que es el caso de algunas enzimas que adoptan el súper-plegamiento de barril  $(\beta/\alpha)_8$ .

El plegamiento de barril  $(\beta/\alpha)_8$  conocido también como barril TIM, es el más común en las bases de datos de estructuras de proteínas. También es visto en muchas superfamilias proteicas diferentes, catalizando reacciones totalmente disimilares. Este plegamiento está constituido por 8 hojas beta rodeadas por 8 hélices alfa. En un barril cerrado es posible encontrar un anillo completo de interacciones electrostáticas; cada hoja beta está formando enlaces de hidrógeno con sus dos vecinas y todas están conectadas. El número de hojas beta varía, pueden ser desde cuatro hasta veinte; estas hojas se encuentran orientadas de forma antiparalela, paralela o mezclada.

El barril TIM clásico consta de ocho repeticiones de unidades  $\beta/\alpha$ , en el cual cada hoja beta en el interior está cubierta por una hélice alfa en el exterior. Este plegamiento se describió por primera vez en la Triosa Fosfato Isomerasa, la enzima encargada de convertir gliceraldehido-3-fosfato en dihidroxiacetona fosfato, dentro de la vía glucolítica. Los sitios activos de todas las enzimas con el plegamiento de barril TIM se encuentran en los extremos carboxilo terminales de las hojas beta, esto implica que la geometría de estas regiones catalíticas está determinada por las ocho asas que siguen después de las betas.

Las hélices alfa y las hojas beta están conectadas por asas que, nombradas en función de su posición, pueden ser asas  $\beta/\alpha$  o  $\alpha/\beta$ . Las  $\beta/\alpha$  pueden ser de tamaño variable pero tienden a ser más largas que las  $\alpha/\beta$  como en la Triosa Fosfato Isomerasa (Nagano 2002). Mediante la modelación y el diseño de estas asas en el extremo carboxilo terminal de cada elemento  $\beta/\alpha$  se han obtenido evidencias de los papeles que juegan ambos tipos de estructuras. Ha sido posible demostrar que las asas que no están implicadas en la catálisis, están jugando un papel fundamental al mantener la estabilidad estructural (Vega 2003). Estas características permiten intercambiar las asas del carboxilo terminal sin comprometer la estabilidad del plegamiento al realizar experimentos de ingeniería de proteínas en la búsqueda de catálisis no naturales.

Las proteínas que presentan el plegamiento de barril TIM son comúnmente enzimas o están relacionadas con ellas. Las actividades de estas proteínas están ligadas a reacciones en el metabolismo energético o de biosíntesis, funciones consideradas entre las más antiguas a lo largo de la evolución de la vida en nuestro planeta. Este plegamiento es de los más representados entre todos los tipos de organismos vivos y es el plegamiento que más veces encontramos en la mayoría de los genomas (Nagano 2002).

Existen 889 estructuras de barril TIM en el CATH (*Class Architecture Topology Homologous Superfamily*) versión 1.7, que representan 147 secuencias no idénticas agrupadas en muchas familias homólogas. Las actividades que presentan las enzimas con el plegamiento de barril TIM comprenden cinco de las seis catalogadas por la Comisión Internacional Enzimática. La única actividad que hasta el momento no ha sido relacionada con este tipo de estructura es la de ligasa.

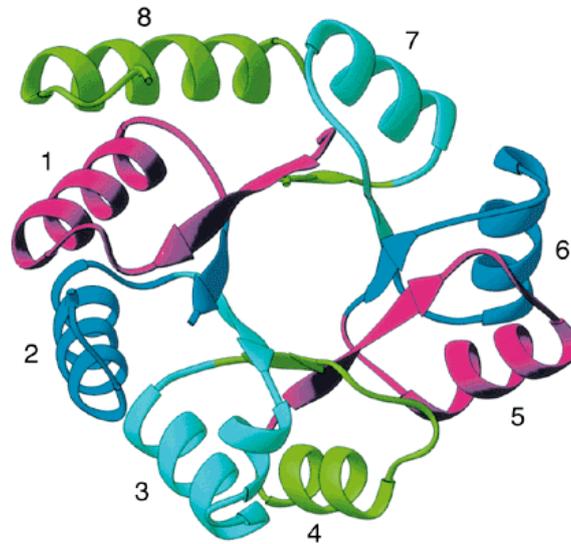


Figura 1. Plegamiento  $(\beta/\alpha)_8$

Estudios tanto de tipo estructural como de secuencia sobre dos proteínas que presentan el plegamiento de barril TIM, HisA y HisF, han demostrado que ambas enzimas, implicadas en la vía de síntesis de aminoácidos, surgieron por eventos de duplicación de genes .

El análisis estructural de este plegamiento en HisF y HisA, dos enzimas involucradas en la vía biosintética de histidina, indica una pseudo-simetría de dos ejes; soportando la hipótesis de que estos genes surgieron de la duplicación de un solo gen que codificaba para una proteína con estructura  $(\beta/\alpha)_4$  que posteriormente se fusionó y diversificó (Hocker 2001). Recientemente, Sterner en conjunto con su grupo de colaboradores realizó la demostración experimental de esta hipótesis al obtener proteínas solubles formadas por unidades  $(\beta/\alpha)_4$  HisFN y HisFC a partir de la disección de la proteína silvestre HisF (Hocker 2004). Cuando estos medios barriles se co-expresan se genera un polipéptido soluble.

Posteriormente, el mismo grupo de investigación logró la expresión de un homodímero estable formado por dos mitades carboxilo terminal de HisF (HisFC\*C), optimizando mediante mutagénesis sitiodirigida la interfase de dimerización entre las mitades idénticas.

Sin embargo, esta proteína sólo pudo ser producida en su forma insoluble y no se despliega cooperativamente en concentraciones bajas de urea, por lo que en un trabajo posterior la sometieron a rondas de mutagénesis al azar y selección, utilizando el sistema CAT como reportero de plegamiento (Seitz 2007).

En este trabajo logran estabilizar la mutante HisFC\*C con sólo algunos cambios obteniendo una proteína que presenta propiedades nativas de empacamiento y solubilidad. Finalmente, ellos logran la obtención de dos proteínas solubles formadas por las mitades amino terminal y carboxilo terminal de HisF y de HisA (HisFA y HisAF), postulando la existencia de un gen primitivo que codificaba para una proteína  $(\beta/\alpha)_4$ . Este conjunto de evidencias nos permite pensar que la duplicación y fusión de genes, seguida de un proceso de optimización natural por mutagénesis al azar, es suficiente para generar plegamientos estables.

Por otra parte, Shukla y Guptasarma (2004) obtuvieron barriles solubles  $(\beta/\alpha)_8$  mezclando las unidades  $(\beta/\alpha)$  de distintas maneras en el plegamiento (2-4-6-8-1-3-5-7), demostrando una autonomía en la subestructura del plegamiento y una plasticidad en las interacciones tanto hidrofóbicas, como de puentes de hidrógeno establecidas en esta estructura. Esta serie de evidencias nos permite inferir que las proteínas se pudieron haber ensamblado a partir de superestructuras secundarias repetitivas que intrínsecamente promueven la estabilidad a través de la recurrencia periódica de interacciones favorables. Tanto la ubicuidad como la diversidad de ciertas superestructuras, como son los elementos  $\beta\alpha\beta$ , podrían sugerir que éstos se han inventado varias veces a lo largo de la evolución.

Los mecanismos más comunes que están implicados en el cambio de los plegamientos a través del tiempo incluyen inserciones, deleciones o sustituciones de elementos de estructura secundaria, permutaciones circulares, invasiones por hojas beta e intercambio de dominios. Muchas estructuras han evolucionado partiendo de estados monoméricos hasta oligomerizarse. Se ha propuesto que la oligomerización es un mecanismo evolutivo que permite el cambio en la estructura cuaternaria de las proteínas, mediante la introducción de pequeños cambios en la secuencia de aminoácidos. Otros estudios sustentan la relevancia del uso recurrente de superestructuras secundarias en el ensamble de las proteínas actuales.

Búsquedas de unidades que sean capaces de plegarse de manera independiente (*foldons*) en grupos de proteínas no homólogas, generalmente identifican un grupo limitado de estructuras recurrentes (Soding 2003). La generación de proteínas quiméricas a partir de la recombinación de polipéptidos homólogos muestra que los fragmentos exitosamente mezclados o *schemas* (Voigt 2002) corresponden esencialmente a elementos conocidos de superestructuras secundarias.

Uno de los antecedentes más impactantes que demuestra la posibilidad de que nuevos dominios proteicos se pueden generar a través de combinatorias de segmentos polipeptídicos, independiente de homología es el desarrollado por Winter en el 2000. En este trabajo se fusiona la parte N-terminal de un dominio de barril  $\beta$  de una proteína de choque frío CspA, utilizándola como un bloque de construcción (*building block*), con fragmentos de ADN genómico de *E. coli* (Riechmann 2000). Posteriormente clonaron el repertorio de polipéptidos quiméricos y los desplegaron en fagos filamentosos tamizando proteínas plegadas por su resistencia a proteólisis con tripsina.

En este trabajo logran seleccionar un polipéptido que presentó el mismo plegamiento que la proteína parental y otro con un plegamiento distinto. La determinación cristalográfica de una de las proteínas quiméricas (1b11) revela un barril beta con seis cadenas que asemeja un tetrámero (de Bono 2005). Ellos demuestran que el barajeo de polipéptidos no homólogos puede generar proteínas plegadas con arquitecturas únicas, y que algunos segmentos pueden actuar como templados alrededor de los cuales el resto de la proteína se pliega [concepto relacionado con la hipótesis de los atractores (Holm 1996)].

En un trabajo posterior (Riechmann 2006), Winter utiliza como modelo un segmento proteico de la lisozima de huevo, que en su condición nativa forma la interfaz de unión con los anticuerpos monoclonales HyHEL5 y F10 ( $K_D = 60$  pM). El segmento expresado de manera aislada no es capaz de plegarse, se degrada rápidamente y se une muy débilmente a los anticuerpos antes mencionados. Ellos utilizan este segmento para fusionarlo con secuencias genómicas al azar y crear un repertorio de polipéptidos quiméricos. Este repertorio es tamizado utilizando despliegue en fagos y anticuerpos anti lisozima buscando proteínas plegadas y activas. Con esta aproximación experimental son capaces de aislar una proteína quimérica plegada que une los anticuerpos con una gran afinidad ( $K_D = 400$  pM). Inesperadamente el dímero incorporó un grupo hemo que aumenta su estabilidad.

La tasa de éxito obtenida por Winter fue de uno en  $10^7$  eventos en el experimento de la proteína de choque frío, lo que significa que dados los enormes tamaños poblacionales, el tiempo generacional y los niveles de recombinación intracelular en los filotipos bacterianos permitirían la generación de decenas de miles de proteínas al año a través de la recombinación de proteínas no homólogas. Aunque sólo una pequeña fracción de estos nuevos polipéptidos se pudiera establecer en la población, el efecto acumulativo en tiempos evolutivos sería significativo, como lo demuestra el hecho de que la secuenciación del genoma completo de la arquea *Aeropyrum pernix* generó un total de 1,500 ORFs (58 % del total del contenido genético) imposibles de reconocer haciendo uso de herramientas bioinformáticas en ningún otro organismo (Doolittle 2002). De hecho se ha postulado que la mayoría de los plegamientos únicos (*unifolds*) revelados en proyectos de secuenciación masiva podrían haberse originado mediante este mecanismo.

Teniendo como marco teórico estas evidencias del posible origen de los dominios proteicos, nosotros desarrollamos una estrategia para recuperar proteínas plegadas fusionando una unidad  $(\beta/\alpha)_4$  con un banco de secuencias genómicas al azar, valiéndonos de CAT (cloramfenicol acetyl transferasa) como sistema reportero de plegamiento.

Muchas veces es posible establecer una asociación directa entre la solubilidad de una proteína y la estabilidad de su plegamiento; en este contexto encontramos descritos en la literatura algunos métodos que nos permiten seleccionar proteínas plegadas. Las fusiones traduccionales son estrategias comúnmente utilizadas en laboratorios de evolución dirigida, ya que permiten tamizar bancos de variantes en función de algún fenotipo que presentan las bacterias que se encuentran expresando las fusiones. El método que utiliza CAT, enzima que confiere resistencia al cloramfenicol a bacterias que la expresan, fue descrito por primera vez en un trabajo de Davidson (Maxwell 1999).

En esta investigación se observa que células que expresan la fusión de un gen que codifica para una proteína blanco insoluble, en el extremo amino terminal del gen de CAT, resisten una cantidad mucho menor de cloramfenicol, que células que expresan fusiones con proteínas solubles. La lógica detrás de este resultado indica que la expresión de proteínas insolubles en el extremo amino terminal de CAT, evita que la enzima adquiera una conformación estable activa, propiciando un decremento en su actividad que se ve reflejado en la resistencia al cloramfenicol que presentan las bacterias transformadas con la fusión.

Por otra parte, la formación del cromóforo de la proteína verde fluorescente (GFP) depende del plegamiento correcto de la misma. Teniendo esta información como marco teórico, Geoffrey en 1999 desarrolla un trabajo en el cual el gen que codifica para una proteína de prueba, es fusionado en el extremo amino terminal del gen que codifica para la GFP. Ellos demuestran que la fluorescencia presente en células de *Escherichia coli* (que se encuentran expresando dicha fusión) está relacionada con el plegamiento correcto de la proteína de prueba que se fusionó a GFP. Estos métodos son ampliamente utilizados en experimentos de evolución dirigida, ya que permiten seleccionar a partir de bancos de variantes, proteínas que sean capaces de adoptar una estructura estable.

La proteína que utilizamos como andamiaje para los experimentos de complementación, es una proteína sintética que fue construida en un trabajo previo en este laboratorio. Este andamiaje es potencialmente termoestable, ya que fue construido utilizando una metodología que se fundamenta en definir una secuencia consenso a partir de un alineamiento múltiple de secuencias de proteínas ortólogas; en este caso PRAI (fosforibosil antranilato isomerasa). Una vez que se diseñó la secuencia se procedió a realizar la síntesis del gen completo. Posteriormente, se expresó de la proteína y se hizo una caracterización parcial de sus parámetros biofísicos.

## **Hipótesis**

Es posible generar estructuras proteicas plegadas mediante la fusión de una unidad  $(\beta/\alpha)_4$  con secuencias genómicas al azar.

## **Objetivos**

### **General**

Fusionar un sub dominio proteico  $(\beta/\alpha)_4$  a un banco de secuencias genómicas al azar, para obtener proteínas plegadas.

### **Particulares**

- Construir un modelo teórico de la proteína que se utilizará como andamiaje.
- Disectar la proteína para obtener un templado que nos permita recuperar proteínas solubles.
- Fragmentar ADN genómico de *Escherichia coli* para obtener secuencias al azar.
- Generar un banco de proteínas quiméricas fusionando las secuencias genómicas con la unidad  $(\beta/\alpha)_4$ .
- Seleccionar proteínas plegadas utilizando como sistema reportero de plegamiento fusiones traduccionales con la cloramfenicol acetil transferasa.
- Identificar los fragmentos insertados mediante métodos bioinformáticos.
- Caracterizar la solubilidad de las nuevas proteínas quiméricas

## Metodología General

1. Generación de un modelo teórico de la proteína que utilizaremos como andamiaje, para analizar un sitio de corte que nos permita disectarla en dos unidades  $(\beta/\alpha)_4$ .
2. Diseño de sitios de restricción que nos permitan partir el barril en dos mitades.
3. Clonación del barril en el vector apropiado para generar la fusión traduccional con el reportero de plegamiento (CAT).
4. Introducción de los sitios de restricción en la secuencia de la proteína consenso, utilizando la técnica megaprimer.
5. Digestión de la proteína consenso generando dos fusiones traduccionales distintas (dependiendo de la mitad que fue removida) con el reportero de plegamiento: N-PRAIcons- $(\beta/\alpha)_4$ -CAT (conservando el extremo amino terminal) y C-PRAIcons- $(\beta/\alpha)_4$ -CAT (conservando el extremo carboxilo terminal). La unidad  $(\beta/\alpha)_4$  que utilizaremos como andamiaje a complementar, deberá ser incapaz de plegarse por sí sola; de esta manera CAT tampoco podrá plegarse eliminando clonas autoligadas con el sistema de selección.
6. Purificación de ADN genómico de *Escherichia coli*.
7. Fraccionamiento del ADN genómico hasta obtener fragmentos de un tamaño aproximado de la mitad del barril que fue removida.
8. Preparación del vector con la construcción y los fragmentos a clonar.
9. Ligación de los fragmentos con el vector (generación del banco de variantes).
10. Selección de clonas resistentes a cloramfenicol. En teoría deberían ser variantes que presentan insertos capaces de complementar el plegamiento de la unidad  $(\beta/\alpha)_4$ .
11. Identificación y análisis bioinformático de los fragmentos que complementaron el medio barril.
12. Expresión de las nuevas proteínas quiméricas fusionadas a CAT (reportero de plegamiento).
13. Análisis bioquímico de las quimeras fusionadas a CAT.
14. Remoción del reportero de la fusión permitiendo la expresión independiente de las quimeras.
15. Análisis de la solubilidad de las quimeras.

## Métodos y Materiales

### Bioinformática

En el diseño del corte del barril se utilizaron los programas *swiss model* con los parámetros *default* para generar un modelo de la proteína consenso. El modelo fue minimizado utilizando *NAMD*. Para predecir la estructura secundaria del barril se utilizaron *Predictprotein* y *Psipred* (Rost 2004) con los parámetros *default*. Para diseñar los sitios de restricción necesarios para disectar el barril en dos mitades se utilizó *DNA strider*. En el análisis de las clonas se hicieron búsquedas con *BLAST* vs la base de datos de proteínas no redundantes del *NCBI* (National Center for Biotechnology Information) para identificar los fragmentos insertados. También se realizaron búsquedas contra el *PDB* para identificar estructuras reportadas de proteínas homólogas a las caracterizadas como fuente de los insertos que complementaron el plegamiento.

### Biología Molecular

La **extracción de plásmido** es una metodología muy frecuente en laboratorios que utilizan sistemas de expresión recombinantes. La lisis alcalina libera el plásmido de la bacteria y la ARNasa elimina el ARN del lisado bacteriano. En presencia de sales caotrópicas (cloruro de guanidino) el ADN plasmídico se une selectivamente a una resina especial y permanece de esa manera mientras se somete a una serie de lavados que remueven residuos bacterianos. Finalmente, se eluye con una solución baja en sales. El proceso no requiere una precipitación del material genético, extracciones orgánicas agresivas o una manipulación excesiva del mismo. Frecuentemente se usa el *High Pure Plasmid Isolation Kit* distribuido comercialmente por la marca *Roche*. El procedimiento completo llevado a cabo cada vez que se purificó plásmido a lo largo de este trabajo, es como sigue: se inoculó una colonia aislada en 3 ml de medio LB con espectinomicina 100 µg/ml y se incubó durante 16 horas. Posteriormente se colectaron células por centrifugación y se purificó de la manera ya descrita. En experimentos de extracción de bandas de plásmido digerido a partir de geles de agarosa se utilizó el kit de *Qiagen* (*QIAquick Gel Extraction Kit*).

Durante la realización de este trabajo fue necesario **purificar los productos de PCR** que se sintetizaron para eliminar residuos de amortiguadores (principalmente sales) y reactivos, entre otros contaminantes. Esta purificación se basa en los mismos fundamentos teóricos de la purificación de plásmido, y se utilizaron soluciones del kit *High Pure PCR Product Purification* también de *Roche*.

Para introducir los sitios de restricción que permitieron disectar el barril en dos mitades se siguió la estrategia denominada **megaprimer**. Esta metodología involucra dos reacciones de *PCR* con un oligonucleótido mutagénico que introduce los sitios de restricción requeridos. En el esquema de la figura 2 podemos encontrar una descripción detallada de la metodología.

## Introducción de sitios de restricción con megaprimer

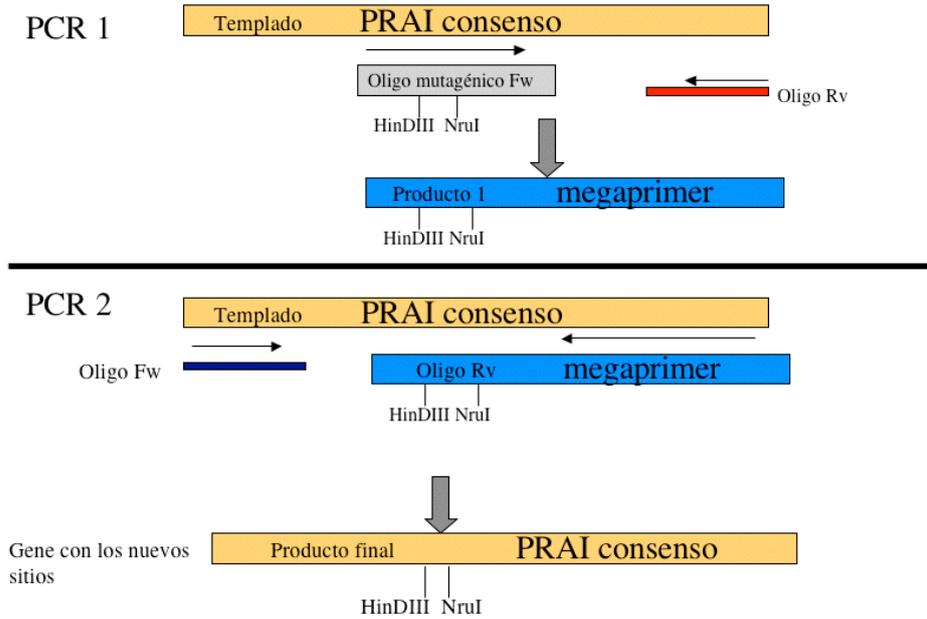


Figura 2. Estrategia de megaprimer. En dos reacciones de *PCR* se obtiene el gen con los nuevos sitios de restricción, la estrategia exige utilizar una cantidad muy pequeña de templado en ambas reacciones.

Para realizar las **digestiones de plásmido** se utilizaron endonucleasas de restricción (v. gr. HindIII o BamHI), estas enzimas reconocen secuencias palindrómicas específicas y cortan enlaces fosfodiéster en cada una de las cadenas de la secuencia nucleotídica. Después de la digestión con estas herramientas moleculares es posible separar los fragmentos resultantes mediante una electroforesis en gel de agarosa. Cada una de las enzimas de restricción requiere condiciones de reacción específicas para lograr una actividad óptima. Entre estas condiciones de reacción se encuentran: concentración de sales, pH y temperatura.

Los **plásmidos** utilizados en este trabajo son: **a) pT4** (construido en el laboratorio del Dr. Hilvert, del ETH en Suiza, donado por el Dr. Joel Osuna) vector de expresión en *E. coli* de 1845 pares de bases que contiene el marcador de resistencia a kanamicina y origen de replicación pUC. Además es de alto número de copias y presenta un promotor ptrc y **b) pBC-blaPZ –CAT** de 6644 pares de bases, alto número de copias, origen de replicación de *Escherichia coli* y promotor lac derivado de un fagémido PBC SK menos de *Stratagene*, que contiene un marcador de resistencia para espectinomina y estreptomina. En este plásmido se clonó el gen que codifica para la proteína reportera de plegamiento CAT (cloramfenicol acetil transferasa) entre los sitios de restricción BamHI y BglII.

Las enzimas de restricción utilizadas fueron: BamHI, NdeI, HindIII, NruI y BglII. También se usó BSA 10 mg/ml en las reacciones de digestión. Todas las enzimas son de *Biolabs*.

<i>IX NEB 2</i>	<i>IX NEB 3</i>	<i>IX NEB 4</i>
50 mM NaCl	100 mM NaCl	50 mM Acetato de potasio
10 mM Tris-HCl	50 mM Tris-HCl	20 mM Tris-acetato
10 mM MgCl <sub>2</sub>	10 mM MgCl <sub>2</sub>	10 mM Acetato de magnesio
1 mM DTT	1 mM DTT	1 mM Dtt
pH 7.9 @ 25° C	pH 7.9 @ 25° C	pH 7.9 @ 25° C

Tabla 1. Amortiguadores utilizados en las reacciones.

El procedimiento generalizado para realizar las digestiones en este trabajo fue el siguiente:

- 30 µl Vector
- 5 µl BSA (casi siempre empleamos BSA en las reacciones de digestión)
- 5 µl Amortiguador recomendado por el catálogo de *Biolabs* para la digestión doble.
- 1.5 µl Enzima 1
- 1.5 µl Enzima 2
- 7 µl H<sub>2</sub>O

Se incubaron por 120 minutos a 37 °C. Al finalizar la digestión se inactivaron las enzimas agregando EDTA hasta alcanzar una concentración final de 20 mM. Cuando no era posible montar una digestión doble, se realizó secuencial purificando el primer producto con el *kit* utilizado para purificar *PCR*. Las enzimas empleadas en cada paso del experimento se detallan en la tabla 2.

Enzimas usadas	Fase del procedimiento
Subclonación de PRAI consenso	Bam HI y NdeI
Remoción del extremo carboxilo terminal del gen que codifica para PRAI consenso	Bam HI y HindIII
Remoción del extremo amino terminal del gen que codifica para PRAI consenso	HindIII y NdeI
Remoción del reportero de plegamiento de las clonas obtenidas en el banco de variantes	BamHI y BglII

Tabla 2. Enzimas utilizadas en el experimento.

Otra técnica empleada frecuentemente fue la **electroforesis en gel de agarosa**. El fundamento teórico de esta metodología es el siguiente: el gel de agarosa consiste en un polisacárido que hace las veces de una matriz que permite separar moléculas en función de su tamaño. El ADN está cargado negativamente debido a los grupos fosfatos que presenta. En función de esta carga esta biomolécula se mueve en dirección al electrodo positivo (ánodo) a una tasa que es inversamente proporcional a su peso molecular. Moléculas grandes se moverán más lento que moléculas pequeñas. La tasa de migración del ADN también es proporcional a la concentración de agarosa con la que se haga el gel; porcentajes bajos de agarosa se utilizaron para resolver moléculas grandes y porcentajes altos para moléculas pequeñas.

Se emplearon geles de agarosa grado ADN (*Biorad*) al 1%, 1.5%, y 2% en solución amortiguadora TAE (Sambrook 2000). Para obtener el material a fragmentar se realizaron **extracciones de ADN genómico** mediante fenol cloroformo a partir de cultivos de *Escherichia coli*. En primera instancia se usó un amortiguador de lisis que rompe las células para liberar el ADN genómico. Posteriormente se agrega proteinasa (usualmente proteinasa K) para digerir las proteínas asociadas al ADN y ARNasa para digerir el ARN contenido en el lisado. En el siguiente paso se separa el ADN genómico de los demás componentes celulares, aplicando una combinación de fenol-cloroformo. En un paso subsiguiente se lava el material genético con una solución de etanol (70%) y agua (30%) para remover la mayoría de sales que pudiesen encontrarse contaminándolo. Finalmente se recupera en agua o un amortiguador que evite su degradación.

Protocolo para la extracción de ADN genómico a partir de un cultivo saturado de *Escherichia coli* utilizando una mezcla de fenol-cloroformo:

Materiales utilizados:

Amortiguador TE  
SDS 10% (peso/volumen)  
Proteinasa K  
ARNasa  
Fenol  
Cloroformo  
Etanol 70% (volumen/volumen)  
Etanol 100% (volumen/volumen)

1. 3-4 ml de cultivo *overnight* (*Escherichia coli*).
2. Centrifugar a velocidad máxima y conservar el *pellet*.
3. Lavar con 400  $\mu$ l de amortiguador TE (10:1).
4. Centrifugar y conservar el *pellet*.
5. Resuspender en 400  $\mu$ l de amortiguador TE.
6. Adicionar:
  - 50  $\mu$ l de SDS 10%.
  - 50  $\mu$ l de Proteinasa K (20  $\mu$ g/ml)
  - 20  $\mu$ l de ARNasa (10  $\mu$ g/ml). Si está muy diluida poner 200  $\mu$ l
7. Incubar a 37° C durante 180 minutos, o hasta que la muestra se vea transparente.
8. Pasar por una jeringa de insulina de 2 a 3 veces.
9. Pasar a un tubo limpio y agregar 200  $\mu$ l de fenol y 200  $\mu$ l de cloroformo.
10. Pasar por vortex.
11. Centrifugar 2 minutos a velocidad máxima (13,000 rpm)
12. Verificar 2 fases visibles, tomar la fase acuosa (la de arriba) en donde se encuentra el ADN.
13. Poner el líquido en un tubo nuevo y repetir los pasos 9, 10, 11 y 12.
14. Poner la muestra en un tubo nuevo y agregar 400  $\mu$ l de cloroformo.
15. Vortex y centrifugar a velocidad máxima (13,000 rpm) por 2 minutos.
16. Pasar la fase acuosa a un tubo nuevo y agregar 1 ml de etanol al 100% frío.
17. Agitar suavemente el tubo, el ADN se hará visible (pelusa blanca).

18. Pescar el ADN con una pipeta *pasteur* sellada y doblada en forma de gancho.
19. Enjuagar el ADN en etanol al 70%.
20. Escorrir bien y resuspenderlo en 200  $\mu$ l de agua tetradesilada.

Nota: El etanol al 100% y al 70% deben estar fríos (-20 °C)

El protocolo de purificación de ADN genómico requiere preparar amortiguador **TE (Tris / EDTA)**. Este amortiguador es utilizado en biología molecular generalmente en procesos que involucran ADN o ARN. Contiene Tris, un amortiguador de pH común y EDTA que funciona para secuestrar cationes divalentes, como el  $Mg^{2+}$ . El propósito del amortiguador es proteger el ADN de degradación ya que las nucleasas requieren estos iones para degradarlo. El protocolo es el siguiente:

- \* 10 mM Tris, llevar a pH 7.5 con HCl
- \* 1 mM EDTA

Posteriormente a la purificación del ADN genómico se procedió a **estimar su concentración** de la siguiente manera:

- Usando celdas de cuarzo (volumen pequeño).
- Siempre utilizando una muestra blanco.
- Diluir la muestra (1/50 ADN genómico)
- Evaluar en el espectro, la medida debe estar entre 0.1 y 1.0 DO (densidad óptica a 595 nm).

$A_{260} = 260 = 1$	Concentración $\mu$ g/ml
<b>ADN doble cadena</b>	50
<b>ADN cadena sencilla</b>	33
<b>ARN cadena sencilla</b>	40
<b>Oligonucleótidos</b>	20 – 30

Tabla 3. Cuantificación de ADN.

Cuantificación de ADN

a) Concentración del ácido nucleico

Medir la absorbancia a 260 nm =  $Abs_{260}$

(50 ml/ml) (0.035) (100) = X

Conversión  $Abs_{260}$  Factor de dilución Concentración

b) Cantidad total de ácido nucleico

(concentración) (volumen de la muestra) = Cantidad en  $\mu$ g

c) Pureza de la muestra

$Abs_{260}/Abs_{280}$  debe ser entre 1.8 y 2.0 (para ADN)

Si existe una absorbancia muy fuerte en 280 nm, hay una contaminación por proteínas. Para **fragmentar el ADN genómico** se empleó la enzima **DNAsa I**. Esta enzima digiere el ADN al azar, teniendo como sustrato último tetranucleótidos. Generalmente se utiliza para degradar ADN contaminante durante la purificación de ARN, también para identificar secuencias de unión en el material genético o para generar librerías de variantes. Es posible regular su actividad manipulando las condiciones de reacción o diluyéndola hasta encontrar las condiciones adecuadas para obtener fragmentos del tamaño deseado. Existen algunos aspectos muy importantes que se deben tener en cuenta cuando se usa la enzima DNAsa I como método de fragmentación:

- Sin importar que la metodología se encuentre previamente estandarizada bajo ciertos parámetros, es importante montar una cinética de digestión.
- La cinética de digestión se deberá realizar aplicando diversas concentraciones de enzima y tomando alícuotas a diversos intervalos de tiempo, para encontrar la concentración ideal de la misma y el tiempo necesario para obtener fragmentos del tamaño requerido.
- El protocolo siguiente se encuentra estandarizado para obtener fragmentos de entre 100 y 400 pares de bases. Lo ideal es emplearlo como punto de partida para la cinética de digestión.

### **Mezcla de reacción**

6 µg	ADN genómico de <i>Escherichia coli</i> (previamente purificado)
5 µl	Amortiguador 10X para PCR-MgSO <sub>4</sub> (marca comercial <i>Biolabs</i> )
5 µl	BSA 10X ( <i>Biolabs</i> )
1 µl	MnCl <sub>2</sub> 100 mM
4 µl	DNAsa 5 µg/ µl
x µl	H <sub>2</sub> O tetradestilada estéril
50 µl	Volumen total de mezcla de reacción

Mezcla de reacción en un volumen 50 µl; incubar durante 10 minutos a 15° C y detener con 10 µl de EDTA 100 mM.

Nota: El ADN genómico debe estar puro, es importante que no esté contaminado con ARN. Si esto sucede, se podría suponer erróneamente que se ha alcanzado el tamaño óptimo de digestión; la contaminación con proteínas impide la fragmentación eficiente del mismo y su correcta cuantificación.

Para la **preparación y fosforilación de los fragmentos a clonar** aplicamos el sistema de *EPICENTRE Biotechnologies End-It DNA-Repair Kit*. Este funciona para convertir el ADN dañado, o con terminaciones 5' incompatibles y/o terminaciones 3' incompatibles, en ADN fosforilado en su terminación 5' y extremos terminales romos listos para clonar, en plásmidos, cósmidos, fósidos y algunos otros vectores de clonación. La transformación en extremos romos del ADN tratado se realiza explotando la actividad de 5'–3' polimerasa y la actividad 3'–5' exonucleasa de la T4 polimerasa.

El *kit* también incluye la T4 polinucleótido cinasa y ATP para la fosforilación de las terminaciones 5' del ADN de extremos romos, para lograr una mayor eficiencia en el subsiguiente proceso de ligación.

Contenido del *kit*:

Mezcla de enzimas T4 DNA Polimerasa T4 Polinucleótido cinasa	20µl
Amortiguador 10X (330 mM Tris acetato [pH 7.8], 660 mM acetato de potasio, 100 mM acetato de magnesio, 5 mM DTT)	100µl
ATP (10 mM)	100µl
dNTPs (2.5 mM de cada uno)	100µl

Almacenar el *kit* a  $-20^{\circ}\text{C}$

El protocolo se desarrolla de la siguiente manera: Primero es necesario purificar el ADN a reparar (ya sea mediante el uso de fenol cloroformo o con el *kit* de limpieza de *PCR* marca comercial *Roche*). Es posible eluirlo en amortiguador TE (10 mM Tris- HCl [pH 7.5] 1 mM EDTA) o en H<sub>2</sub>O. Utilizando un tubo de microcentrífuga, combinar los siguientes componentes para una reacción estándar. Un volumen de reacción de 50 µl es suficiente para reparar hasta 5 µg de ADN. La mezcla estándar es como sigue:

34 µl DNA a ser reparado (hasta 5 µg)
5 µl 10X Amortiguador
5 µl Mezcla 2.5 mM dNTPs
x µl H <sub>2</sub> O tetradestilada estéril
1 µl Mezcla de enzimas
50 µl Volumen total de mezcla de reacción

Incubar a temperatura ambiente por 45 minutos. Para detener la reacción calentar el tubo a  $70^{\circ}\text{C}$  por 10 minutos.

**Nota:** Aun después de haber incubado a  $70^{\circ}\text{C}$  por 10 minutos es posible que la T4 polinucleótido cinasa no haya quedado completamente desactivada, lo que conlleva en un alto porcentaje de clonas no recombinantes debido a la fosforilación en el extremo 5' y autoligación del vector de clonación durante la ligación de ADN. Para reducir el porcentaje de autoligación podría ser necesario realizar una extracción con fenol cloroformo y precipitar con etanol antes de montar la reacción de ligación. El ADN reparado puede ser utilizado para ligar sin ser purificado previamente.

La reacción de **ligación** implica combinar fragmentos de ADN provenientes de diferentes fuentes en una molécula nueva. En este trabajo ligamos los fragmentos obtenidos mediante la fragmentación del ADN genómico de *Escherichia coli* (con la DNAsaI), con dos unidades  $(\alpha/\beta)_4$  diferentes (carboxi y amino) generadas al disectar el gen que codifica para la proteína PRAI consenso en dos mitades. Específicamente, la ligación involucra crear enlaces fosfodiéster entre el extremo 3' hidroxilo de un nucleótido con el extremo 5' fosfato de otro. La enzima usada para ligar fragmentos de ADN es la T4 ligasa proveniente del bacteriófago T4. Esta enzima es capaz de ligar tanto fragmentos romos, como cohesivos. Sin embargo, se recomienda emplear una concentración mayor de la enzima cuando se trata de fragmentos romos. Es importante notar que ésta es una de las reacciones más importantes cuando se está trabajando en la generación de un banco de variantes.

#### Procedimiento y materiales

Enzima T4 DNA ligasa de *Roche* 1 U/ $\mu$ l  
 Amortiguador 10X  
 660 mM Tris – HCl  
 50 mM MgCl<sub>2</sub>  
 50 mM DTT  
 10 mM ATP  
 pH 7.5 (20° C)

El ATP no es estable, concentraciones menores de esta molécula afectan en gran medida la eficiencia de ligación, es preferible hacer alícuotas del amortiguador de ligasa para descongelar sólo el necesario cada vez que se monte una reacción de ligación.

Componente	<i>Extremos cohesivos</i>	<i>Extremos romos</i>
<b>ADN templado</b>	Hasta 1 $\mu$ g de ADN digerido	Hasta 1 $\mu$ g de ADN digerido
<b>Amortiguador de ligación 10X</b>	3 $\mu$ l	3 $\mu$ l
<b>T4 ADN ligasa</b>	1 – 5 U	3- 5 U
<b>H<sub>2</sub>O</b>	Llevar hasta 30 $\mu$ l	Llevar hasta 30 $\mu$ l
<b>Incubación</b>	4-16 °C <i>overnight</i>	16- 25 °C <i>overnight</i>

Tabla 4. Mezcla de reacción para ligación.

Se utilizaron las siguientes cepas de *Escherichia coli*: a) para la expresión de las proteínas mutantes JM101, b) para la selección del banco XL1Blue y MC1061. También se prepararon células electrocompetentes de todas las cepas (Sambrook 2000).

Los antibióticos usados en soluciones stock (1000x) fueron los siguientes:

Kanamicina (kan) 25  $\mu$ g/ml.  
 Espectinomicina (sp) 100  $\mu$ g/ml.  
 Cloramfenicol (cm) 10, 15, 25 y 35  $\mu$ g/ml.  
 Ampicilina (ampi) 200  $\mu$ g/ml

Para determinar la **concentración de proteínas** totales en solución aplicamos el ensayo **Bradford** (marca comercial *Biorad*). Este sistema es sensible y rápido, aunque puede presentar alguna variabilidad de respuesta entre diferentes proteínas purificadas. Las proteínas usadas en este ensayo son desnaturalizadas irreversiblemente. El intervalo de sensibilidad comprende de 25 a 200  $\mu\text{g}$  de proteína en solución.

Las soluciones para el Bradford *stock* (estable) son:

100 ml 95% Etanol  
 200 ml 88% Ácido fosfórico  
 350 mg *Serva Blue G*

Para el Bradford de trabajo:

425 ml de agua destilada  
 15 ml 95% Etanol  
 30 ml 88% Ácido fosfórico  
 30 ml Bradford Stock

Se pasa a través de un filtro Whatman 1 y la curva patrón se mide a 595nm (DO).

<i>Proteína (<math>\mu\text{g}</math>)</i>	<i>Solución estándar 1 mg/ml de BSA</i>	<i>Amortiguador experimental Bradford (<math>\mu\text{l}</math>)</i>	<i>Bradford (<math>\mu\text{l}</math>)</i>
<b>0</b>	0	800	200
<b>2.5</b>	2.5	797.5	200
<b>5</b>	5	795	200
<b>10</b>	10	790	200
<b>15</b>	15	785	200
<b>20</b>	20	780	200

Tabla 5. Mezcla curva patrón Bradford.

<i>Muestra</i>	<i>Amortiguador experimental (<math>\mu\text{l}</math>)</i>	<i>Bradford (<math>\mu\text{l}</math>)</i>
<b>20</b>	780	200
<b>50</b>	750	200

Tabla 6. Mezcla para leer la muestra.

Para analizar las muestras de proteínas totales se hicieron electroforesis en geles de poliacrilamida bajo condiciones desnaturalizantes **SDS PAGE** (por sus siglas en inglés). El gel de poliacrilamida funciona como una matriz que permite separar las proteínas. El SDS (dodecil sulfato de sodio) confiere a estas biomoléculas una carga neta uniforme; en consecuencia, el único factor que va a determinar la migración diferencial de las especies proteicas será su tamaño.

<i>Gel 1.5mm</i>	<i>7.5%</i>	<i>10%</i>	<i>11.3%</i>	<i>12.5%</i>	<i>15%</i>	<i>20%</i>	<i>Stacking Gel</i>
<b>LGS</b>	8.0	8.0	8.0	8.0	8.0	8.0	-
<b>UGS</b>	-	-	-	-	-	-	3.75
<b>Acrilamida 30%</b>	8.0	10.6	12.5	13.4	16.0	21.3	2.25
<b>H<sub>2</sub>O</b>	16.0	13.4	11.5	10.6	8.0	2.7	9.0
<b>10% APS</b>	160µl	160µl	160µl	160µl	160µl	160µl	45µl
<b>TEMED</b>	15µl	15µl	15µl	15µl	15µl	15µl	15µl

Tabla 7. Mezcla utilizada para preparar geles de poli-acrilamida de 1.5 mm de grosor.

<i>Gel 0.75mm</i>	<i>7.5%</i>	<i>10%</i>	<i>11.3%</i>	<i>12.5%</i>	<i>15%</i>	<i>20%</i>	<i>Stacking Gel</i>
<b>LGS</b>	2.0	2.0	2.0	2.0	2.0	2.0	-
<b>UGS</b>	-	-	-	-	-	-	1.25
<b>Acrilamida 30%</b>	2.0	2.65	3.125	3.35	4.0	5.0	0.75
<b>H<sub>2</sub>O</b>	4.0	3.35	2.875	2.65	2.0	1.0	3.0
<b>10% APS</b>	40µl	40µl	40µl	40µl	40µl	40µl	15µl
<b>TEMED</b>	4µl	4µl	4µl	4µl	4µl	4µl	5µl

Tabla 8. Mezcla utilizada para preparar geles de poli-acrilamida de 0.75 mm de grosor.

En el proceso de caracterización de las quimeras generadas en este trabajo, se realizaron inmunodetecciones (*Western Blot*). Para las pruebas de solubilidad de las quimeras fusionadas a CAT se utilizaron anticuerpos anti-CAT. En la caracterización de las quimeras expresadas independientemente del reportero, usamos anticuerpos anti-6xhis. El procedimiento completo que se siguió se enlista a continuación:

1. Inocular 3 ml de LB con una colonia aislada en el antibiótico del plásmido. Dejar crecer *overnight* (16 horas).
2. Inocular nuevamente 2 ml de LB con el antibiótico del vector con 500 µl del anterior *overnight*.
3. Crecer a 220 rpm 30° C *overday* evaluando la densidad óptica cada hora (600 nanómetros).
4. Inducir con 1mM de IPTG cuando el cultivo haya alcanzado una densidad óptica de 0.4 (600 nanómetros).
5. Detener el cultivo cuando haya alcanzado una densidad óptica de 0.9 (600 nanómetros).
6. Centrifugar durante 1 minuto a 14,000 rpm para colectar las células.
7. Resuspender en 500 µl de amortiguador PBS, añadir inhibidor de proteasas (PMSF).

8. Sonicar utilizando la micropunta en un tubo *ependorff* de 1.5 ml. Aplicar 3 rondas de doce pulsos a potencia 3. Dejar en hielo durante 2 minutos entre cada ronda.
9. Centrifugar durante diez minutos a 9,000 rpm en frío (6 °C) para obtener la fracción soluble.
10. Posteriormente a la obtención de la fracción soluble, tomar una muestra y agregarle amortiguador de carga (previamente adicionar beta-mercapto-etanol fresco al amortiguador de carga).
11. Calentar las muestras por 5 minutos a 95 °C.
12. Cargar las muestras en un gel de poliacrilamida al porcentaje necesario para obtener la resolución máxima, en función del tamaño de las proteínas esperadas.
13. Correr el gel a corriente constante máximo 10 mA para que lo haga correctamente.

#### **Método de transferencia semi-seco:**

1. Colocar sobre la cámara de transferencia 2 pedazos de papel *Whatman* húmedos de amortiguador ánodo 1.
2. Sobre estos pedazos de papel, colocar otros 2 pedazos de papel *Whatman* mojados con amortiguador ánodo 2.
3. Sobre ellos se dispone la membrana de nitrocelulosa húmeda con amortiguador ánodo 2, sobre la cual se ubica el gel de acrilamida y finalmente sobre éste se colocan otros cuatro pedazos de papel *Whatman*, mojados con amortiguador cátodo.
4. Se transfiere entre 60 y 90 minutos aplicando amperaje constante (0.8 mA /cm<sup>2</sup>).
5. Bloquear la membrana de nitrocelulosa *overnight* en 5% de leche y TBST.
6. Lavar 3 veces con TBST, 10 minutos cada lavado (utilizar el anticuerpo en la dilución adecuada en TBST al 0.1% de leche).
7. Incubar con el anticuerpo primario durante una hora (agitación ligera).
8. Lavar con TBST 3 veces 10 minutos cada vez (utilizar el anticuerpo en la dilución adecuada en TBST al 0.1% de leche).
9. Incubar con el anticuerpo secundario durante una hora.
10. Lavar 3 veces con TBST, 10 minutos cada vez.
11. Adicionar BCIP/NTB (sustrato para revelar) 1 ml de cada uno y llevar a 10 ml con agua miliQ.
12. Incubar en el sustrato hasta que se distingan las bandas del control.

13. Detener la reacción con EDTA/H<sub>2</sub>O, 5 mM o enjuagando bien con suficiente agua.

Nota: es posible que la concentración de proteína no sea suficiente para verla en el gel de acrilamida, en ese caso ajustar el volumen del cultivo al tamaño correcto. Cuantificar la concentración de proteína utilizando Bradford.

Una vez lograda la expresión de una proteína heteróloga es necesario separarla del resto de las proteínas del hospedero, es decir, **purificarla**. Obtener **proteínas puras** es una meta sumamente difícil en cualquier proyecto de investigación. Muchas de las veces la purificación de una proteína se convierte en el paso limitante de un trabajo científico. Existen muchas variantes en la purificación de proteínas. Los métodos de separación modernos recaen fuertemente en la separación cromatográfica. En todos ellos una mezcla de las sustancias a ser fraccionadas se solubiliza en un fluido líquido o gaseoso que toma el nombre de fase móvil. La solución resultante se filtra en una columna que consiste en una matriz sólida, conocida como fase estacionaria.

La interacción entre los solutos individuales y la matriz estacionaria resulta en un avance diferencial de los solutos en función de sus propiedades fisicoquímicas individuales. Una de las propiedades de los polipéptidos más utilizada en la separación cromatográfica es el tamaño. En las columnas de exclusión por tamaño las moléculas se separan en función de su tamaño y forma. La fase estacionaria en este tipo de columnas consiste en resinas de material esponjoso que contiene poros en un rango relativo de tamaños. Si una solución acuosa es pasada por una columna que presenta estos poros de diferentes tamaños, los solutos presentarán una elusión diferencial (Sambrook 2000).

Las nuevas técnicas de ADN recombinante permiten la construcción de proteínas fusionadas a marcas de afinidad específicas; éstas facilitan la purificación de la secuencia polipeptídica de interés a través de columnas cromatográficas especiales. La marca más común que se agrega a las proteínas que se pretenden purificar, es una serie de 6 residuos de histidinas. Esta marca le confiere gran afinidad por una matriz de níquel en un sistema de NI NTA (*Nitrilotriacetic acid-Níquel*). Este sistema está montado en sefarosa CL-6B. Las proteínas que presentan esta marca permanecen unidas al níquel, mientras que los demás polipéptidos que no están fusionados, pasan por la columna separándose de la proteína problema. Ésta se recupera en un gradiente creciente de imidazol al eluir las diferentes fracciones

## Resultados

### Diseño del corte de la proteína consenso.

La selección del andamiaje constituye el paso más importante dentro de toda la metodología de este trabajo. El plegamiento que utilizamos como modelo es un barril  $(\beta/\alpha)_8$ ; consta de ocho repeticiones de unidades  $(\beta/\alpha)$ . En la realización del experimento fue necesario definir con exactitud en dónde terminan las primeras cuatro unidades  $(\beta/\alpha)$  es decir, partir el barril en dos unidades  $(\beta/\alpha)_4$ . Para disectar la proteína en dos mitades construimos un modelo teórico, ya que aunque no contamos con un cristal de PRAI consenso, la identidad de su secuencia con PRAI de *Escherichia coli* (existe la estructura determinada) es de 75%, por lo que confiamos en la verosimilitud del modelo. Para este propósito usamos el servidor en línea *swiss-model* valiéndonos del PDB 1PII (resolución 2.0 Å) como templado.

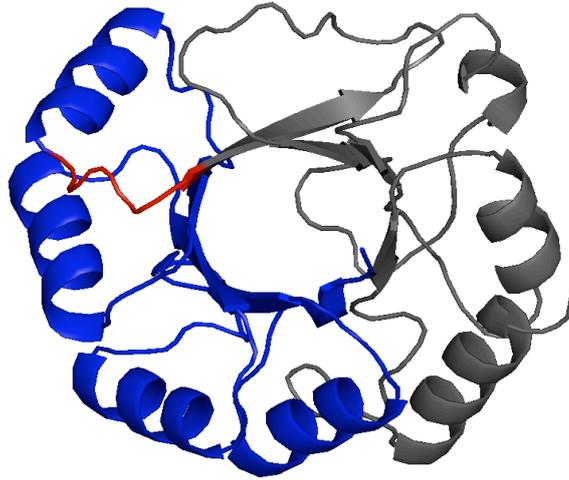


Figura 3. Modelo teórico de PRAI consenso generado con *swiss-model* (Torsten S. 1993). En azul se muestra la mitad amino terminal, en gris la mitad carboxilo terminal. Se introdujeron dos sitios de restricción, en el gen que codifica para PRAI consenso, cercanos a la mitad del barril. El sitio HindIII se encuentra al final de la cuarta hélice alfa (inicio del color rojo) y el sitio NruI se encuentra al principio de la quinta hoja beta (inicio del color gris).

En la figura 3 se muestra un modelo teórico de PRAI consenso, en color azul encontramos la mitad amino terminal y en color gris la mitad carboxilo terminal. En el modelo tridimensional se evaluó la posición correcta de los aminoácidos catalíticos así como los residuos de unión a ligando, con base en la información biológica obtenida a partir de PRAI de *Escherichia coli*. Posteriormente a la generación del modelo, realizamos la búsqueda de los posible sitios de restricción que se introducirían cerca de la mitad del barril. Para este propósito empleamos el programa de visualización y edición de secuencias *DNA Strider 1.4f6*. Encontramos que era posible introducir dos sitios de corte, para las enzimas HindIII y NruI. Decidimos introducir ambos sitios en el barril para contar con puntos alternativos de corte en las cercanías de la mitad del gen.

Nuestra precaución se debe a que contamos con algunos antecedentes (Soberon 2004) que confirman que las mitades del barril TIM son capaces de plegarse independientemente. En este caso nosotros necesitamos que la mitad que utilizemos como templado, sea incapaz de plegarse; al introducir dos sitios de restricción aumentamos nuestras probabilidades de lograrlo. El sitio *HindIII* se encuentra en el codón que codifica para el residuo 96L a la mitad del gen, justo al finalizar la secuencia de la cuarta hélice alfa; el sitio *NruI* esta un poco más adelante en la secuencia de nucleótidos, en el codón que codifica para el residuo 101I (en la figura 4 se muestra el sitio *HindIII* en la unión azul/rojo y el sitio *NruI* en la unión rojo/gris). Para introducir los sitios de corte se sintetizó un oligonucleótido mutagénico con los cambios necesarios y se siguió la estrategia megaprimer. En el apéndice de este trabajo se encuentran los detalles de las mitades generadas con cada una de las endonucleasas utilizadas.

## Resultados Biología Molecular

La disección del barril TIM en dos unidades  $(\beta/\alpha)_4$  se diseñó mediante el análisis de un modelo teórico, y estudiando un consenso generado a partir de las predicciones de estructura secundaria arrojadas por varios programas *on line*. La estrategia experimental involucró explorar qué sitios de restricción existían en la secuencia cercana a la zona de corte, o cuáles se podían incorporar de manera silenciosa (sin modificar la secuencia de aminoácidos). No fue posible encontrar sitios de restricción cerca de la zona seleccionada. Utilizamos el programa *DNA Strider 1.4f6* para buscar sitios silenciosos que insertar lo más cerca de la zona de digestión. El programa nos permitió diseñar la incorporación de los sitios de restricción *NruI* y *HindIII* a la mitad del gen del barril.

La proteína consenso se subclonó en el vector pBC – blaPZCAT utilizando los sitios de restricción *BamHI* y *NdeI*. El vector tiene las siguientes características: 6644 pb, alto número de copias, origen de replicación de *Escherichia coli* y promotor *lac*. El vector es un derivado de un fagémido PBC SK- de *Stratagene*, además contiene un *cassette* de resistencia para espectinomicina y otro para estreptomycinina. Entre los sitios *XbaI* y *NdeI* existe un sitio de unión a ribosoma (GAAGG).

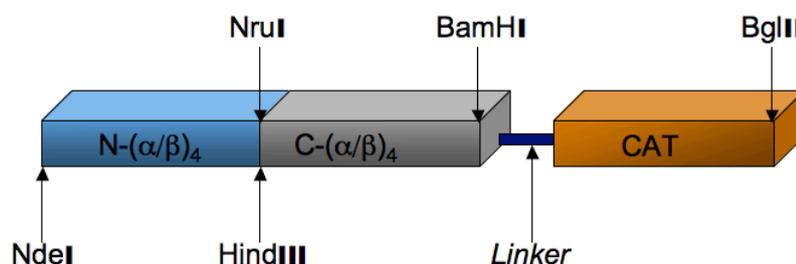


Figura 4. Mapa de restricción de la construcción. Se muestra el gen de la proteína PRAI consenso fusionado al gen del reportero de plegamiento CAT mediante un *linker*. Se muestran los dos sitios alternativos de restricción introducidos para diseccionar el barril en dos mitades *NruI* y *HindIII* (ver detalles en apéndice).

La inserción de los sitios se realizó aplicando la estrategia denominada megaprimer, descrita con detalle en los métodos de este trabajo. Decidimos introducir dos sitios de restricción para aumentar nuestras probabilidades de obtener una unidad  $(\beta/\alpha)_4$  incapaz de plegarse. De esta manera podíamos explorar dos tipos de digestiones distintas, si la mitad generada con NruI y BamHI era capaz de plegarse independientemente, cabía la posibilidad de utilizar HindIII y BamHI. Una descripción detallada de los sitios de digestión se encuentra en el apéndice de este trabajo. En la siguiente figura se describe la estrategia de generación del banco y las construcciones que podrían obtenerse en función de las enzimas utilizadas.

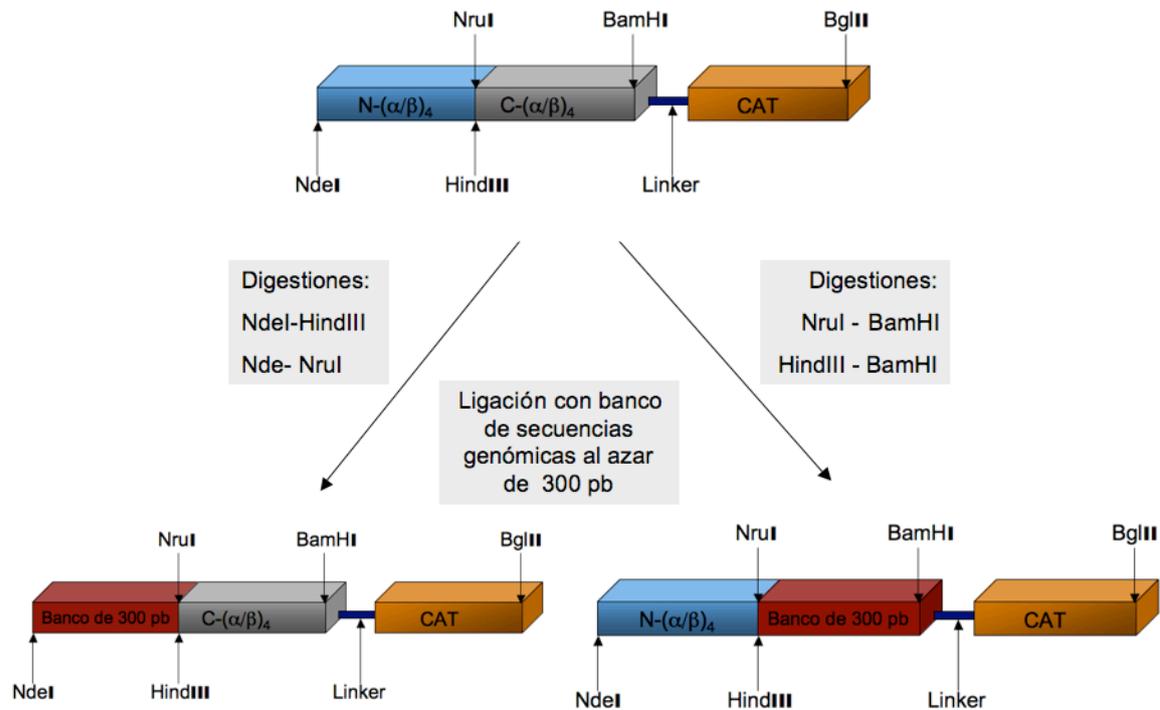


Figura 5. Esquema de la generación del banco de variantes: a) se digiere el gen de la proteína con las enzimas indicadas, b) se obtiene el ADN genómico a partir de un cultivo de *E. coli*, c) se fragmenta y se clona en el lugar de la unidad  $(\beta/\alpha)_4$  removida del barril.

Para lograr una descripción clara y coherente de los pasos a seguir en la construcción del banco de variantes, en el cual se seleccionaron las proteínas quiméricas generadas en este trabajo, hemos dividido los resultados en tres partes principales: a) preparación de los fragmentos a clonar para complementar el barril (se requerían fragmentos de 300 pares de bases, ya que la mitad removida del gene de PRAI consenso era de este tamaño), b) preparación del vector en donde clonamos los fragmentos y c) la estrategia seguida para montar la reacción de ligación y su posterior selección.

## Preparación de los fragmentos a clonar

Se estandarizó un protocolo de extracción de ADN genómico (descrito en la sección de métodos de este trabajo) que funcionó bastante bien, permitiéndonos obtener grandes cantidades de ADN para estandarizar mecanismos de fragmentación del mismo. El método requiere: a) inocular medio LB con una colonia aislada de *E. coli*, b) crecer el cultivo por 16 horas, y c) extraer el ADN genómico con fenol cloroformo. Después de estas extracciones usualmente fue necesario darle un tratamiento con RNAsa al genómico extraído, ya que el primer tratamiento con esta enzima parece que no da buenos resultados cuando el ADN no se encuentra purificado. Era necesario tratar con RNAsa el genómico purificado ya que frecuentemente estaba contaminado con ARN ribosomal, este contaminante no nos permitía monitorear el proceso de fragmentación al que era sometido.

Para la fragmentación del genómico probamos tres estrategias distintas. Primero tratamos con nebulización, una técnica que nos permite fragmentar el material genético imprimiéndole presión al colocarlo en un nebulizador y hacer pasar por él nitrógeno. En función de la cantidad de presión que ejerzamos al modular el tanque de nitrógeno, es posible obtener poblaciones de fragmentos de diversos tamaños.

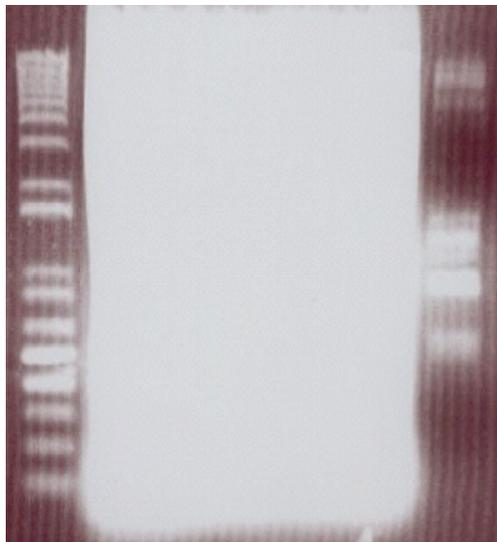


Figura 6. Electroforesis en gel de agarosa de ADN genómico nebulizado. Se muestra el ADN nebulizado teñido con bromuro de etidio flanqueado por marcadores de peso molecular.

En la figura 6 se aprecia el resultado de nebulizar ADN genómico. Se puede observar como se genera un gradiente de tamaños que recorren casi toda la superficie de resolución del gel de agarosa. La nebulización se realiza en amortiguador TE; en función de la presión ejercida en el ADN y el tiempo que ésta se mantenga, se obtienen fragmentos de diferentes tamaños. Al aumentar la presión y el tiempo la fragmentación mecánica del ADN se incrementa; sin embargo, existe un límite que no puede ser sobrepasado y si se ejerce demasiada presión sólo se logra dañar el sustrato. Este fue el primer método empleado para fragmentar el ADN genómico. Desafortunadamente los fragmentos generados nebulizando, no fueron útiles para obtener el banco de variantes.

El segundo método empleado para fragmentar el ADN fue la sonicación. Esta técnica mecánica de fragmentación rompe el ADN utilizando la misma fuerza sónica empleada para romper las células cuando se preparan fracciones totales. Se utiliza un sonicador normal, el sustrato a sonicar se puede mezclar con un amortiguador común o en agua. En función de la potencia del sonicador y del tiempo que la muestra sea sometida a la acción del mismo, se obtienen fragmentos de un tamaño determinado.

Realizamos algunas cinéticas de sonicatedo para determinar el tiempo ideal y la energía mecánica necesaria para obtener fragmentos de 100 – 400 pb. Una de las ventajas que este método tiene si lo comparamos con la nebulización, es que no es necesario precipitar el material genético fragmentado una vez que se ha sometido a sonicación. El procedimiento se puede realizar en agua tetradesilada por lo que no se necesita limpiar la muestra una vez terminada la fragmentación. Es un hecho bien establecido que la manipulación excesiva del ADN reduce su potencial para generar librerías de variantes.



Figura 7. Cinética de sonicatedo. Es posible apreciar cómo se va reduciendo el tamaño de los fragmentos hasta estabilizarse en un tamaño mínimo máximo. Carriles: 1) Marcador de peso molecular, 2) ADN genómico, 3) al 15) cinética de sonicatedo, cada uno de los carriles es un pulso de 10 segundos a potencia 2 en la misma muestra.

La cinética de sonicatedo reveló que 10 pulsos de 10 segundos a energía 2 eran suficientes para obtener fragmentos del tamaño deseado. Someter el ADN a más pulsos se traducía en un daño mecánico sin obtener una reducción significativa de su tamaño. De la misma manera que la nebulización, los fragmentos generados por este método no fueron útiles para nuestros propósitos. A pesar de que la nebulización y la sonicación son métodos frecuentemente utilizados para generar librerías de variantes en proyectos de secuenciación masiva, el uso de estas técnicas en nuestras condiciones de trabajo no dio buenos resultados. Es posible que las muestras se dañaran demasiado cuando aplicamos estos métodos de fragmentación, y aun con los métodos de reparación implementados, no logramos obtener fragmentos viables.

La digestión con DNAsaI, el último método implementado, dio buenos resultados. Esta enzima tiene la capacidad de digerir el ADN al azar (no muestra especificidad por alguna secuencia en particular), teniendo como sustrato último tetranucleótidos. Es posible regular su actividad al manipular las condiciones de reacción: concentración final de la enzima, temperatura, tiempo y tal vez lo más importante para nuestros fines, el tipo de catión que se agrega al amortiguador de reacción. En presencia de magnesio la enzima rompe las cadenas de ADN independientemente generando terminaciones de tipo saliente. Si el ion que se encuentra disponible en la mezcla de reacción es manganeso, la enzima rompe ambas cadenas simultáneamente, generando fragmentos con terminaciones romo, que son del tipo que necesitamos para la construcción del banco de variantes.

Las condiciones finales de reacción que resultaron óptimas para el trabajo son las que se detallan en los métodos. Es preciso mencionar que hay dos puntos críticos muy importantes en el procedimiento de digestión con DNAsaI. Tenemos que estar seguros de que la enzima ha quedado inactivada; para lo cual se recomienda utilizar calor y EDTA, además de purificar los fragmentos obtenidos utilizando columnas de afinidad. También es importante realizar una cinética nueva cada vez que se efectúa el experimento, ya que la concentración de la enzima puede variar durante su almacenaje (la concentración de enzima viable se reduce), la temperatura ambiental también influye en el resultado. Es posible emplear el protocolo descrito en este trabajo como punto de partida para estandarizar las condiciones de reacción cada vez que se necesite digerir el ADN genómico.

Una vez que se ha fragmentado el ADN genómico es preciso purificarlo. Para este propósito se realiza una electroforesis preparativa en gel de agarosa. Después que el ADN fragmentado ha migrado lo suficiente, se procede a cortar de banda la región que corresponde al tamaño deseado; para extraerlo de la agarosa se utiliza un *kit* común de extracción (en este trabajo se usó para ese fin el *High Pure PCR Product* de Roche). Posterior a la purificación, es necesario reparar y fosforilar los fragmentos para aumentar su viabilidad en la ligación.

Los mecanismos de fragmentación mecánica antes descritos dañan, generando mellas y extremos de cadena sencilla, en gran medida el ADN; no así el tratamiento con DNAsaI. Aunque teóricamente los fragmentos generados con este método enzimático deberían estar listos para clonarse, nosotros los sometimos a un proceso de reparación para aumentar su viabilidad, asegurándonos de obtener la mayor cantidad de fragmentos con extremos romos. No obstante que en teoría la enzima corta las dos cadenas simultáneamente, es posible que se generara una cantidad importante de fragmentos con extremos salientes.

A lo largo de la estrategia experimental se hicieron esfuerzos para clonar bancos con fragmentos generados mediante nebulización, sonicación y digestión con DNAsaI. Sin embargo muchas veces no dieron resultado, el fracaso en la obtención de los bancos se atribuyó en gran medida a la eficiencia de reparación de los fragmentos. Se utilizaron muchos protocolos distintos de reparación, no obstante éste fue uno de los puntos débiles de la metodología ya que nunca se obtuvieron los resultados esperados.

Las mezclas de reacción incluían enzimas como la T4 DNA polimerasa y Klenow, con ATP como fuente de energía para catalizar la incorporación de nucleótidos en los extremos salientes; se realizaron modificaciones en las concentraciones de las enzimas, en los tiempos de incubación, en la concentración de los nucleótidos y en los amortiguadores utilizados, sin obtener resultados satisfactorios.

Posteriormente se adquirió un *kit* especialmente diseñado para este propósito (*End-It DNA- Repair Kit*), que dio muy buenos resultados al reparar fragmentos obtenidos mediante digestión con DNAsaI. Los fragmentos que se lograron clonar son los reparados con este *kit*, que además también los fosforila para facilitar su ligación. Otra de las ventajas que encontramos al utilizar este *kit* es que no es necesario purificar los fragmentos antes de montar la reacción de ligación.

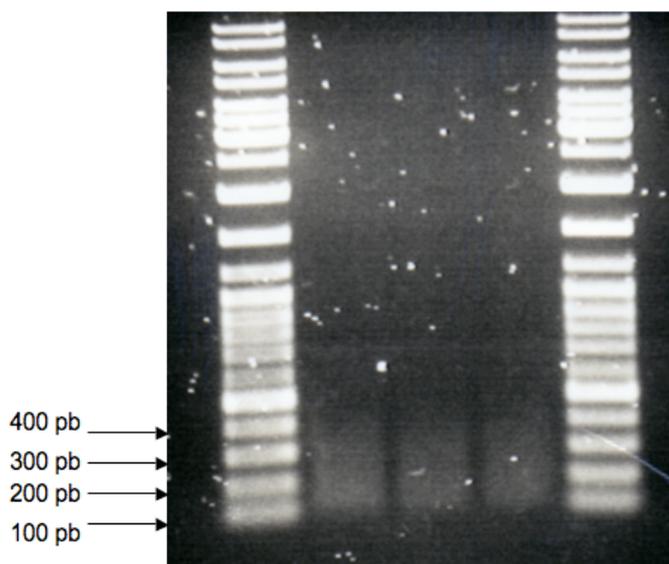


Figura 8. Se muestran fragmentos de ADN genómico de *E. coli* flanqueados por marcadores de peso molecular. Estos fragmentos se encuentran reparados, fosforilados y purificados, listos para la reacción de ligación.

En la figura anterior se muestra el aspecto de los fragmentos antes de utilizarlos en las reacciones de ligación. Como se puede apreciar usamos fragmentos de entre 100 y 400 pares de bases, para tratar de reconstituir el tamaño de la mitad del barril que se había removido ( $\approx 300$  pb). Es necesario decir que se trató de clonar el banco muchas veces y ninguna dio resultado hasta que se utilizó la DNAsaI como método de fragmentación, y el *kit End-It*. Una vez que hemos revisado el proceso de obtención de los fragmentos viables pasaremos a explicar el trabajo realizado con el vector.

### **Preparación del vector**

El vector podía ser digerido con cuatro combinaciones de enzimas distintas que permitían remover la mitad carboxilo terminal o la mitad amino terminal. Las digestiones NdeI – NruI y NdeI- HindIII remueven el extremo amino terminal. Las digestiones HindIII- BamHI y NruI-BamHI remueven el extremo carboxilo terminal.

La digestión con cualquier combinación de enzimas resultaba en extremos cohesivos en el vector; este tipo de extremos no es compatible con nuestra estrategia experimental ya que planeamos clonar fragmentos con extremos romos para evitar cualquier tipo de tendencia en las secuencias a insertar. Para transformar los extremos cohesivos en romos contábamos con dos opciones, digerir los extremos sobresalientes con una exonucleasa, o convertir la cadena sencilla en doble al incorporar los nucleótidos faltantes con una polimerasa.

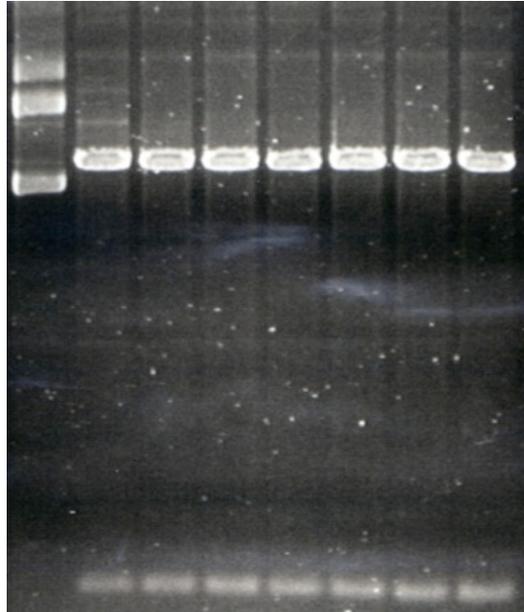


Figura 9. Cinética de digestión del vector con NruI - BamHI. En el carril uno se muestra el vector superenrollado. En los subsiguientes se muestra el progreso de la digestión tomando muestras cada hora. Se aprecia que en la hora tres el vector está completamente digerido.

Se realizaron cinéticas de digestión del vector para estimar el tiempo óptimo de incubación con las enzimas; posteriormente se purificó mediante una electroforesis en gel de agarosa y extracción de banda con el mismo *kit* empleado en la purificación de los fragmentos. Una vez que obtenemos el vector purificado lo sometemos al proceso de reparación usando el *End-It DNA-Repair Kit*. Se trató de defosfatar el vector para reducir el porcentaje de clonas autoligadas. Sin embargo, no se logró estandarizar el método degradando el vector durante el procedimiento, por lo que se decidió no defosfatarlo en reacciones subsiguientes.

De esta manera preparamos vector y fragmentos para montar la reacción de ligación y generar el banco de variantes que nos permitió seleccionar proteínas plegadas utilizando CAT como reportero de plegamiento.

## Ligación

Construir un banco de variantes es considerada una de las tareas más difíciles en biología molecular. La preparación de los insertos y del vector tienen que hacerse lo más rápido posible y bajo condiciones de extrema limpieza. El éxito en la clonación del banco está en función de muchas variables, desde que se inicia con la extracción del plásmido a digerir, hasta cuando se transforma la ligación en las células para realizar la selección de las variantes que nos interesan. A lo largo de todo el experimento se efectuaron muchos intentos de clonar el banco. Recordemos que existía la posibilidad de remover el extremo amino terminal para usar como templado el carboxilo terminal o viceversa, remover el extremo carboxilo terminal para usar como templado el amino terminal.

Decidimos remover el extremo carboxilo terminal para utilizar como templado la mitad amino [N( $\beta/\alpha$ )<sub>4</sub> prai cons-cat]. Mezclamos el vector digerido con las enzimas NruI – BamHI con los fragmentos en una proporción de 1:10, incubamos la reacción de ligación por 16 horas a 16 °C, purificamos la ligación utilizando una columna de afinidad y por último, la transformamos en células electrocompetentes *XL1Blue*. Recuperamos las células transformadas incubándolas en 1 ml de medio SOC por 1 hora a 37 °C. Plaqueamos 100  $\mu$ l en una placa de LB sólido con espectinomicina (para estimar número de transformantes), y plaqueamos los 900  $\mu$ l restantes del SOC en placas con cloramfenicol (20  $\mu$ g/ml). Todo este procedimiento se realizó de la misma manera con un control de autoligación (la misma mezcla de reacción de la ligación sin adicionar fragmentos).

Las placas seleccionadas en cloramfenicol de la ligación y de la autoligación se incubaron a 37 °C por 16 horas. Se obtuvieron colonias incontables en las placas de espectinomicina para la ligación y para el control de autoligación. En las placas seleccionadas en cloramfenicol obtuvimos un gran número de colonias, tanto para el control como para la ligación.

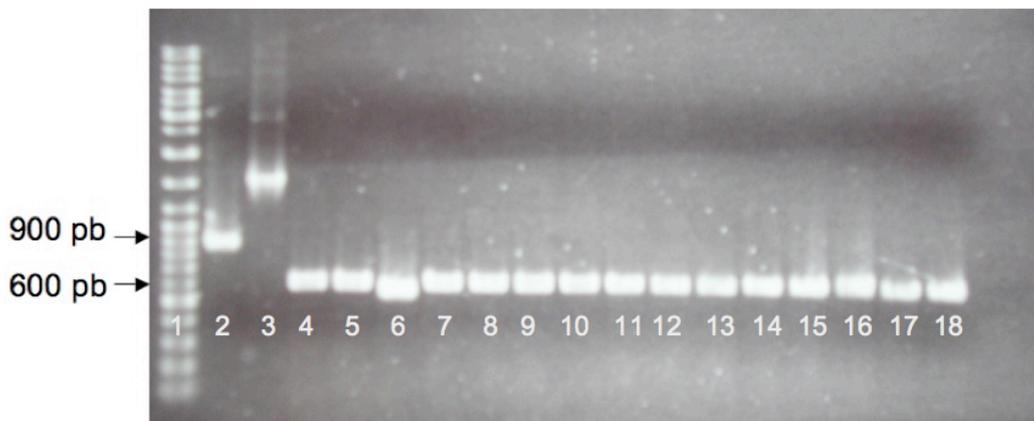


Figura 10. *PCR* de colonia de clonas resistentes a cloramfenicol del experimento de ligación. Carriles: 1) Marcador de peso molecular, 2) Control positivo (*PCR* de proteína consenso completa + un fragmento del gene de CAT), 3) plásmido superenrollado utilizado como templado para el control positivo, 4) al 18) *PCR* de las clonas picadas.

Se realizaron *PCR*'s de colonia para discriminar clonas autoligadas de clonas con inserto entre las colonias seleccionadas de la placa de la ligación. El *PCR* de colonia se diseñó utilizando un oligonucleótido *fw* que hibrida justo al inicio de la secuencia que codifica para la proteína consenso (en el sitio *NdeI*), y un oligonucleótido *rv* que hibrida en la secuencia que codifica para el reportero de plegamiento (CAT) después de 100 bases del inicio de la transcripción de la misma. De esta manera obtenemos un producto de 900 pares de bases si el templado del *PCR* corresponde a *prai* consenso completa + *cat*, y de 600 pares de bases si el templado corresponde a la unidad  $N-(\beta/\alpha)_4$  *prai* consenso + *cat* (autoligada). En la figura 10 podemos apreciar que todas las clonas picadas (carriles 4 al 16) son autoligadas. Se hicieron 90 *PCR*'s diagnóstico entre los cuales se encontraron 4 clonas sin digerir (plásmido superenrollado no digerido) y todas las demás resultaron ser autoligaciones.

El análisis realizado con los *PCR*'s de colonia arrojó dos resultados muy importantes: a) muchas de las colonias cloramfenicol resistentes habían sido transformadas con plásmido superenrollado que no se había digerido por completo, y b) que al remover la unidad  $C-(\beta/\alpha)_4$  digiriendo el gen del barril con las enzimas *NruI* – *BamHI*, la unidad  $N-(\beta/\alpha)_4$  restante es capaz de plegarse independientemente. Necesitamos una unidad  $(\beta/\alpha)_4$  incapaz de plegarse por sí sola para discriminar clonas autoligadas de las que tienen inserto, utilizando CAT como reportero de plegamiento. En consecuencia, procedimos a digerir el plásmido con las enzimas *HindIII* – *BamHI*, siguiendo el mismo procedimiento que se describió anteriormente montamos una reacción de ligación y selección de clonas. El resultado fue el mismo, la mitad amino terminal se puede plegar independientemente al digerir el barril con *HindIII* – *BamHI*. Una descripción detallada de las mitades generadas se encuentra en el apéndice de este trabajo.

El uso de cualquier mitad amino terminal de la proteína consenso como templado para recuperar proteínas plegadas no resultó viable, por lo que decidimos utilizar la mitad carboxilo terminal y clonar los fragmentos corriente arriba de la unidad  $C-(\beta/\alpha)_4$ . Para este propósito empleamos las enzimas *NdeI* y *HindIII*, realizamos cinéticas de digestión para encontrar el tiempo óptimo de incubación. Se procedió con la reacción de ligación bajo las siguientes condiciones:

Los fragmentos y el vector fueron reparados con el *End-It DNA-Repair Kit*, y purificados con el *Kit High Pure PCR Product*.

- a) Proporción 1:20 vector/inserto.
- b) 840 ng totales de ADN.
- c) 5 U de T4 DNA ligasa de *Roche*.
- d) 20 mM ATP.
- e) Volumen de reacción 120  $\mu$ l.
- f) 16 horas 16 ° C.

Nota: Las mismas condiciones para la reacción de autoligación (sin insertos).

Una vez terminado el tiempo de incubación de la ligación, se procedió a limpiarla con el *kit High Pure PCR Product*. Se electroporó en 4 viales de células electrocompetentes y se inoculó en placas de LB con cloramfenicol y espectinomicina, por separado. Al transformar esta ligación, implementamos un control de transformación que consistía en electroporar vector superenrollado (construcción que contiene a *prai* consenso completo + *cat*) para evaluar la eficiencia de las células electrocompetentes.

<b>BANCO</b>	<b>CLONAS POR CAJA SP</b>	<b>CLONA POR CAJA CM<sub>25</sub></b>
<b>Superenrollado</b>	Incontables	Incontables
<b>Vector auto ligado</b>	1770	0
<b>Ligación</b>	5085	120

Tabla 9. Conteo de colonias del banco generado al insertar fragmentos al azar corriente arriba de la unidad C-( $\beta/\alpha$ )<sub>4</sub>, utilizando CAT como reportero de plegamiento. En teoría, las 120 colonias que resisten cloramfenicol, se encuentran expresando una fusión traduccional que incluye a CAT en su forma biológicamente activa.

Observando la tabla 9, inmediatamente salta a la vista que no existen clonas autoligadas en cloramfenicol (25  $\mu\text{g/ml}$ ), este hecho se traduce en un buen funcionamiento del sistema de selección. Si analizamos las cifras podremos apreciar que existe un número bastante grande de colonias en la ligación cuando seleccionamos las transformantes en espectinomicina, y este número se reduce en gran medida cuando la selección se realiza en cloramfenicol (este comportamiento es teóricamente esperado).

En un principio podríamos pensar que las 120 clonas seleccionadas en cloramfenicol contienen un inserto que complementa el plegamiento de la unidad C-( $\beta/\alpha$ )<sub>4</sub>. Sin embargo, a la luz de los resultados del banco anterior (muchas veces las colonias cloramfenicol resistentes son producto de la transformación de plásmido superenrollado contaminante), es necesario analizar más a fondo cada una de las clonas antes de poder concluir algo acerca de ellas. Aunque un dato relevante, es que las células tienen una buena eficiencia de electroporación, al observar el elevado número de colonias en las cajas transformadas con plásmido superenrollado.

El siguiente paso en el análisis del banco consistió en hacer *PCR*'s de cada una de las 120 clonas seleccionadas en cloramfenicol. Para este propósito se picaron todas las colonias, inoculando medio líquido LB, incubándolas por 16 horas a 37 °C. Se extrajo plásmido que se utilizó como templado. Con los *PCR*'s buscamos productos por arriba de 500 pb (producto esperado en el caso de clonas autoligadas) y alrededor de 800 pb (el tamaño del gen que codifica para la proteína consenso completa + 200 pb del vector).

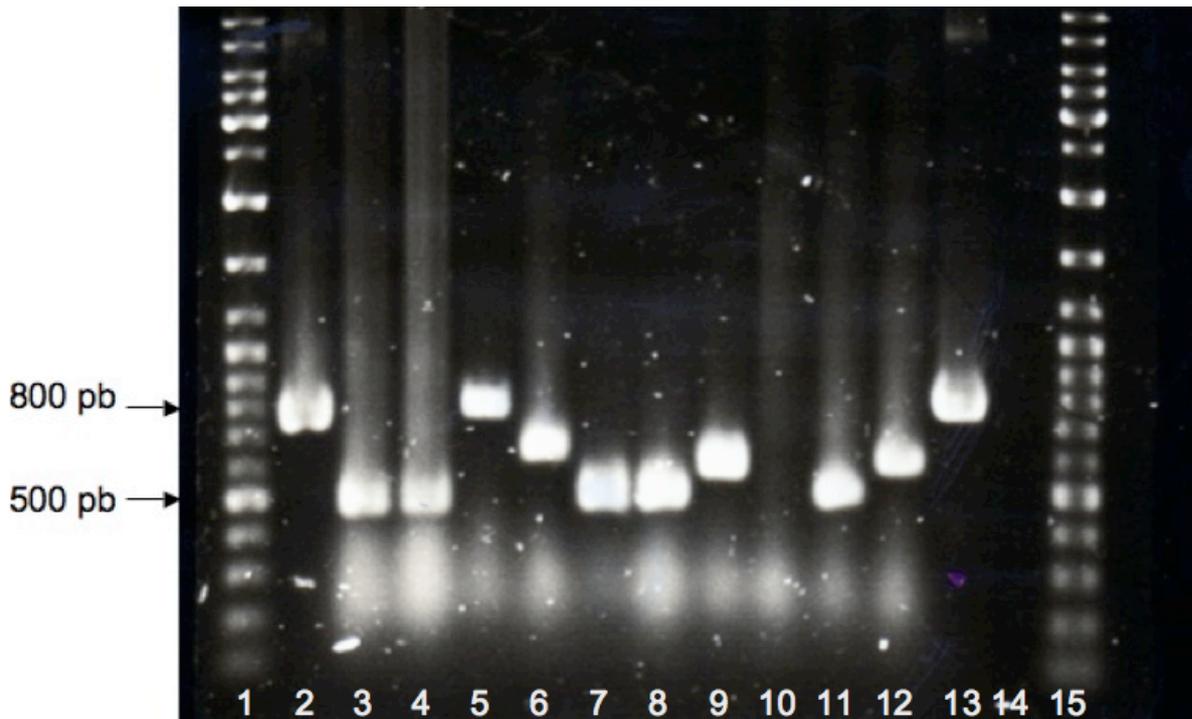


Figura 11. Electroforesis en gel de agarosa de PCR's de 10 clonas de las 120 seleccionadas en cloramfenicol. Carriles: 1) Marcador de peso molecular, 2) Control positivo del PCR (800 pb), 3) al 12) productos de PCR de las clonas seleccionadas, 13) control positivo y 15) marcador de peso molecular.

En la figura 11 se muestra el resultado del PCR de 10 de las clonas seleccionadas en cloramfenicol. Podemos apreciar que los productos de los carriles 3, 4, 7, 8 y 11 corresponden a los pesos de clonas autoligadas (500 pb), y que el producto del carril 5 pertenece al peso del PCR del gen de la proteína completa (800 pb). Aquí es momento de hacer un análisis de lo que podríamos estar encontrando. Es claro que los productos de 500 pares de bases corresponden a clonas que no tienen inserto (clonas autoligadas). Estas clonas de alguna manera codifican para una fusión traduccional en la cual CAT puede alcanzar su conformación biológicamente activa. Podemos explicar el producto de 800 pares de bases como plásmido superenrollado contaminante que no se digirió completamente.

También podemos explicarlo suponiendo que el fragmento que presenta esta clona es exactamente del mismo tamaño que el medio barril que se removió. Por último, los productos de PCR de los carriles 6, 9 y 12 tienen una magnitud intermedia entre lo que sería el tamaño de una clona autoligada (500pb) y el del plásmido superenrollado (800 pb); estas clonas potencialmente presentan insertos de dimensión variable. Un patrón similar se repitió en el análisis de las 120 clonas, encontrando muchos productos correspondientes a clonas autoligadas; otros más a plásmido superenrollado, y algunos otros correspondientes a tamaños intermedios. Un dato curioso es que no logramos identificar ninguna clona que tuviera un inserto más grande que la proteína consenso completa; es decir, no obtuvimos ningún producto de PCR de más de 800 pb en ninguna de las 120 clonas seleccionadas.

Posteriormente se mandaron a secuenciar todas las clonas que presentaban fragmentos de tamaños variables. También las que presumiblemente eran autoligadas o plásmido superenrollado; sin embargo, no podíamos asegurarlo debido al poder de resolución del gel. Posiblemente el fragmento que se insertó, era justamente del mismo tamaño de la mitad del barril. Finalmente, enviamos a secuenciar las clonas que no dieron producto de *PCR* (como es el caso del carril 10 de la figura 12) conformando un total de 70 clonas secuenciadas.

<i>ANÁLISIS DEL BANCO</i>	<i>CLONAS SECUENCIADAS</i>
<i>Superenrollado</i>	20
<i>Autoligadas</i>	30
<i>Inserto en fase con lac</i>	12
<b>Inserto positivo</b>	8
<b>Total</b>	70

Tabla 10. Resultados de la secuenciación de 70 clonas.

En la tabla 10 se aprecia el resultado de la secuenciación de las clonas seleccionadas, el análisis arrojó datos muy interesantes. A pesar de haber seleccionado previamente las clonas a secuenciar, encontramos 20 que son contaminación por plásmido superenrollado y 30 autoligadas. Estas últimas de alguna manera permiten que CAT se pliegue; abordaremos con más detalle este resultado en la discusión del trabajo. Ahora explicaremos que significan las 12 clonas con inserto en fase con lac y las 8 clonas con inserto positivo (tabla 11).

Clona	Marco de lectura en la proteína original <sup>(a)</sup>	Tamaño del inserto (aminoácidos)	Proteína de origen	Tipo de plegamiento
Meg	-	57	Gen identificado para la proliferación celular	ND
17s	+ (N)	52	Hidrolasa hipotética	ND
N14	+ (N)	52	Endopeptidasa	ND
Q1	+ (N)	57	yfjQ es inferida por homología con función desconocida	ND
Q5	+	24	Transportador yeiM	ND
1a	+	39	Componente 1 de la Antranilato sintasa	Sándwich alfa/beta 4
14s	+	44	Isocorismato sintasa	Sándwich alfa/beta 4
20s	+	31	Carboxilasa sub-unidad ATPasa	Sándwich alfa/beta 3

Tabla 11. Clonas positivas. Siete de ocho insertos se fusionaron a la unidad  $(\beta/\alpha)_4$  en la misma fase en la que se encontraban en su proteína de origen, sólo uno se fusionó en una fase distinta<sup>(a)</sup>; tres insertos corresponden a los extremos amino terminal en sus proteínas de origen, se encuentran indicados en la tabla con una 'N'.

Las clonas con inserto en fase con lac son el resultado de una fusión triple; la clonación de una parte del banco de secuencias genómicas en nuestra construcción, dio como resultado que una región de lac quedara fusionada en fase con los insertos, que a su vez se fusionaron con la unidad C-( $\beta/\alpha$ )<sub>4</sub> formando un ORF completo que incluye al reportero de plegamiento. El codón que codifica para metionina, en las fusiones triples, se encuentra en la región fusionada de lac.

Es importante mencionar que en la clonación de los fragmentos no incluimos un codón de inicio en la construcción. Lo esperado era recuperar insertos que ya presentaran ese triplete. Las 8 clonas con inserto positivo presentaron fragmentos que tienen un codón de inicio y están en fase con la unidad C-( $\beta/\alpha$ )<sub>4</sub> formando un ORF completo con el gene de CAT (una descripción detallada de las construcciones se encuentra en el apéndice de este trabajo). Estas clonas son el objetivo principal de esta tesis.

Más adelante hablaremos de la identificación de los genes de donde proceden los insertos y de la caracterización bioquímica (solubilidad) de las nuevas proteínas quiméricas. El reportero de plegamiento fue removido de estas construcciones (excepto de la clona Q1) mediante una digestión doble con BamHI y BglII. Estas clonas se secuenciaron y se procedió con su caracterización como se describe en la siguiente sección de este trabajo.

## Resultados Bioquímica

Después del análisis del banco final se lograron obtener 8 clonas positivas. Consideramos clonas positivas aquellas que reúnen las siguientes características: a) presentan un inserto en fase con la mitad del barril que se utilizó como templado, b) este inserto incluye un codón de inicio (Met) que permite su traducción, c) la construcción no muestra tener codones de paro que interrumpan el marco de lectura que se establece desde el inicio del inserto y hasta el final del reportero de plegamiento.

Las clonas que se consideraron positivas se retransformaron y seleccionaron en  $cm_{35}$  para constatar que no se tratará de falsos positivos causados por algún malfuncionamiento del antibiótico. Posteriormente procedimos a expresar las proteínas para evaluar su perfil de expresión, estimando la proporción de las mismas que se encontraba tanto en la fracción soluble, como en la insoluble. Como control positivo transformamos, en la cepa de expresión (JM101), la construcción que codifica para la proteína consenso completa fusionada al reportero. Para el control negativo utilizamos la cepa de expresión sin plásmido. Iniciamos el procedimiento de expresión picando colonias de todas las clonas, la metodología fue la siguiente:

- a) Se retransformaron las clonas positivas secuenciadas en la cepa JM101 de *E. coli*.
- b) Se picaron colonias aisladas de placas recién estriadas.
- c) Se preparó un preinóculo de 3 ml en el antibiótico del vector (sp) incubando 16 horas.
- d) Se inocularon 25 ml de LB (sp) con 1 ml del preinóculo.
- e) Se incubó el cultivo creciendo a 37 °C , 250 rpm hasta alcanzar una densidad óptica de 0.4 (600 nm).
- f) Se adicionó al cultivo 0.5 mM de IPTG incubando a 25 °C 220 rpm hasta alcanzar una densidad óptica de 0.8 (600 nm).
- g) Se colectaron las células centrifugando 20 minutos a 4,500 rpm, resuspendiéndose en amortiguador NaCl 300 mM, fosfato de potasio 20 mM, 10% glicerol, pH 7.5
- h) Las células se rompieron mediante sonicación: Tres rondas de 10 pulsos a potencia 2, punta chica, adicionando PMSF (inhibidor de proteasas) después de la segunda ronda.
- i) Se separaron las fracciones en solubles e insolubles centrifugando el sonificado a 10,000 rpm (rotor JA14 Beckman) por 25 minutos.

La centrifugación diferencial permite separar los componentes del sonificado en fracciones solubles e insolubles. La fracción insoluble está integrada principalmente por membranas y proteínas que no reunieron los requerimientos fisicoquímicos para solubilizarse en el amortiguador de trabajo. La fracción soluble está compuesta principalmente por proteínas que se encuentran solubilizadas en el mismo amortiguador. Posteriormente al fraccionamiento del contenido celular en las dos partes antes mencionadas, se mezclaron con amortiguador de carga y se sometieron a 95 °C por 5 minutos para provocar su desnaturalización.

## Electroforesis en gel de poliacrilamida desnaturalizante

La fracción soluble de cada uno de los cultivos de las clonas mutantes se analizó mediante electroforesis desnaturalizante en gel de poliacrilamida.

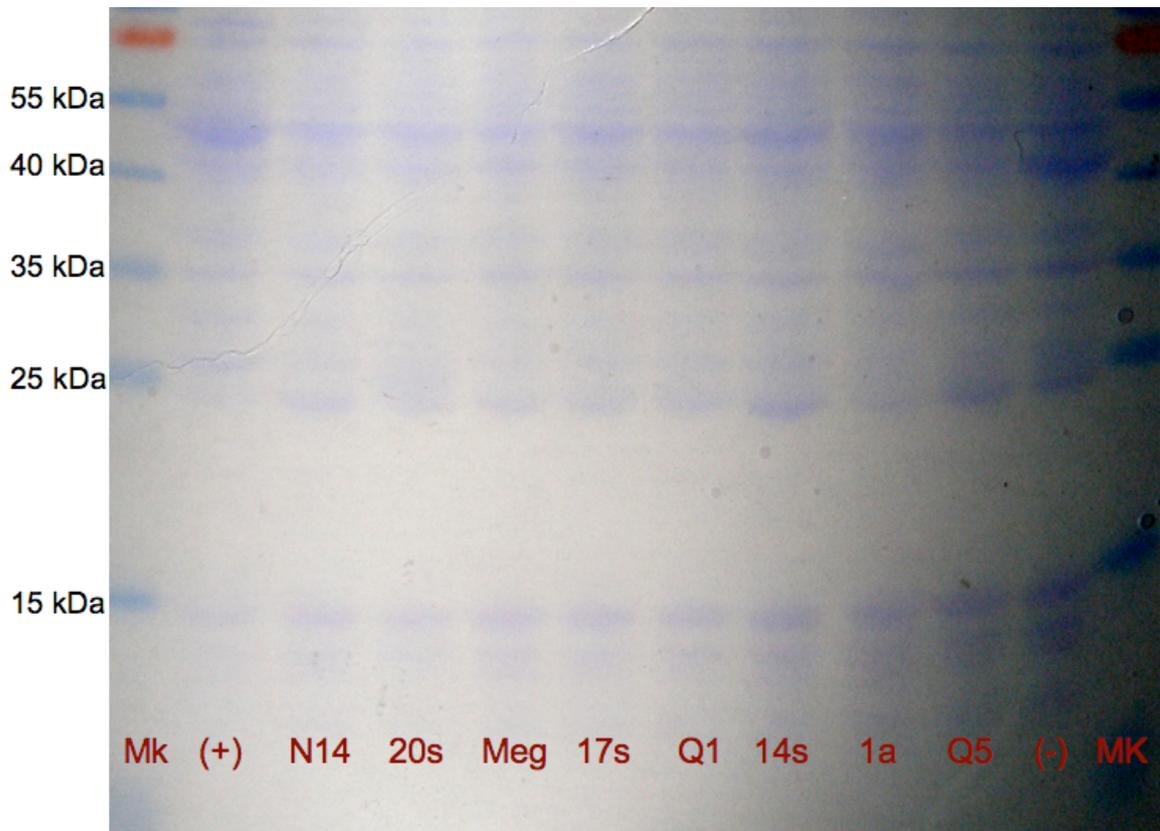


Figura 12. Electroforesis en gel de poli-acrilamida de las fracciones solubles de cada una de las clonas mutantes. Se muestran también los controles positivo y negativo flanqueados por marcadores de peso molecular. El peso esperado de las fusiones es de menos de 47 kDa.

La electroforesis en gel de poliacrilamida no reveló ninguna diferencia apreciable al comparar la cantidad de proteína soluble entre las clonas mutantes y los controles. Sólo fue posible observar una banda particularmente más intensa que las demás en el carril del control positivo. La banda que se espera de la fusión de la proteína consenso completa con el reportero es de 47 kDa. A pesar de que se probaron algunos amortiguadores distintos y condiciones de expresión variables, no se logró observar cambio en el perfil de expresión.

En vista de que la simple inspección del gel de proteínas no permitió establecer si las nuevas mutantes fusionadas al reportero estaban presentes en la fracción soluble, fue necesario montar un ensayo de inmunodetección. El ensayo se realizó utilizando en primera instancia un anticuerpo anti-cat acoplado a dioxigenina. En un paso subsiguiente, se utilizó un anticuerpo secundario anti-dig acoplado a fosfatasa alcalina. El *western blot* se reveló con el sustrato de la enzima mostrando el siguiente patrón de inmunodetección:

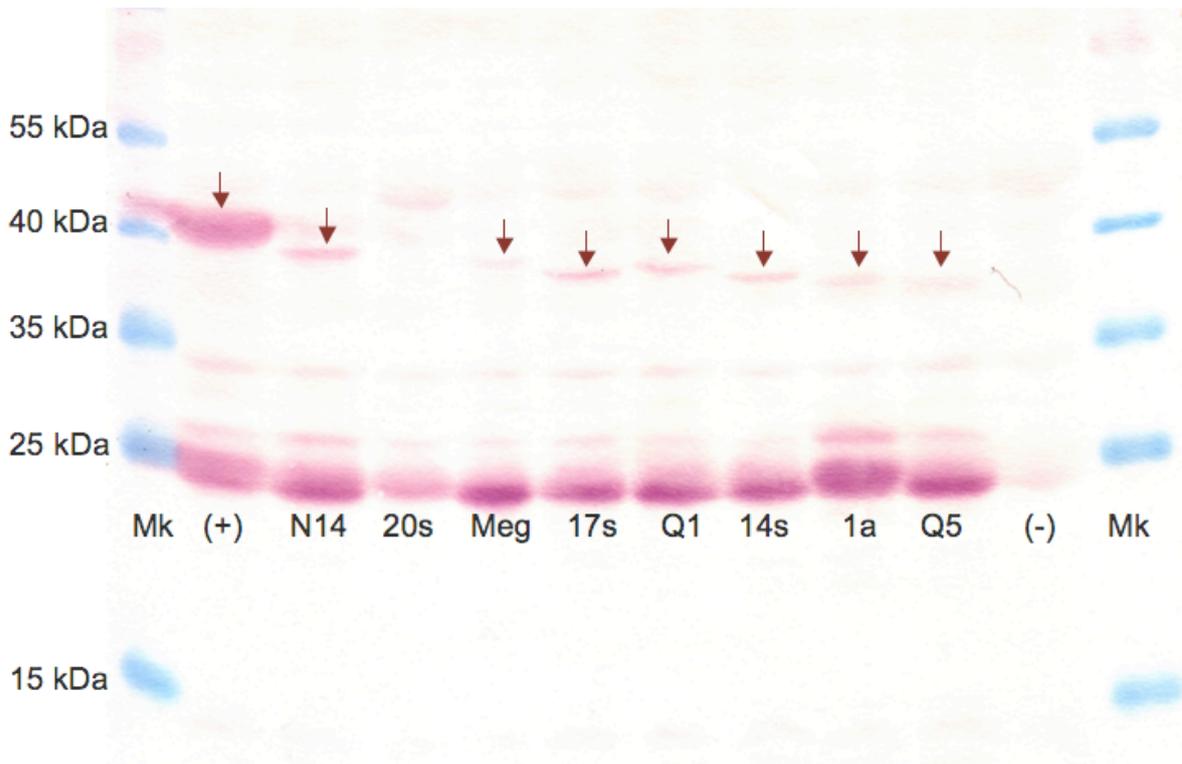


Figura 13. *Western blot* anti-cit de las clonas mutantes. En la inmunodetección están indicadas las bandas correspondientes a las fusiones. La banda del control positivo es muy intensa en comparación con las fusiones y no fue posible detectar la banda de la clona 20s. Es posible apreciar que existe una gran cantidad de reportero (CAT), en todos los carriles excepto en el control negativo, expresado por separado de la fusión en el peso correspondiente a los 25 kDa (CAT tiene un peso de 27 kDa).

El *western blot* arroja datos muy interesantes acerca del perfil general de expresión de las proteínas mutantes. El dato más importante es que fue posible detectar, en la fracción soluble, casi todas las bandas que corresponden a las nuevas proteínas mutantes fusionadas al reportero, con excepción de la clona 20s.

Al analizar la secuencia de los fragmentos que se insertaron complementando el medio barril en nuestro sistema, nos dimos cuenta que son más pequeños que la mitad que se removió del barril: el más pequeño es el de la clona Q5 de 24 residuos y los más grandes son los de las clonas Meg y Q1, ambos de 57 residuos. Es posible observar que existe una banda muy intensa en el peso correspondiente a los 25 kDa; esta banda es visible en todos los carriles, con excepción del control negativo. Nuestra hipótesis es que esta banda es un producto de proteólisis. Cada una de las quimeras se expresa fusionada al reportero; sin embargo, tan pronto como éstas alcanzan el estado plegado, una fracción importante de la población total de las proteínas fusionadas se ve sometida a la actividad de proteasas. Es por esto que podemos encontrar CAT expresándose separada de la fusión.

Entre la banda de la fusión y la banda que corresponde a CAT expresándose por separado, también encontramos algunas otras bandas que se presentan al mismo nivel en todos los carriles.

Presumiblemente, otros productos de degradación son derivados de la fusión y están siendo detectados por el anticuerpo. Inferimos esto porque estas bandas no están presentes en el carril del control negativo (cepa de expresión que no contiene el vector con la fusión).

La observación de que el reportero se encuentra en la fracción soluble sin estar fusionado con la proteína tiene implicaciones muy importantes. Podríamos concluir que la capacidad para resistir el antibiótico se debe únicamente a la fracción de CAT que no está fusionada. En ese caso, todas las clonas que pudiéramos ser capaces de seleccionar serían falsos positivos (no serían capaces de alcanzar una conformación estable sin estar fusionadas al reportero). Sin embargo, también podemos detectar las bandas de las proteínas fusionadas, lo que significa que existe una fracción de la fusión traduccional que permanece íntegra y soluble en el citoplasma de las bacterias. Una vez que el reportero de plegamiento ha funcionado para seleccionar proteínas que en teoría son capaces de plegarse, procedemos a removerlo de la fusión. Para poder realizar estudios biofísicos con las nuevas proteínas quiméricas, es necesario expresarlas independientemente del reportero de plegamiento.

Como se ha mencionado en los resultados de biología molecular, el siguiente paso en la metodología fue escindir el gen de cat de las clonas seleccionadas mediante una digestión doble con las endonucleasas de restricción BamHI y BglII. Al realizar esta digestión se remueve este gen de la construcción y simultáneamente se inserta una marca que codifica para 6xHis. Esta marca nos permite detectar las mutantes utilizando un anticuerpo que reconoce este motivo, así como la purificación de las mismas valiéndonos de una columna de níquel.

No fue posible remover el reportero de la clona Q1 ya que en el inserto que complementó la unidad  $(\beta/\alpha)_4$  se encontró un nuevo sitio BamHI. Habría sido necesario remover ese sitio con una mutagénesis sitiodirigida; sin embargo, en esta fase del experimento implementar ese método no resultaba trivial, así que se decidió continuar trabajando con las 7 quimeras restantes. Después de digerir las clonas con las enzimas de restricción, se verificó la integridad de las construcciones mediante su secuenciación.

Las proteínas quiméricas contenidas en la fracción soluble construidas en este estudio, son el objetivo primario a caracterizar tanto, bioquímica como biofísicamente. Más adelante, en la parte correspondiente a la discusión, abordaremos de manera amplia el fundamento de esta aseveración. No es un problema de fácil solución obtener una cantidad considerable de proteína soluble que nos permita realizar experimentos de dicroísmo circular o desnaturalización térmica (metodologías que nos permiten obtener información del plegamiento que adoptan estas especies proteicas) utilizando sistemas de expresión recombinantes. No hay que olvidar que en este trabajo estamos tratando de caracterizar proteínas quiméricas artificiales, en muy pocos casos estas especies se van a comportar como lo harían proteínas nativas que han estado sujetas a procesos evolutivos durante algunos millones de años.

Las nuevas proteínas quiméricas se transformaron en células electrocompetentes, que previamente habían sido transformadas con un plásmido que codifica para un sistema de chaperonas (DnaK y DnaJ) (Acebron 2007). Se ha demostrado que es posible aumentar la solubilidad de proteínas expresadas en *E. coli* co-expresándolas con estas chaperonas. Las condiciones de expresión que utilizamos fueron las mismas que se usaron para expresar las proteínas fusionadas. Se prepararon fracciones solubles a partir de estos cultivos y se analizaron mediante una electroforesis en gel de poliacrilamida.

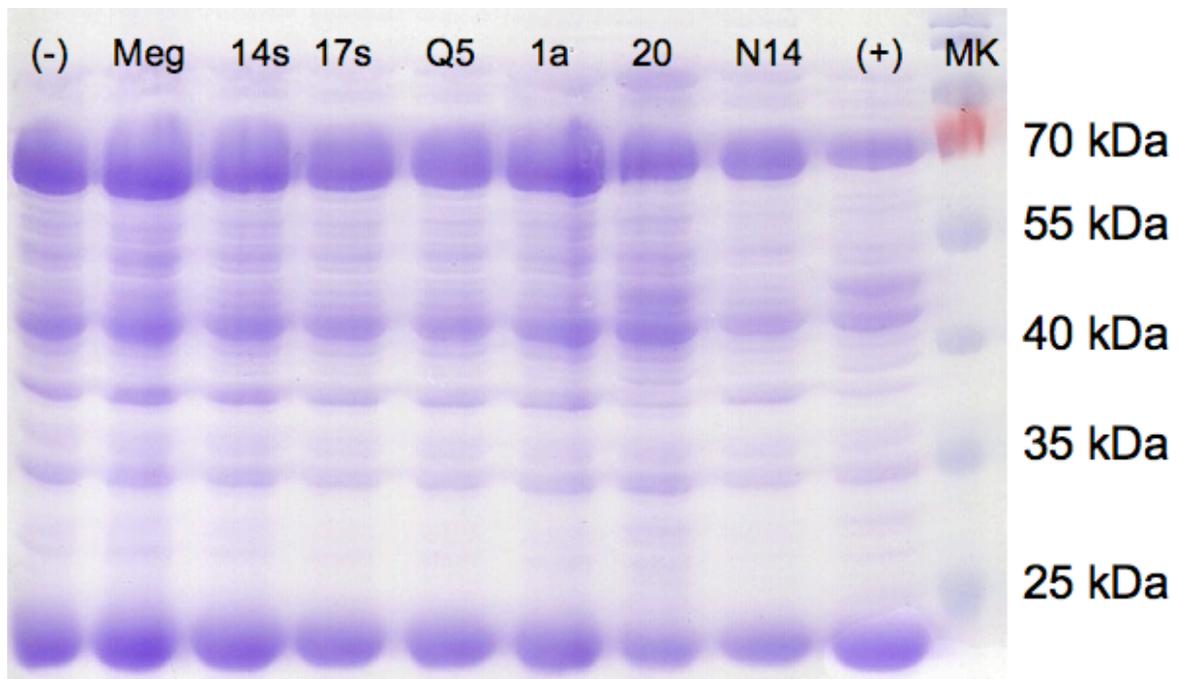


Figura 14. Se muestran las proteínas quiméricas junto con los controles positivo y negativo. En el gel de acrilamida se aprecian dos bandas muy intensas que corresponden a las chaperonas: DnaK de 70 kDa y DnaJ de alrededor 25 kDa. Desafortunadamente, no es posible distinguir las bandas que podrían corresponder a las quimeras o al control positivo por dos razones: 1) la banda que corresponde a la chaperona más pequeña la encontramos justamente en el peso esperado para las mutantes, esto impide su visualización y/o 2) que la cantidad de proteína soluble sea muy poca.

La electroforesis en gel de poliacrilamida no permitió distinguir las bandas que podrían corresponder a las proteínas quiméricas, debido a la presencia mayoritaria de la banda correspondiente a DnaJ justo en el peso esperado para nuestras mutantes. Sin embargo podemos apreciar que ambas chaperonas están en una proporción importante de la fracción soluble. Monitorear la expresión de proteínas mutantes con el simple análisis de un *SDS PAGE* puede ser una tarea muy difícil. Son pocos los casos en que podemos observar bandas muy intensas que correspondan a la expresión de proteínas mutantes. En consecuencia, procedimos a realizar una inmunodetección de las proteínas quiméricas utilizando un anticuerpo anti 6xHis. Sin embargo, no logramos realizar este experimento con éxito ya que no fue posible detectar el control positivo ni tampoco ninguna de las quimeras. Aunque el experimento se realizó por triplicado no se llegó a obtener un resultado concluyente.

Es posible que los anticuerpos que utilizamos no se encontraran en buenas condiciones o que la cantidad de proteína existente en la fracción soluble no fuera detectable por los anticuerpos bajo estas condiciones. A pesar del resultado anterior, seleccionamos dos quimeras para montar su purificación a partir de la fracción soluble de los cultivos, utilizando una columna de níquel( Meg y N14). Era posible que la cantidad de proteína expresada en los cultivos anteriores, fuera muy poca para ser detectada. Montando una purificación a partir de un litro de cultivo podríamos obtener proteína concentrada detectable con el anticuerpo, o en cantidades suficientes para realizar experimentos de caracterización biofísica.



Figura 15. Electroforesis en gel de poliacrilamida de la purificación montada para la clona N14. Carriles: A) Marcador de peso molecular, B) y C) lavados, D) al L) fracciones colectadas. No es posible encontrar la banda correspondiente a la clona N14. Debería encontrarse entre los marcadores de peso molecular que indican 25 kDa y 15 kDa. Solamente se aprecia una banda que se encuentra más arriba de 25 kDa, probablemente a causa del pegado inespecífico de alguna proteína presente en el lisado.

Como se puede apreciar en la figura 15, no fue posible obtener proteína pura en la experimento realizado. El mismo resultado obtuvimos cuando se llevó a cabo la purificación de la clona Meg. Estos datos nos indican que tal vez el motivo 6XHis no se encuentra expuesto, y no fue capaz de interactuar con la columna de níquel ni con los anticuerpos.

## Resultados Bioinformática II

### Identificación de los fragmentos que forman las nuevas quimeras.

Como resultado de la selección a partir del banco de variantes, logramos identificar 8 clonas positivas que permitían obtener fusiones traduccionales solubles. Las fusiones seleccionadas contenían distintos fragmentos de genes de *Escherichia coli*, que fueron capaces de complementar el plegamiento de la unidad C-( $\beta/\alpha$ )<sub>4</sub> utilizada como templado. Para identificar las proteínas de las cuales provienen los fragmentos realizamos BLAST con cada uno de ellos.

#### 1. Meg

Este fragmento no pudo ser identificado haciendo un BLAST con su secuencia de aminoácidos vs. el genoma de *E. coli*, ya que el marco de lectura que presentaba en la fusión traduccional con la mitad del barril, era distinto al marco de lectura de su proteína de origen. Para poder identificar el fragmento insertado realizamos BLAST con la secuencia de nucleótidos del fragmento (171); encontramos que el fragmento identifica 9 *hits* con una identidad del 100%, un *score* de 340 y un *e-value* de 9e-82.

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
<a href="#">AP009048.1</a>	Escherichia coli W3110 DNA, complete genome	315	315	100%	2e-85	100%
<a href="#">U00096.2</a>	Escherichia coli K12 MG1655, complete genome	315	315	100%	2e-85	100%
<a href="#">L10328.1</a>	E. coli; the region from 81.5 to 84.5 minutes	315	315	100%	2e-85	100%

Tabla 11. (Tomada de NCBI) Identificación del fragmento de la clona Meg haciendo BLAST con la secuencia de nucleótidos. Encontramos 3 *hits* con los mismos valores en distintos genomas secuenciados de *E. coli*.

Todos los *hits* encontrados son regiones del genoma de *E. coli* que corresponden a genes no caracterizados. Uno de ellos se ha identificado como un gen que se requiere para la proliferación celular. Podríamos realizar predicciones de estructura secundaria para tratar de tener una idea del tipo de elemento que logra complementar el plegamiento de nuestra unidad utilizada como templado. Sin embargo, la secuencia de nucleótidos que se insertó en la clona positiva está en un marco de lectura distinto al de la proteína original, por lo que una predicción de estructura secundaria con la proteína original no revelaría mucho acerca de la posible estructura secundaria que adoptó el fragmento en la quimera nueva.

#### 2. N14

Se realizó un BLAST con la secuencia de aminoácidos del fragmento de 52 residuos vs el genoma de *E. coli* y contra genomas de procariontes del EMBL-EBI (para identificar secuencias homólogas con estructura), encontrando varios *hits* con un 100% de identidad. Todos éstos reconocen la proteína de origen de este fragmento como una *O-sialoglycoprotein* endopeptidasa hipotética.

El fragmento que se insertó corresponde al extremo amino terminal de la proteína; en otras palabras, complementamos nuestra unidad carboxilo terminal con un fragmento amino. El homólogo más cercano (38% identidad) con estructura determinada es 2ivn que es un ortólogo de la proteína universal Kae1, una metaloproteína que tiene actividad atípica de unión a ADN y actividad de endonucleasa apurínica. Aunque estos tipos de proteínas han sido anotadas como endopeptidasas su función es desconocida. En la referencia principal de la descripción de esta proteína se le atribuyen funciones importantes para mantener la integridad del genoma en las células.

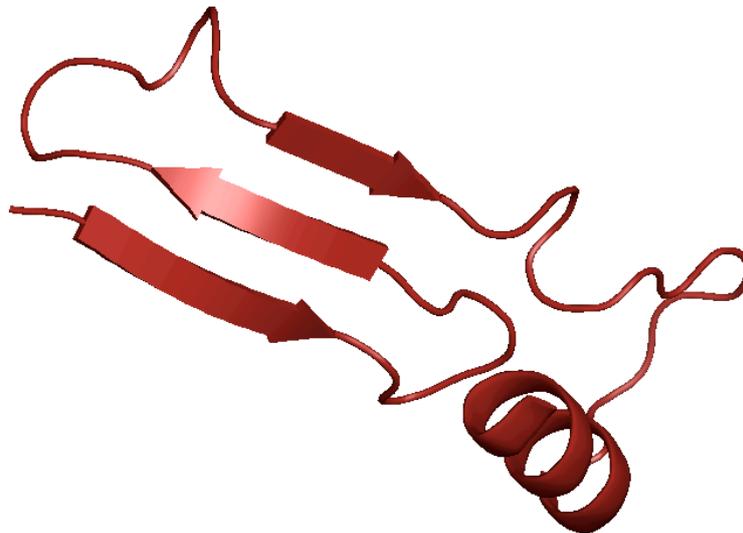


Figura 16. En rojo se muestra el segmento de 52 aminoácidos que se insertó en la clona N14 (la estructura con la cual se generó el modelo de la proteína hipotética es 2ivn) produciendo una quimera con el extremo carboxilo terminal del barril.

### 3. Q1

El BLAST con el fragmento de 52 aminoácidos de esta clona nos permitió identificarla como una proteína hipotética yfjQ (identificador de *swiss-prot* YFJQ\_ECOLI) perteneciendo a la familia de proteínas de proteobacterias con función desconocida. No se pudo localizar algún homólogo con estructura.

Se trató de realizar un modelo utilizando swiss-model, sin embargo no al existir templados viables no fue posible generarlo. Se procedió a realizar una predicción de estructura secundaria con psipred. Un dato importante es que nuevamente se clonó el extremo amino terminal de la proteína de origen, lo cual indica que complementamos el extremo carboxilo que utilizamos como templado, con un amino que le permitió plegarse.





## 7. a1

El BLAST con un fragmento de 39 aminoácidos insertado en esta clona encuentra un *hit* principal con el componente I de la antranilato sintasa. Esta enzima es en algunos organismos parte de un complejo multifuncional en la biosíntesis de triptofano. Al no contar con la estructura resuelta de la misma realizamos un modelo teórico utilizando como templado la estructura de la antranilato sintasa de *Salmonella typhimurium* (1i1qA).

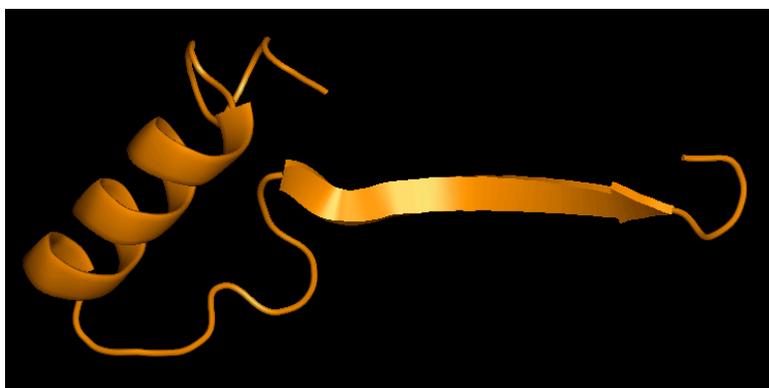


Figura 21. Fragmento insertado en la clona a1. El modelo teórico tridimensional se realizó con swiss-model utilizando el templado 1i1qA; éste, presenta un 90% de identidad con la proteína de la cual proviene el inserto. El extremo amino del fragmento se encuentra en el asa que precede el alfa hélice. Este fragmento está entre las posiciones 421 a 459 en su proteína original.

## 8. 17s

Esta clona presenta un inserto de 52 aminoácidos. Haciendo BLAST con este fragmento encontramos una hidrolasa hipotética con identificador de swiss-prot A1AHZ4\_ECOK1. Al revisar el alineamiento del fragmento con su proteína de origen, encontramos una mutación (P17L). El fragmento insertado en la clona corresponde al extremo amino terminal del polipéptido de origen. Para constatar si la mutación se debía a algún error de secuenciación, verificamos con una segunda secuenciación encontrando la misma mutación. No contamos con una estructura resuelta para esta proteína así que intentamos realizar un modelo teórico sin éxito, ya que parece ser que el extremo amino terminal que se encuentra en la proteína que origina el fragmento no se encuentra en los homólogos. *Swiss-model* nos devolvió un modelo tomando como templado la estructura 1nrwA, que corresponde a una Dehalogenasa haloacídica *like* hidrolasa de *Bacillus subtilis*.

La estructura que utilizamos como templado presenta un 28% de identidad con la hidrolasa hipotética identificada. Al revisar el modelo constatamos que la región amino terminal que se había insertado en la clona 17s no pudo ser modelada. El aminoácido 42 de la proteína hipotética alinea con el aminoácido 2 en el templado, sugiriendo que el extremo amino es una inserción en la hidrolasa hipotética identificada. Tratando de tener una idea del tipo de estructura que complementó el plegamiento de la unidad  $(\beta/\alpha)_4$ , generamos una predicción de estructura secundaria con la secuencia de la hidrolasa hipotética.

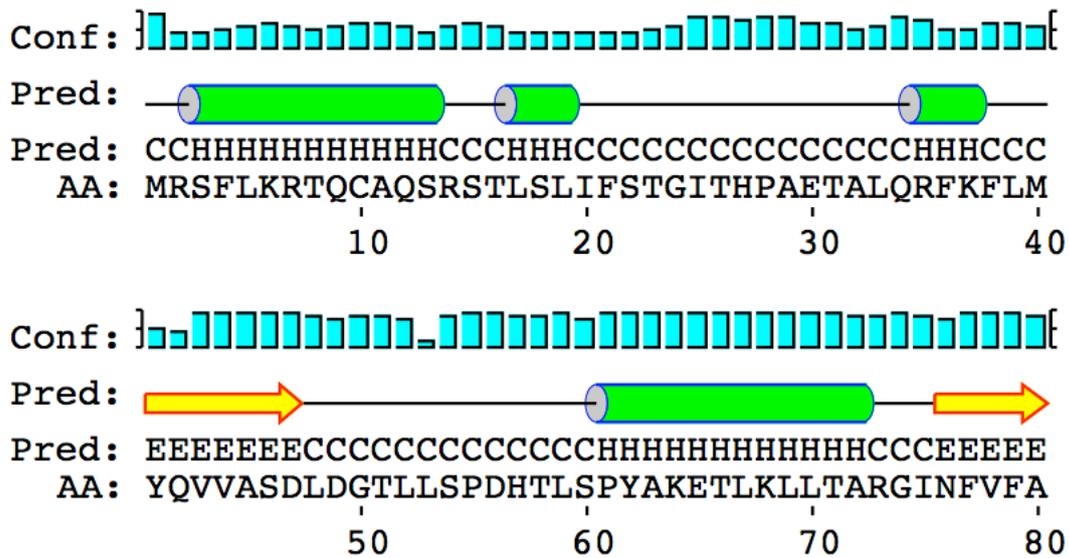


Figura 22. Estructura secundaria generada con la secuencia del fragmento insertado en la clona 17s; utilizamos como templado la secuencia de la hidrolasa hipotética identificada con BLAST. El fragmento insertado en la clona inicia en la posición 1 y termina en la 52.

A partir de la predicción de estructura secundaria, podemos observar que el fragmento insertado comprende tres hélices alfa, un asa y una hoja beta; sin embargo, el nivel de confianza alcanzado con esta predicción es bajo como se puede apreciar en la figura 22.

## Discusión y Perspectivas.

El origen de los dominios proteicos es una de las más grandes interrogantes que existe en ciencia de proteínas. En la búsqueda de información que nos permita aumentar el conocimiento generado alrededor de este enigma, desarrollamos un trabajo en el cual, aunque no se conoce su estructura, obtuvimos proteínas plegadas utilizando la información estructural contenida en la unidad C-( $\beta/\alpha$ )<sub>4</sub> de PRAI consenso como templado, para buscar secuencias que complementarían el plegamiento de este subdominio. Como estrategia de tamizaje construimos una fusión traduccional con un reportero de plegamiento (CAT).

Se han realizado diversos trabajos de investigación acerca de los dominios proteicos para postular teorías que permitan explicar su origen. Existe una hipótesis muy importante que ha sido desarrollada por varios grupos de investigación en los últimos años, principalmente el de Andrei Lupas. Si observamos los plegamientos que adoptan las proteínas, podremos darnos cuenta que existen motivos super estructurales que se repiten en muchos polipéptidos aunque posiblemente no tienen un origen común. Esto se ha explicado de dos maneras, principalmente: podríamos estar observando casos de convergencia o divergencia por causas funcionales, o podríamos postular que los dominios surgen a partir de cadenas polipeptídicas pequeñas que se fusionan para formar estructuras más complejas (Ponting 2001). En el caso de los barriles TIM, la repetición de estructuras  $\beta/\alpha$  es una evidencia que sugiere este mecanismo.

En dos antecedentes principales de este trabajo, encontramos evidencias afirmando que es posible generar dominios polipeptídicos estables a partir de ladrillos constructores. En un primer acercamiento (Riechmann 2000), Winter fusiona secuencias genómicas al azar a un núcleo de plegamiento, obteniendo proteínas estables. Posteriormente (Winter 2006), complementa un péptido que aislado de su contexto original no es estable (pero conserva información suficiente para unir anticuerpos) con secuencias al azar, para generar una proteína que aumentó su afinidad y es soluble.

En este punto podemos observar dos vertientes muy interesantes; por un lado se utiliza un fragmento que contiene información estructural capaz de plegar secuencias al azar; y por el otro se complementa un péptido aislado que contiene cierta información funcional con una secuencia que posiblemente aportó información que le permitió plegarse de manera estable (es posible que se haya seleccionado un núcleo de plegamiento). Las dos estrategias responden preguntas diferentes, aunque ambas reconstruyen el posible origen de los dominios.

En la actualidad contamos con algunos métodos muy poderosos para analizar estabilidad de proteínas. Entre las metodologías que nos permiten estudiar esta característica *in vivo*, encontramos el sistema reportero que utiliza CAT en una fusión traduccional con una proteína blanco. Es pertinente analizar los alcances y limitaciones que este sistema representó para nuestro estudio, en el contexto del banco de variantes generado.

La construcción de la librería de variantes fue el punto más complejo a lo largo de toda la estrategia experimental de esta tesis; la dificultad radica en la gran cantidad de variables involucradas. El estado del material genético es una de las variables que más influye en el resultado final del experimento; por esta razón, la selección del método de fragmentación y la optimización de los procesos de reparación resultaron determinantes. Otra variable fundamental es el tiempo en el que se realiza el ensayo, hay que clonar el banco en el menor tiempo posible, ya que la eficiencia se reduce significativamente cuando las muestras se refrigeran por muchos días. En conjunto, la obtención de bancos de variantes lo bastante grandes como para lograr una exploración suficiente del espacio de secuencia, es una tarea sumamente difícil. Entre tanto discutiremos los resultados obtenidos con el sistema de selección (CAT).

Si observamos la tabla de resultados 10 podremos notar que hallamos una gran cantidad de falsos positivos. De 120 clonas seleccionadas en cloramfenicol, solamente 20 presentaron un inserto, a pesar de que teóricamente las clonas autoligadas no son capaces de conferir resistencia al antibiótico cuando utilizamos el extremo carboxilo de PRAI consenso como templado. Entre los falsos positivos encontramos 40 clonas que contienen el gen de la proteína consenso completa, como resultado de una digestión parcial y posterior contaminación del banco a seleccionar con el plásmido silvestre. Este tipo de falsos positivos no pueden ser atribuidos al sistema de selección, ya que se originan por motivos distintos de la fusión traduccional (digestiones parciales y extracciones de geles de agarosa poco eficientes). También encontramos 60 clonas autoligadas seleccionadas como cloramfenicol resistentes, que sí se originan en el sistema de selección. Este fenómeno puede ser explicado en razón de que debido al azar en el procedimiento de reparación del vector, se reconstruyeron codones de inicio en algunas de las clonas autoligadas.

Encontramos doce falsos positivos entre las 20 clonas con inserto que logramos seleccionar. Consisten en clonas que codifican para una quimera triple, compuesta por el extremo amino terminal de lac, el fragmento insertado a partir del banco de secuencias genómicas, y el extremo carboxilo terminal de PRAI consenso. Nosotros no esperábamos obtener esta clase de proteínas quiméricas, son el resultado azaroso del diseño de la construcción (artefactos). Por esta razón descartamos su caracterización, ya que no consideramos que sean relevantes para nuestra hipótesis. Solamente nos enfocamos en las clonas que contenían un inserto en fase con la unidad C-  $(\beta/\alpha)_4$ .

Evaluando la cantidad de verdaderos positivos *vs* falsos positivos, podríamos argumentar que a este nivel de selección el sistema es poco eficiente. Ahora bien, el procedimiento nos permite seleccionar clonas con fragmento a partir de bancos relativamente pequeños. El que los fragmentos que se insertaron complementen la unidad  $(\beta/\alpha)_4$  para generar quimeras estables sin encontrarse fusionadas al reportero, es un aspecto completamente distinto. El plegamiento *in vivo* de proteínas es un proceso sumamente complejo en el que se ven involucradas diversas variables (ribosomas, grupos prostéticos, chaperonas y composición iónica intracelular).

Comúnmente, durante la expresión en sistemas heterólogos (sobrepoducción de productos proteicos), se producen fallas en el transcurso de plegamiento, ocasionando que una gran proporción de los polipéptidos expresados se agreguen y acumulen en cuerpos de inclusión intracelulares. Muchas veces estas proteínas pueden ser recuperadas en conformaciones funcionales *in vitro*, utilizando agentes desnaturizantes como la urea o el cloruro de guanidino. De esta manera se logró realizar la caracterización biofísica de la mutante HisF-C\*C (Hocker 2004). En un trabajo posterior del mismo grupo de investigación (Seitz 2007) se logró solubilizar la misma mutante, al someterla a rondas de evolución dirigida. Creemos que es necesario discutir los resultados experimentales de esa investigación, ya que se encuentran íntimamente relacionados con nuestro trabajo.

En ese experimento, desarrollan una metodología a través de la cual generan variantes al azar del extremo N terminal de la mutante HisF-C\*C para aumentar su solubilidad. Como recordamos, esta quimera solo pudo ser purificada al re-solubilizarla a partir de cuerpos de inclusión. Para seleccionar las variantes plegadas utilizan CAT como sistema reportero de plegamiento, obteniendo los siguientes resultados: 100% de las variantes seleccionadas en 400 µg/ml y cerca del 95% de las variantes seleccionadas en 350 µg/ml y en 300 µg/ml de cloramfenicol resultaron falsos positivos. Éstos consistieron en versiones deletadas de la proteína HisF-C\*C fusionadas en fase con el extremo amino terminal de CAT (la fusión es muy parecida a la que nosotros utilizamos en este trabajo). Parece ser que la solubilidad de CAT no se vio comprometida, a pesar de encontrarse fusionada con versiones incompletas de la quimera HisF-C\*C. Más interesante aún es el hecho de que 7 de 11 mutantes completas seleccionadas con CAT, al escindir el reportero y expresarlas independientemente, no mostraron mayor solubilidad o estabilidad que la proteína parental.

Teóricamente una de las ventajas que tenemos al utilizar CAT como sistema reportero de plegamiento, es que seleccionando bacterias en un gradiente de concentraciones de antibiótico, seríamos capaces de obtener proteínas plegadas a distintos niveles. En otras palabras, si seleccionamos con una concentración alta del antibiótico, estaríamos obteniendo probablemente proteínas muy estables; en cambio, si seleccionamos con una cantidad baja de antibiótico, el resultado serían proteínas menos estables. Claramente, los datos obtenidos por el grupo de Sterner nos indican que el sistema no se comporta de esta manera. En nuestro caso, fuimos capaces de detectar con un anticuerpo anti-cat las fusiones compuestas por las quimeras y el reportero. No obstante, cuando CAT fue removida y las quimeras se expresaron independientemente fue imposible, detectarlas utilizando un anticuerpo anti 6xhis. A continuación se montaron purificaciones de dos de las clonas positivas, utilizando columnas de níquel en las condiciones descritas en la sección de resultados de este estudio, sin obtener un resultado exitoso. Es muy posible que la fusión con CAT estabilice las quimeras; este tipo de fenómeno ya había sido descrito en un trabajo anterior realizado por nuestro grupo de investigación (Escobar 2007).

La confiabilidad de los métodos de tamizaje que utilizan fusiones traduccionales puede comprometerse, debido a que se establecen interacciones imprevistas entre el reportero y la proteína blanco, que modifican el resultado del experimento. Creemos que para lograr una mayor tasa de éxito utilizando CAT como sistema de selección, es muy importante encontrar el intervalo óptimo de concentración de antibiótico.

Cabe la posibilidad de que los fragmentos que complementaron la unidad  $(\beta/\alpha)_4$  en la fusión traduccional con CAT, no aportaran la información necesaria para que las estructuras quiméricas generadas se mantuvieran en la fracción soluble bajo las condiciones de expresión descritas en la sección de resultados. Podríamos suponer que el problema que presentan las quimeras radica únicamente en el transcurso del plegamiento. Sin embargo, la información contenida en su secuencia primaria (complementada por los fragmentos insertados) sería suficiente para mantener una conformación estable.

La identificación bioinformática de los fragmentos insertados en las ocho clonas positivas, reveló que no existe ninguna tendencia por algún elemento de estructura secundaria. Encontramos dos datos muy importantes, 7 de 8 insertos que complementaron el barril están en la misma fase que su proteína de origen, lo que confirma que estas secuencias ya habían sido previamente seleccionadas para plegarse. El caso de la clona Meg es distinto, esta secuencia se encuentra en una fase diferente a la de su proteína de origen. Es interesante que sólo una de las clonas incluya un fragmento fuera de fase de su proteína original, lo que indica que en la complementación se reclutaron secuencias que aportaron información relevante. No encontramos ningún patrón en las estructuras secundarias que complementaran el plegamiento de nuestro templado. Sin embargo, hallamos tres extremos amino terminales que se fusionaron con nuestro extremo carboxilo. Dado el tamaño tan pequeño del banco no podemos afirmar (no es estadísticamente significativo) que al utilizar un templado carboxilo terminal es mucho más probable recuperar aminos.

Como una perspectiva para continuar con el desarrollo de este trabajo se podrían manipular las condiciones de plegamiento *in vitro* para salvar el problema existente durante ese proceso. También contamos con la opción de resolubilizar las mutantes para obtener proteína en estado nativo y realizar pruebas biofísicas, como fue el caso de la quimera HisF-C\*C (Hocker 2004). Otro camino viable, es optimizar la construcción de los bancos de variantes y establecer el rango óptimo de concentración de antibiótico para encontrar una mayor proporción de verdaderos positivos. Una estrategia alternativa sería someter las mutantes a rondas de evolución para aumentar su estabilidad y solubilidad; esto pudo haber ocurrido naturalmente con muchos dominios estructurales existentes. Al analizar la información que se ha generado en los trabajos realizados alrededor del origen de los dominios, podemos darnos cuenta que la recombinación no homóloga es capaz de generar proteínas meta-estables que van estabilizándose gradualmente de forma natural por procesos de mutación espontánea.

La investigación aquí descrita, forma parte de un proyecto global que trata de responder varias preguntas. Por una parte, se explora la plasticidad estructural del barril TIM al remover una unidad  $(\beta/\alpha)$  y complementarla con fragmentos al azar (Sánchez 2006). Por la otra, mimetizamos lo que pudo ser el camino evolutivo de este plegamiento al complementar la unidad  $(\beta/\alpha)_4$  con secuencias que aportaron la información faltante para alcanzar una estructura estable. Desde el punto de vista básico aportamos información relevante para desentrañar el origen de los dominios como unidades evolutivas, y en términos de biología aplicada, aumentamos nuestras posibilidades de generar estructuras versátiles que sean útiles en futuros experimentos de evolución dirigida.

## Conclusiones

- Es posible obtener proteínas plegadas usando una unidad  $(\beta/\alpha)_4$  como templado en nuestro sistema de selección.
- Los elementos de estructura secundaria que complementaron nuestro medio barril resultaron muy diversos; aunque seleccionamos más extremos amino terminales.
- Los resultados confirman que los dominios proteicos pueden haber surgido de la fusión de fragmentos más pequeños.
- La recombinación no homóloga es capaz de generar diversidad estructural.
- Con nuestro sistema logramos seleccionar secuencias que presentaron información relevante para alcanzar una estructura tridimensional, ya que la mayoría de ellas complementaron el medio barril en la misma fase que en su proteína original.

## Apéndice

Inserto	Secuencia nucleótidos	Secuencia aminoácidos
N14	ATGCGTGTACTGGGTATTGAAACTTCC TGCGATGAAACCGGCATCGCCATTTAC GACGATGAAAAAGGTTTGTAGCCAAC CAATTGTATAGTCAGGTGAAATTGCAC GCTGACTACGGCGGCGTCGTGCCTGAA CTGGCCTCCCGCGATCAA	MRVLGIETSCDETGIAIYDDEKGLLANQ LYSQVKLHADYGGVPELASRDQ
Q1	ATGAGGAGTCTGCTTATGACCCGTCTG GCTTCGCGCTTTGGCGCAGCAAACCTT ATCCGTCGCGACCGTCCGTTAACCCGT GAAGAGCTGTTTCGCGTAGTGCCCAGT GTATTCAGTGAGGACAAACACGAGTCC CGTAGTGAGCAATCGCCAGATAACGGT ACATCTCCA	MRSLLMTRLASRFGAANLIRDRPLTRE ELFRVVPVSVFSEDKHESRSEQSPDNGTSP
14s	ATGAATGCTTTTGACCCGTCGCAGGGC AATTTACTTTTACCCCGCCTGGAATGG CGACGCTGTGGCGGTAAAGCCACGCTG CGGCTGACGCTATTCAGCGAAAGCTCC CTTCAGCACGATGCGATTTCAGGCA	MNAFDPSQGNLLLPRLEWRRCGGKATL RLTLFSESSLQHDAIQA
20s	ATGCCTTTTCAACAAAGCGTGAT TACCGCTGAGATAGAACGCTGGC CGGAAACCGCATTAAACCCGCGAG CTGGCGCGCCATCCGGCCTTTGTA	MPFQQSVITAEIERWPETALTRELARHP AFV

Tabla suplementaria 1. Secuencia de los insertos encontrados en 4 de las clonas generadas en este trabajo (N14, Q1, 14s, 20s).

Inserto	Secuencia nucleótidos	Secuencia aminoácidos
Q5	ATGAATCGCAACGAGTTGTTTAC CGCAATTTGTAGCGGGATGGCGT CCATTGCTGGTTCGATGATGATTGGA GCT	MNRNELFTAICSGMASIAGSMMIGA
1a	ATGGGGACGTTAAGCGGTGCGCC GAAAGTACGCGCTATGCAGTTAA TTGCCGAGGCGGAAGGTCGTCGC CGCGGCAGCTACGGCGGCGCGGT AGTTATTTCAACCGCGCATGGCGAA	MGTLSGAPKVRAMQLIAEAEGRRRGSY GGAVGYFTAHE
Meg	ATGCGGTGGCAGAGTGATGGATTGCT TCAACTGACCAAACGAGGAGATTTCC AGCGGTTTCTCGCCAGCGTTCTGCACG TTGTAGTTGACGTTGACAGCGTAATC ACCACGTTTCAGGACAAACGTTTTGG TAAACGTGTTGCCTGCCGCGTCGGTAT ACGTCATCGGCA	MRWQSDGLLQLTKRGDFQRFLASVLHV VVDVDSVITTFQDKRFGRVACRVGIR HRQ
17s	ATGAGATCCTTTTTGAAAAGGACGCA ATGCGCTCAGTCGCGCTCCACGCCAT CGTTGATTTTTTCAACAGGCATAACTC ACCCAGCGGAAACCGCTCTACAGAGG TTTAAATTTCTTATGTACCAGGTTGTT GCGTCTGATTTAGATGGCACGTTA	MRSFLKRTQCAQSRSTPSLIFSTGITHPA ETALQRFKFLMYQVVASDLDTL
PRAI Consenso	ATGCTGGAGAATAAGGTATGTGGCCT GACGCGTGGGCAAGATGCTAAAGCAG TATATAAGGCGGGCGGATTTACGGT GGGTTGATTTTTGTTAAGTCATCAAAA CGTTATGTCAGCGTTAAACAGGCGCA GGAAATCATGGCTGCGGCACCGTTGC GCTTTGTTGGCGTGTTCAAAATCACG ATATTGCCGATATCGTGGACATTGCTA AGAAGTTATCGCTGTACGCAGTGCAA CTGCATGGTAATGAAGATCAGGAGTA TATCGATCAACTGCGTAA AGCTTTGCCAAAGCATATCGCGATCT GGAAAGCAATCAGCGTCGGTAGCAC CCTGCCCGCCCTGAATTTTGATAATA ATGATAAATATATCTTCGACAGCCAA CAGGGTGGAAAGCGGGAAGAGCTTTGA CTGGTCACTATTAATGGTCAAATCCT TGGCAACGTTATTCTGGCGGGGGGCT TAAATGCAGATAACTGCGTGCAGGCG GCAAAGCTGGGCTGCCTGGGACTTGA TTTTAATTCTGGGGTAGAGTCGAGCCC GGGCATCAAAGACGCAAAGAAGATC AAGTCGGTTTTCCAGAAGCTGCGCGC ATAC	MLENKVCGLTRGQDAKAVYKAGAIYG GLIFVKSSKRYVSVKQAQEIMAAAPLRF VGVFQNHDIADIVDIKKLSLYAVQLHG NEDQEYIDQLRKALPKHIAIWKAISVGS TLPALNFDNNDKYIFDSQQGGSGKSF WSLLNGQILGNVILAGGLNADNCVQAA KLGCLGLDFNSGVESPPGIKDAKKIKSV FQKLAY

Tabla suplementaria 2. Secuencia de los insertos encontrados en 4 de las clonas generadas en este trabajo (Q5, 1a, Meg y 17s), y PRAI consenso, la proteína que utilizamos como andamiaje.

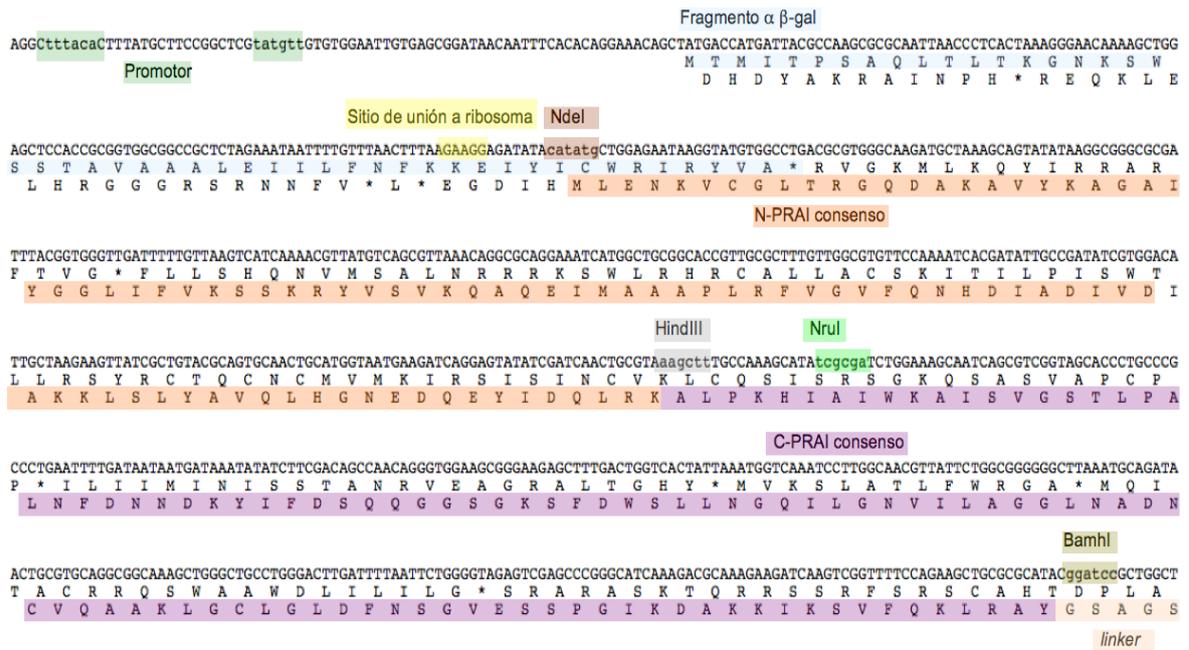


Figura suplementaria 1. Región corriente arriba de la construcción empleada en este trabajo. En código de colores se muestran los elementos reguladores, así como los sitios de restricción utilizados, y dos marcos de lectura distintos. También se muestran los sitios HindIII y NruI que permiten digerir el gen que codifica para PRAI consenso en dos mitades distintas. El sitio elegido para generar las unidades ( $\beta/\alpha$ )<sub>4</sub> fue HindIII. Se muestra un fragmento del *linker* (rosa claro) que une la región carboxilo terminal del gene de PRAI consenso con la región amino del gene que codifica para el reportero de plegamiento CAT

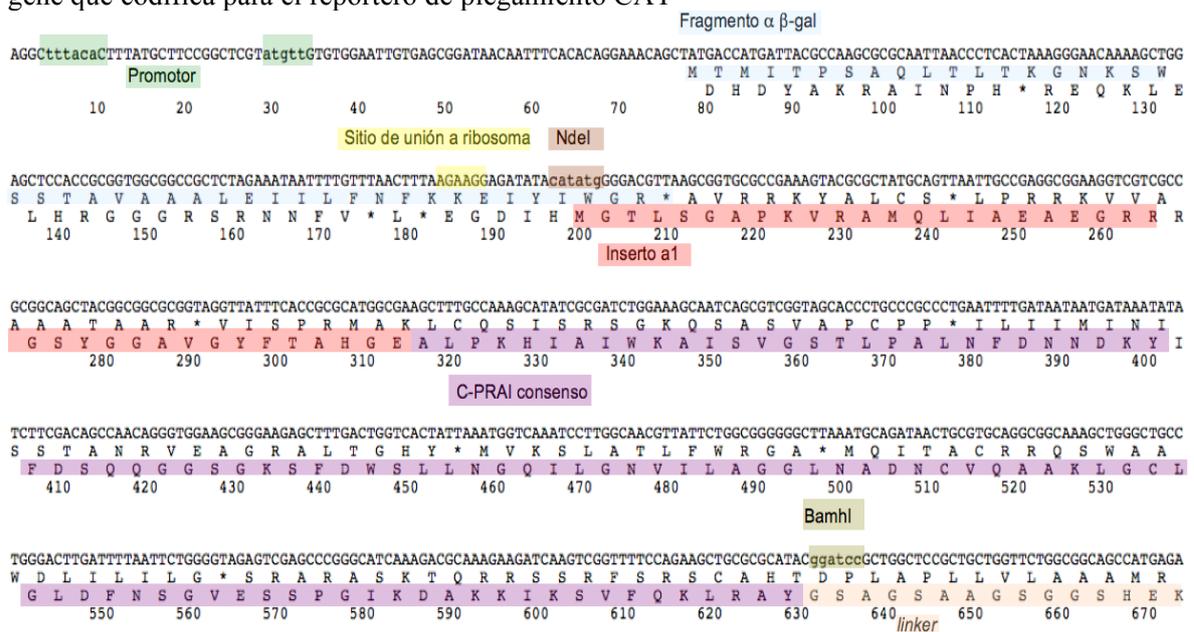


Figura suplementaria 2. Aspecto de la región corriente arriba de la clona positiva a1. Al digerir el gen que codifica para PRAI consenso con las enzimas NdeI y HindIII se remueve la mitad amino terminal del mismo; es en esta región donde se clonó el banco de fragmentos. En rojo claro se muestra la traducción del inserto encontrado en la clona a1. También se muestra un fragmento de la traducción del *linker* que une la región carboxilo terminal del gen de PRAI consenso con la región amino del gen que codifica para el reportero de plegamiento CAT.

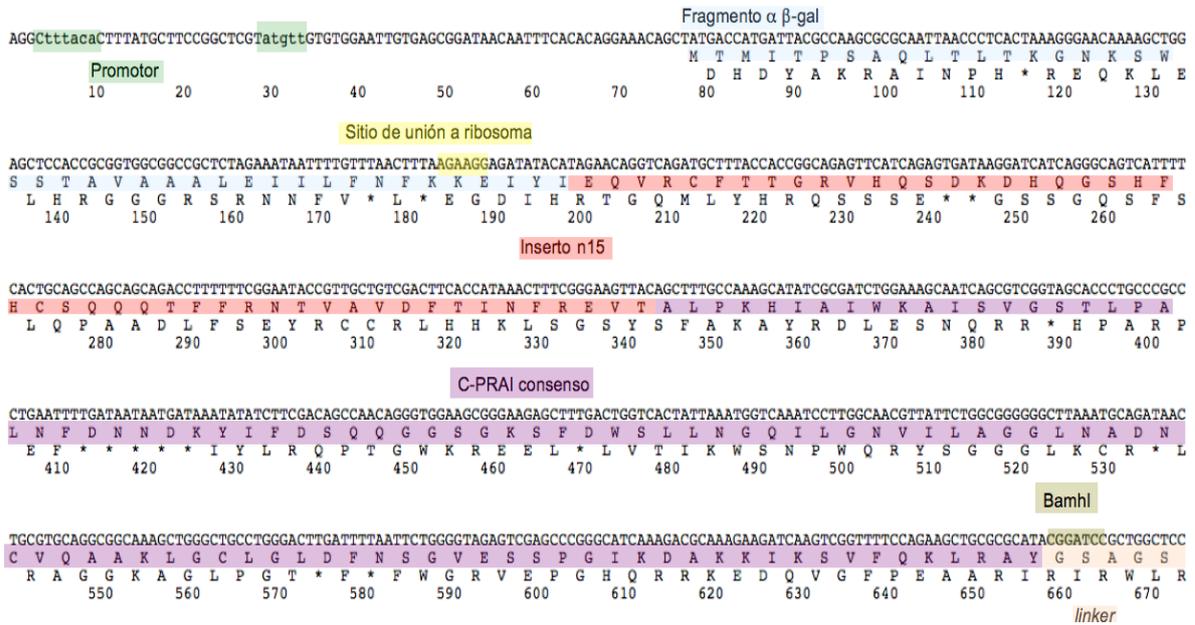


Figura suplementaria 3. Se muestra la región corriente arriba de una de las clonas fusionadas a lac. En azul claro se muestra el fragmento de lac que se fusiona en fase con el inserto clonado a partir del banco de secuencias al azar. La fusión comprende: a) fragmento de lac, b) inserto clonado a partir de las secuencias al azar, c) región carboxilo terminal del gen que codifica para PRAI consenso, d) gen que codifica para el reportero de plegamiento CAT.

## Referencias

- Acebron, S. P., V. Fernandez-Saiz, et al. (2007). "DnaJ recruits DnaK to protein aggregates." J Biol Chem.
- Alexander, P. A., Y. He, et al. (2007). "The design and characterization of two proteins with 88% sequence identity but different structure and function." Proc Natl Acad Sci U S A **104**(29): 11963-8.
- Alexey G. Murzin, J.-M. C., Antonina Andreeva, Dave Howorth, Loredana Lo Conte, Bartlett G. Ailey, Steven E. Brenner, Tim J. P. Hubbard, and Cyrus Chothia (2007). SCOP: Structural Classification of Proteins.
- B. Rost, G. Y. a. J. L. (2004). "The PredictProtein Server. Nucleic Acids Research 32." from <http://www.predictprotein.org/>.
- Coulson, A. F. and J. Moult (2002). "A unifold, mesofold, and superfold model of protein fold use." Proteins **46**(1): 61-71.
- de Bono, S., L. Riechmann, et al. (2005). "A segment of cold shock protein directs the folding of a combinatorial protein." Proc Natl Acad Sci U S A **102**(5): 1396-401.
- Doolittle, R. F. (2002). "Biodiversity: microbial genomes multiply." Nature **416**(6882): 697-700.
- Escobar, V. (2007). Exploración del espacio de secuencia de la Beta lactamasa: Obtención de mutantes resistentes a cefotaxima por supresión de mutantes inactivantes. . F. J.
- Fulop V, Jones DT. (1999)  $\beta$  Propellers: structural rigidity and functional diversity. Curr Opin Struct Biol 9:715 – 721.62.
- Geoffrey S. Waldo, Blake M. Standish, Joel Berendzen & Thomas C. Terwilliger (1999) Rapid protein-folding assay using green fluorescent protein Nature Biotechnology 17, 691 - 695
- Gerlt, J. A. and P. C. Babbitt (2001). "Barrels in pieces?" Nat Struct Biol **8**(1): 5-7.
- Grant, A., D. Lee, et al. (2004). "Progress towards mapping the universe of protein folds." Genome Biol **5**(5): 107.
- Grishin NV. (2001) KH domain: one motif, two folds. Nucleic Acids Res. Feb 1;29(3):638-43.
- Hocker, B., S. Beismann-Driemeyer, et al. (2001). "Dissection of a (betaalpha)<sub>8</sub>-barrel enzyme into two folded halves." Nat Struct Biol **8**(1): 32-6.
- Hocker, B., J. Claren, et al. (2004). "Mimicking enzyme evolution by generating new (betaalpha)<sub>8</sub>-barrels from (betaalpha)<sub>4</sub>-half-barrels." Proc Natl Acad Sci U S A **101**(47): 16448-53.
- Holm, L. and C. Sander (1996). "Mapping the protein universe." Science **273**(5275): 595-603.
- Hughes, A. L. (2005). "Gene duplication and the origin of novel proteins." Proc Natl Acad Sci U S A **102**(25): 8791-2.
- Koonin, E. V., Y. I. Wolf, et al. (2002). "The structure of the protein universe and genome evolution." Nature **420**(6912): 218-23.
- Lupas, A. N., C. P. Ponting, et al. (2001). "On the evolution of protein folds: are similar motifs in different protein folds the result of convergence, insertion, or relics of an ancient peptide world?" J Struct Biol **134**(2-3): 191-203.
- Maxwell, K. L., A. K. Mittermaier, et al. (1999). "A simple in vivo assay for increased protein solubility." Protein Sci **8**(9): 1908-11.
- Nagano, N., C. A. Orengo, et al. (2002). "One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions." J Mol Biol **321**(5): 741-65.
- Orengo, C. A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B., and Thornton, J.M. (2008). CATH Protein Structure Classification.
- Petsko, G. and Ringe, D. Protein Structure and Function 2004. Press Ltd.
- Riechmann, L. and G. Winter (2000). "Novel folded protein domains generated by combinatorial shuffling of polypeptide segments." Proc Natl Acad Sci U S A **97**(18): 10068-73.
- Riechmann, L. and G. Winter (2006). "Early protein evolution: building domains from ligand-binding polypeptide segments." J Mol Biol **363**(2): 460-8.
- Rost, B. (2002). "Did evolution leap to create the protein universe?" Curr Opin Struct Biol **12**(3): 409-16.
- Rostand, J. (1985). Introducción a la Historia de la Biología Barcelona Planeta Agostini
- Sambrook, P. M., David Russell (2000). Molecular Cloning: A Laboratory Manual.
- Sanchez, L. (2006). Obtención de proteínas plegadas por complementación con secuencias genómicas Ingeniería celular y Biocatálisis Cuernavaca Morelos Universidad Nacional Autónoma de México **Bachelor**.

- Seitz, T., M. Bocola, et al. (2007). "Stabilisation of a (beta/alpha)<sub>8</sub>-barrel protein designed from identical half barrels." *J Mol Biol* **372**(1): 114-29.
- Shukla, A. and P. Guptasarma (2004). "Folding of beta/alpha-unit scrambled forms of *S. cerevisiae* triosephosphate isomerase: Evidence for autonomy of substructure formation and plasticity of hydrophobic and hydrogen bonding interactions in core of (beta/alpha)<sub>8</sub>-barrel." *Proteins* **55**(3): 548-57.
- Soberon, X., P. Fuentes-Gallego, et al. (2004). "In vivo fragment complementation of a (beta/alpha)<sub>8</sub> barrel protein: generation of variability by recombination." *FEBS Lett* **560**(1-3): 167-72.
- Soding, J. and A. N. Lupas (2003). "More than the sum of their parts: on the evolution of proteins from peptides." *Bioessays* **25**(9): 837-46.
- Torsten S., M. C. P. N. G. (1993). "SWISS-MODEL An Automated Comparative Protein Modelling Server", from <http://swissmodel.expasy.org/SWISS-MODEL.html>.
- Vega, M. C., E. Lorentzen, et al. (2003). "Evolutionary markers in the (beta/alpha)<sub>8</sub>-barrel fold." *Curr Opin Chem Biol* **7**(6): 694-701.
- Venter, J. C., K. Remington, et al. (2004). "Environmental genome shotgun sequencing of the Sargasso Sea." *Science* **304**(5667): 66-74.
- Vogel, C., M. Bashton, et al. (2004). "Structure, function and evolution of multidomain proteins." *Curr Opin Struct Biol* **14**(2): 208-16.
- Voigt, C. A., C. Martinez, et al. (2002). "Protein building blocks preserved by recombination." *Nat Struct Biol* **9**(7): 553-8.
- Wierenga, R. K. (2001). "The TIM-barrel fold: a versatile framework for efficient enzymes." *FEBS Lett* **492**(3): 193-8.
- Wolfe, K. H. and D. C. Shields (1997). "Molecular evidence for an ancient duplication of the entire yeast genome." *Nature* **387**(6634): 708-13.