



UNIVERSIDAD NACIONAL AUTÓNOMA DE
MÉXICO

FACULTAD DE CIENCIAS

OBTENCIÓN DE PROTEÍNAS PLEGADAS POR
COMPLEMENTACIÓN CON SECUENCIAS
GENÓMICAS

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

BIÓLOGA

P R E S E N T A :

LORENA PAULINA SÁNCHEZ SÁNCHEZ



Tutor: Dr. Lorenzo Patrick Segovia Forcella

2006



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

FACULTAD DE CIENCIAS

División de Estudios Profesionales



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

ACT. MAURICIO AGUILAR GONZÁLEZ
Jefe de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Por este medio hacemos de su conocimiento que hemos revisado el trabajo escrito titulado:

Obtención de proteínas plegadas por complementación con secuencias genómicas
realizado por Lorena Paulina Sánchez Sánchez
con número de cuenta 402049474, quien cubrió los créditos de la licenciatura en Biología

Dicho trabajo cuenta con nuestro voto aprobatorio.

Tutor (a) Propietario Dr. Lorenzo Patrick Segovia Forcella

Propietario Dr. Daniel Alejandro Fernández Velasco

Propietario Dr. Luis José Delaye Arredondo

Suplente Quim. Viviana Escobar Sánchez

Suplente Dr. Víctor Manuel Valdés López

Atentamente
"POR MI RAZA HABLARÁ EL ESPÍRITU"
Ciudad Universitaria, D.F., a 9 de noviembre del 2006
CONSEJO DEPARTAMENTAL DE BIOLOGIA

Dr. Zenón Cano Santana

FACULTAD DE CIENCIAS



UNIVERSIDAD DE ENSEÑANZA

A mi papá, que me enseñó a ser mejor persona, y que aunque no está ya físicamente en este mundo, lo llevo siempre conmigo.

AGRADECIMIENTOS

A mi mamá, por darme ánimos y apoyarme siempre

A mi familia, en especial a mi hermano y mis abuelos

A mis amigos de la Facultad y del IBT, por hacerme pasar muy buenos momentos

A mi tutor, Lorenzo, por haberme aceptado en su grupo y tener tan buenas ideas

Y a mis compañeros en Cuernavaca, en especial a Viviana y a Luis por haberme enseñado lo necesario para que mi proyecto saliera adelante.

ÍNDICE

RESUMEN.....	1
INTRODUCCIÓN.....	2
1. EVOLUCIÓN DE PROTEÍNAS	2
1.1 Evolución del Proteoma.....	2
1.2 Origen de los Dominios Proteicos.....	4
1.3 Diversificación de los Dominios Proteicos.....	6
2. BARRILES (β/α) ₈	9
2.1 Generalidades.....	9
2.2 Fosforibosil Antranilato Isomerasa.....	12
OBJETIVOS.....	15
MATERIALES Y MÉTODOS.....	16
PLÁSMIDOS Y CEPAS.....	16
1. PERFIL DE INTERACCIONES Y ALINEAMIENTOS MÚLTIPLES....	16
2. CONSTRUCCIÓN DEL BANCO.....	16
2.1 Determinación de la CMI de la PRAI y PRAI-CAT silvestres.....	16
2.2 Vector.....	17
2.3 Inserto.....	18
2.4 Ligación.....	20
2.5 Transformación de Células por Electroporación	20
2.6 Selección.....	20
2.7 Secuenciación.....	21
3. CARACTERIZACIÓN DE LAS CLONAS.....	21
3.1 Caracterización Bioinformática.....	21
3.2 CMI de las Clonas.....	22
4. EXTRACCIÓN DE PROTEÍNAS TOTALES Y WESTERN BLOT.....	22
5. PROTOCOLO PARA DESFUSIONAR CAT.....	23
6. ACTIVIDAD <i>IN VIVO</i> DE PRAI.....	23

RESULTADOS Y DISCUSIÓN.....	25
1. ELECCIÓN DE UN FRAGMENTO DE LA PRAI A SUSTITUIR.....	25
2. OBTENCIÓN DEL BANCO.....	27
3. ANÁLISIS DE LAS CLONAS OBTENIDAS.....	29
4. CARACTERIZACIÓN DE LA PROTEÍNA C5.....	34
 CONCLUSIONES.....	 39
 ANEXO 1.....	 41
 BIBLIOGRAFÍA.....	 42

RESUMEN

La recombinación de genes no homólogos ha tomado gran importancia en los últimos años ya que se ha encontrado evidencia tanto a nivel teórico como experimental que este proceso es una herramienta eficiente de innovación estructural, dando lugar tanto a nuevos plegamientos como a funciones *de novo* en las proteínas. Mediante eventos de recombinación, podrían ser movilizados fragmentos de DNA para generar proteínas plegadas solubles, lo cual incrementaría el potencial evolutivo de los genomas.

El objetivo de este trabajo es el de simular estos eventos de recombinación no homologa con el fin de sustituir con secuencias de DNA genómico un fragmento de la Fosforibosil Antranilato Isomerasa (PRAI) de *Escherichia coli* y obtener proteínas quiméricas plegadas.

Se eligió la PRAI de *E. coli* para llevar a cabo este proyecto ya que es parte de la superfamilia de los barriles $(\beta/\alpha)_8$, un plegamiento particularmente prominente en términos de su abundancia y diversidad funcional, lo que los hace un andamiaje ideal como punto de partida para el campo de la ingeniería de proteínas.

Se generó un banco en donde se encontró una gran diversidad de secuencias que son capaces de complementar el plegamiento del barril $(\beta/\alpha)_8$. De 10^3 recombinaciones el 2% de ellas fueron exitosas dando lugar a un nuevo dominio plegado.

Aunque este trabajo se centra en entender la importancia de la recombinación no homologa en la generación de nuevas estructuras proteicas, los resultados obtenidos facilitarán también el desarrollo de proteínas *in vitro* útiles para el ser humano.

INTRODUCCIÓN

La importancia de las proteínas se encuentra implícita en el significado de su nombre, proveniente del griego *proteios*, que significa “en primer lugar”. Las proteínas desempeñan papeles cruciales en prácticamente todos los procesos biológicos y son los instrumentos moleculares a través de los cuales la información genética es expresada.

Además de ser las macromoléculas más abundantes dentro de las células, constituyendo más de la mitad del peso seco en la mayoría de los organismos, las proteínas son enormemente diversas en cuanto a secuencia y versatilidad de funciones; si tomamos en cuenta la cantidad de especies que existen (10 millones) y que cada una de éstas contiene miles de genes que codifican para una proteína (5,000 genes en promedio), tendríamos un total de 5×10^{10} proteínas con secuencias únicas (Koonin *et al.*, 2002).

Dentro del vasto mundo de las proteínas, se encuentran las enzimas, catalizadores altamente eficientes y específicos que realizan la mayoría de las reacciones químicas que se llevan a cabo dentro de las células. Es importante no sólo entender como la estructura, función y estabilidad de las enzimas están ligadas, sino también comprender como surgieron las proteínas modernas a partir de precursores más sencillos. La ingeniería de proteínas es un campo que está creciendo rápidamente y que nos permitirá dar respuesta a éstas y muchas otras preguntas, que tendrán impacto tanto a nivel de investigación básica como al nivel de la ciencia aplicada.

1. EVOLUCIÓN DE PROTEÍNAS

1.1. EVOLUCIÓN DEL PROTEOMA

Debido a que las proteínas son los agentes sobre los que se basa la vida, su evolución ha sido objeto de un estudio exhaustivo.

Los dominios proteicos representan la unidad evolutiva básica que forma a las proteínas. Un dominio se puede definir como una unidad estructural que tiene la

capacidad de plegarse independientemente de otras unidades (Koonin *et al.*, 2002). Sin embargo las definiciones varían y en algunos casos los dominios pueden ser descritos como regiones proteicas con funciones asignadas experimentalmente.

Los dominios pueden encontrarse en la naturaleza ya sea de manera aislada o como parte de un complejo más grande conformando una proteína multidominio o proteína mosaico. Los dominios encontrados en estas proteínas son evolutivamente móviles, lo que quiere decir que se han expandido durante el transcurso de la evolución y en la actualidad están presentes en proteínas que no están relacionadas. Estos dominios están frecuentemente involucrados en funciones específicas que contribuyen a la actividad de la proteína completa. (Kolkman y Stemmer, 2001)

Se cree que las proteínas multidominio han surgido por la duplicación y el barajeo por recombinación de dominios existentes, probablemente las dos fuerzas evolutivas más importantes que dirigen la evolución del proteoma. La duplicación resulta en la expansión de un dominio en términos de su abundancia y ha sido asociada al incremento en la complejidad evolutiva de un organismo; mientras que la recombinación de dominios podría ser el mecanismo más importante por el cual se modifica la función de una proteína dándole versatilidad al proteoma. (Vogel *et al.*, 2002). (Fig 1)

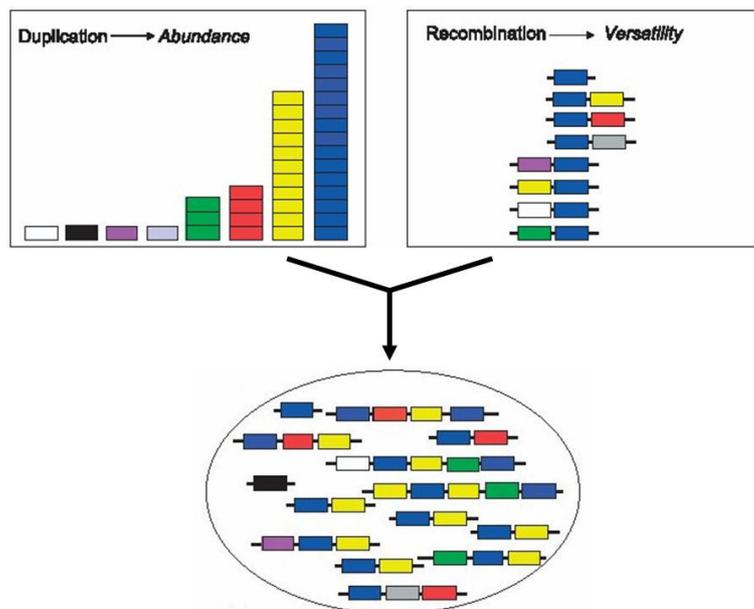


Fig. 1. Esquema en donde se representa el efecto de la duplicación y recombinación en la formación de proteoma. (Modificada de Vogel *et al.*, 2002)

1.2. ORIGEN DE LOS DOMINIOS PROTEICOS

Así como se ha propuesto que las proteínas multidominio surgieron por la fusión de estructuras más sencillas, se ha sugerido que los primeros dominios proteicos surgieron por la recombinación y ensamblaje de un conjunto limitado de fragmentos recurrentes que funcionan como elementos estructurales capaces probablemente, de plegarse de manera autónoma. Estas estructuras son llamadas estructuras supersecundarias, de los cuales los “ $\beta\beta$ -hairpins”, los “ $\alpha\alpha$ -hairpins” y los elementos $\beta\alpha\beta$ son los más comunes (Söding y Lupas, 2003) (Fig. 2), abarcando en un plegamiento promedio aproximadamente el 60% de la estructura. Se ha observado que un cuarto de todos los dominios con estructura conocida, asumen uno de los diez plegamientos más frecuentes. Uno de estos plegamientos es el barril $(\beta/\alpha)_8$ que cuenta con alrededor de un 80% de estructura supersecundaria. (Salem *et al.*, 1999).

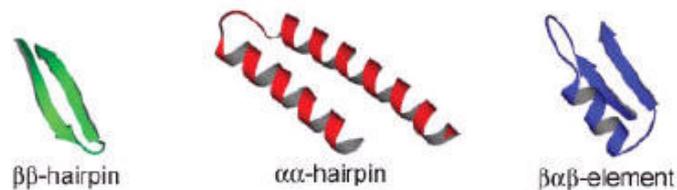


Fig. 2. Representación de las 3 estructuras supersecundarias más abundantes dentro de los dominios. (Tomada de Salem *et al.*, 1999)

Con la formación de estas estructuras locales se esperaría facilitar el plegamiento de toda la proteína ya que se restringiría el espacio conformacional que tendría que ser muestreado. Sin embargo, aunque se han identificado experimentalmente estas estructuras, ha sido difícil vislumbrar el rol que juegan en la evolución de los dominios proteicos. (Bono *et al.*, 2005)

La hipótesis antes mencionada suena atractiva ya que por una parte resulta mucho más probable que el proceso evolutivo haya comenzado con la síntesis de pequeñas cadenas polipeptídicas que con largas cadenas de residuos de aminoácidos (Keller *et al.*, 1994). Además, resulta más eficiente armar dominios por el ensamblaje de pequeños módulos optimizados individualmente que la evolución *de novo* de un dominio completo; lo que sustenta la idea de que la modularidad en los sistemas naturales existe en muchos

niveles incluyendo el nivel proteico. Por último, se han encontrado un gran número de fragmentos provenientes de proteínas no homólogas que son similares en secuencia, estructura y en algunos casos en función, los cuales podrían representar remanentes estructurales de un mundo pre-dominio. Uno de estos casos es el “Asp-box” un “ β hairpin” que se ha encontrado en más de 9 familias de proteínas. Este motivo presenta una similitud significativa tanto a nivel de secuencia como a nivel estructural, a pesar de estar localizados en estructuras proteicas con distintas secuencias y plegamientos (Kinch y Grishin, 2002).

Se ha sugerido que la distribución de los plegamientos está dominada por 5 regiones densamente pobladas, llamadas atractores (*Fig. 3*). Estos atractores representan por una parte vías dominantes de plegamiento y por otra pozas evolutivas que son el resultado de limitaciones físicas. Ésta hipótesis se sustenta en que los dominios contienen una composición similar en estructura secundaria y motivos topológicos característicos. Aproximadamente un 40% de todos los dominios conocidos están incluidos en 16 clases de plegamiento (“fold classes”), todas estas clases mapean en una de las 5 regiones atractoras.

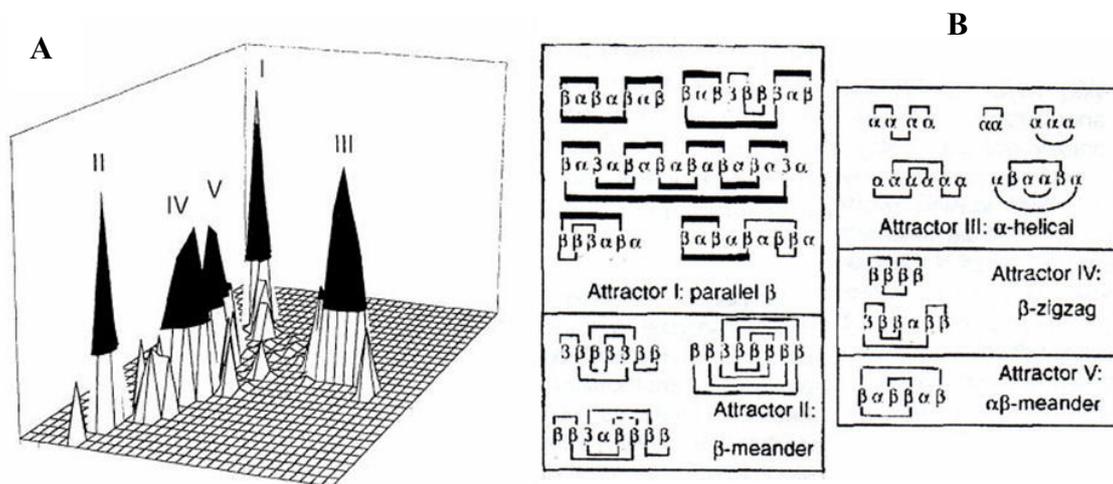


Fig. 3. Atractores del espacio de secuencia. A) La altura de los picos en la gráfica representa la densidad poblacional de estructuras proteicas para cada atractor. B) Cada atractor está definido por una cierta composición de estructuras secundarias similares y motivos topológicos característicos. (Tomada de Holm y Sander, 1996).

Todos los plegamientos que contienen hojas beta paralelas están incluidas dentro del atractor I, éstos plegamientos contienen una unidad $\beta x \beta$, en donde el segmento intermedio (x) es requerido para invertir la dirección de la cadena con el fin de que las hojas beta queden de manera paralela. El atractor II contiene una variedad de plegamientos helicoidales. La conectividad de los elementos en los plegamientos de los atractores III y IV sugiere el colapso de un “hairpin” largo, ya sea solamente de hebras beta o de hebras beta alternadas con una estructura $(\beta \alpha \beta)_2$. El motivo β zigzag del atractor V es una serie de conexiones de “hairpins” antiparalelos entre hebras adyacentes. (Holm y Sander, 1996).

1.3. DIVERSIFICACIÓN DE LOS DOMINIOS PROTEICOS

La diversificación de dominios existentes se puede explicar mediante una variedad de procesos que incluyen mutaciones puntuales, deleciones, inserciones, transferencias horizontales y recombinación tanto de genes homólogos como no homólogos (Bogarad y Deem, 1999)

La recombinación de genes no homólogos ha tomado gran importancia en los últimos años ya que se ha encontrado evidencia tanto a nivel teórico como experimental que este proceso es una herramienta eficiente de innovación estructural, dando lugar tanto a nuevos plegamientos como a funciones *de novo* en las proteínas. (Bogarad y Deem, 1999; Riechmann y Winter, 2000; Cui *et al.*, 2002).

Durante el curso del proceso evolutivo, las proteínas se han quedado atrapadas en mínimos de energía local. Para romper o salir de esas regiones y encontrar así nuevos plegamientos que conlleven a nuevas funciones, son necesarios movimientos dramáticos, como el ensamblaje y/o intercambio de pequeños fragmentos. Sin embargo, estos movimientos bruscos son usualmente deletéreos (*Fig. 4*). En la naturaleza, los mecanismos han evolucionado para incrementar las posibilidades de intercambios exitosos generando así una diversidad de plegamientos que se reflejan en las proteínas que conocemos hoy en día. Justamente la recombinación de genes no homólogos así como otros rearrreglos génicos permiten la transición de un plegamiento a otro y la diversificación de los dominios. (Bogarad y Deem, 1999)

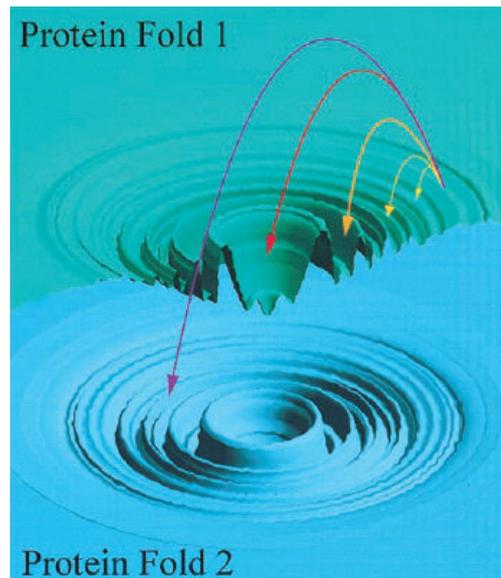


Fig. 4. Diagrama en donde se representa la habilidad de un proceso evolutivo a nivel molecular de traspasar las barreras energéticas que llevan a la diversificación de dominios. Se resalta la importancia de la recombinación no homóloga en la búsqueda del espacio de secuencia global (flecha morada), en comparación a otros eventos como las mutaciones puntuales (flecha amarilla), la recombinación homóloga a nivel funcional (flecha naranja) y estructural (flecha roja) que hacen una búsqueda del espacio de secuencia local.

Se ha observado que la tasa evolutiva a causa del “DNA swapping” es mayor que la que es provocada por mutaciones puntuales. Aun con el hecho de que un evento de “DNA swapping” es menos tolerado de lo que es una mutación puntual (Bogarad y Deem, 1999).

Durante la evolución de proteínas se movilizan fragmentos de DNA mediante eventos de recombinación para generar proteínas plegadas solubles, lo cual incrementa el potencial evolutivo de los genomas (Kolkman y Stemmer, 2001).

Recientemente se ha demostrado experimentalmente que se pueden generar nuevos dominios a través de combinatorias de segmentos polipeptídicos, independientemente de si éstos son homólogos o no. Riechmann y Winter recombinaron el DNA que codifica la parte N-terminal de la proteína A de choque frío (CspA: Barril β -antiparalelo) de *E. coli* con fragmentos al azar del DNA genómico de esta misma bacteria. Posteriormente clonaron los genes que codifican para el repertorio polipéptidos quiméricos y los desplegaron en fagos filamentosos seleccionando proteínas plegadas por su resistencia a proteólisis con tripsina (Riechmann y Winter, 2000). En este trabajo logran seleccionar, entre otros, un polipéptido que presentó el mismo plegamiento que

la proteína parental y otra con un plegamiento distinto. La resolución cristalográfica de ésta última, llamada 1b11 (Fig. 5), revela un barril beta con seis cadenas formando un tetrámero (Bono *et al.*, 2005).

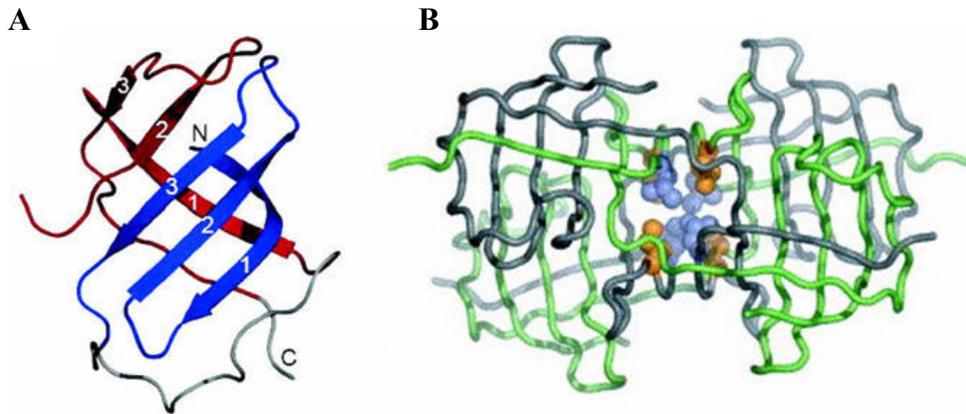


Fig. 5. Estructura de la proteína quimérica 1b11. A) Se muestra el barril en su forma monomérica (azul: segmento de la proteína proveniente de la CspA; en rojo: segmento proveniente del proteoma de *E. coli*). B) Tetrámero 1b11. (Tomada de Bono *et al.*, 2005)

Estos autores 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 pliegue. Asimismo, sus resultados sugieren que cualquier segmento de una secuencia del genoma tiene el potencial para producir dominios plegados con muchos otros segmentos del mismo genoma. En este experimento se lograron obtener entre 4 y 8 nuevas proteínas en 10^8 recombinaciones. Esto quiere decir, que dada la cantidad de enormes variaciones en cuanto al tamaño poblacional, el tiempo generacional (tiempo de duplicación) y los niveles de recombinación intracelular, una especie bacteriana sería capaz de crear cientos de miles de nuevas proteínas al año a través de la recombinación no homóloga de genes. Aunque sólo una pequeña fracción de estas proteínas se llegarán a establecer en la población, el efecto acumulativo a través del tiempo a escalas evolutivas sería substancial (Riechmann y Winter, 2000).

Aunque el panorama es complejo al tratar de explicar los mecanismos mediante los cuales ha surgido la gran diversidad de proteínas con las que contamos hoy en día; el entender la historia evolutiva de éstas nos acercará cada vez mas al diseño de proteínas con funciones biológicas nuevas.

2. BARRILES ($\beta\alpha$)₈

2.1 GENERALIDADES

Aunque no se sabe con exactitud, se ha estimado que existen en la naturaleza alrededor de 1000 familias de proteínas estructuralmente distintas (Chotia, 1992). De entre todas éstas, los barriles TIM o barriles ($\beta\alpha$)₈ son particularmente prominentes en términos de su abundancia y de su diversidad funcional, lo que los hace un andamiaje ideal como punto de partida para el campo de la ingeniería de proteínas.

Los barriles ($\beta\alpha$)₈ son el plegamiento más recurrente en el PDB (Protein Data Bank) y se ha estimado que entre el 10 y el 12% de todas las enzimas incluyen un dominio barril ($\beta\alpha$)₈ (Patrick y Blackburn, 2005). Asimismo, se han descrito barriles (β/α)₈ participando como oxidoreductasas, transferasas, liasas, hidrolasas e isomerasas, 5 de las 6 clases de reacciones enzimáticas (Nagano *et al.*, 2002). Las proteínas que poseen esta arquitectura están ampliamente distribuidas a lo largo de las vías del metabolismo central; en particular, en la vía de síntesis de triptofano tres de las seis reacciones enzimáticas son catalizadas por representantes del plegamiento barril ($\beta\alpha$)₈. (Fig. 6)

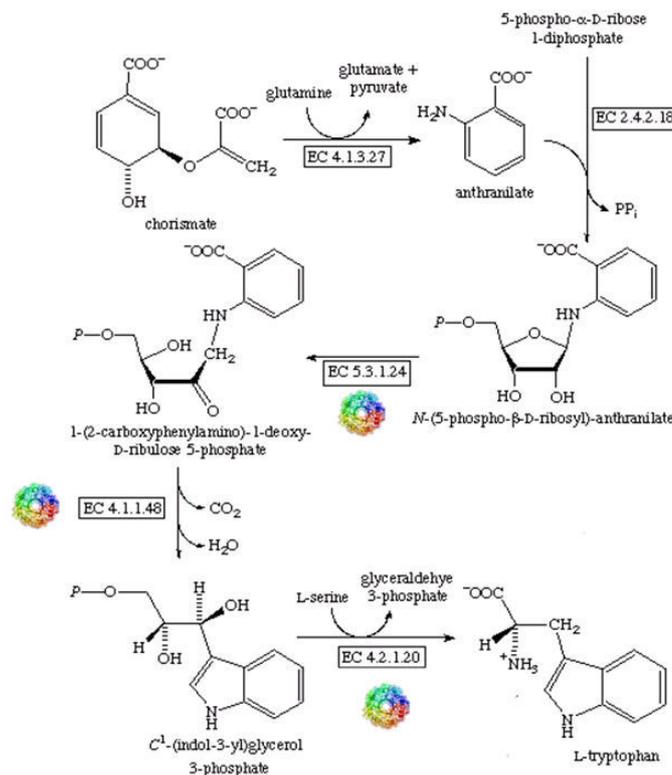


Fig. 6. Ruta metabólica de la biosíntesis de triptofano. Se muestran las reacciones catalizadas por barriles (β/α)₈.

La diversidad de funciones entre las proteínas barril ($\beta\alpha$)₈ se ha atribuido a la construcción aparentemente modular de su plegamiento. Los barriles TIM presentan una estructura de repeticiones de hebras beta formando un barril cerrado, los cuales están constituidos por 8 hebras β plegadas (capa interna) rodeadas por 8 hélices α (en contacto con el solvente) (Fig. 7A). En un barril (β/α)₈ cerrado se detecta un anillo completo de puentes de hidrogeno, cada hebra beta está formando puentes de hidrogeno con sus dos vecinas, todas están conectadas, arreglándose generalmente de manera paralela (Wierenga, 2001). Los sitios activos están localizados en la “cara” catalítica del barril (Fig. 7B), que comprende los extremos carboxilo terminales de las hebras β y las asas que unen las hebras β con las hélices α subsiguientes; por otro lado, las asas que unen las hélices α con las hebras β sucesivas, localizadas en la cara opuesta del barril, son importantes en la estabilidad del plegamiento (Höcker *et al.*, 2001).

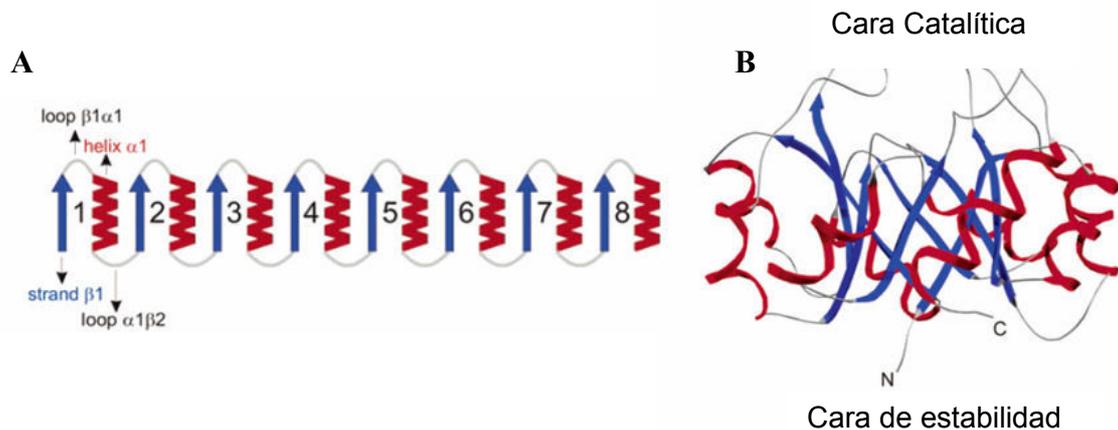


Fig. 7. Representación de un barril (β/α)₈. A) Esquema lineal de las 8 repeticiones (β/α) y las asas que las conectan. B) Estructura terciaria de un barril (β/α)₈. (Tomado de Höcker *et al.*, 2001)

El tamaño de un barril TIM varía de entre unos 150 a 500 residuos de aminoácidos con un promedio de 298. Se caracterizan por poseer una alta estabilidad intrínseca junto con una alta actividad catalítica (Höcker, 2001). Existen representantes monoméricas, aunque pueden ser utilizados en diferentes combinaciones junto con otros dominios para crear proteínas funcionales multidominio. Estos barriles tienen dominios adicionales que preceden, interrumpen y suceden al barril, lo que indica eventos de barajeo de dominios durante la evolución, aproximadamente el 8% del total de los barriles (β/α)₈ conocidos tienen insertado un dominio dentro de su estructura (Nagano *et al.*, 2002).

Parece ser que el plegamiento barril $(\beta/\alpha)_8$ no es dictado por detalles presentes en la secuencia, sino por características más generales como la distribución de residuos polares y no polares así como de aminoácidos cargados. Alineamientos estructurales revelan “clusters” de residuos similares en posiciones topológicamente equivalentes que podrían dirigir y estabilizar el patrón de plegamiento en los barriles $(\beta\alpha)_8$ (Sterner y Höker, 2005)

Se cree que la evolución divergente a partir de un ancestro común explica de mejor manera la evolución de los barriles TIM o por lo menos de la mayoría de las familias dentro del plegamiento (Copley y Bork, 2000). La presencia común del sitio de unión al fosfato es una evidencia importante que sustenta la monofilia en éste plegamiento.

Se ha sugerido que el plegamiento de barril TIM pudo haber surgido de manera modular a partir del ensamblaje de pequeños subdominios, (β/α) ancestrales, capaces de plegarse de manera autónoma. Experimentos llevados a cabo por el grupo de Guptasarma confirman que el plegamiento barril $(\beta/\alpha)_8$ puede tolerar profundas alteraciones en su estructura, debido a que existe una autonomía en la formación de las subestructuras (β/α) y una plasticidad en las interacciones tanto hidrofóbicas como de puentes de hidrogeno dentro del barril; ya que al barajar el orden original de las unidades (β/α) de la TIM (Triosa Fosfato Isomerasa) perteneciente a *S. cerevisiae*, se obtuvieron los siguientes barriles plegados: (2-4-6-8-1-3-5-7) y (1-3-5-7-2-4-6-8), e inclusive se obtuvieron mitades de barril (2-4-6-8) y cuartos de barril (1-3) que mantuvieron las interacciones estructurales del plegamiento a nivel terciario y cuaternario. (Shukla y Guptasarma, 2004)

Se ha propuesto que el ancestro de los barriles $(\beta/\alpha)_8$ pudo haber surgido por la duplicación de la mitad de un barril (*Fig. 8*). A partir de este punto duplicaciones y diversificaciones posteriores pudieron haber dado lugar a enzimas con actividades catalíticas distintas. Esta hipótesis se basa en un estudio llevado a cabo con dos barriles $(\beta/\alpha)_8$ HisA y HisF, que catalizan pasos consecutivos en la vía de síntesis de histidina; al comparar estas dos enzimas se encontró un patrón en cuanto a la secuencia y estructura común a las dos mitades en las dos enzimas (Lang *et al.*, 2000).

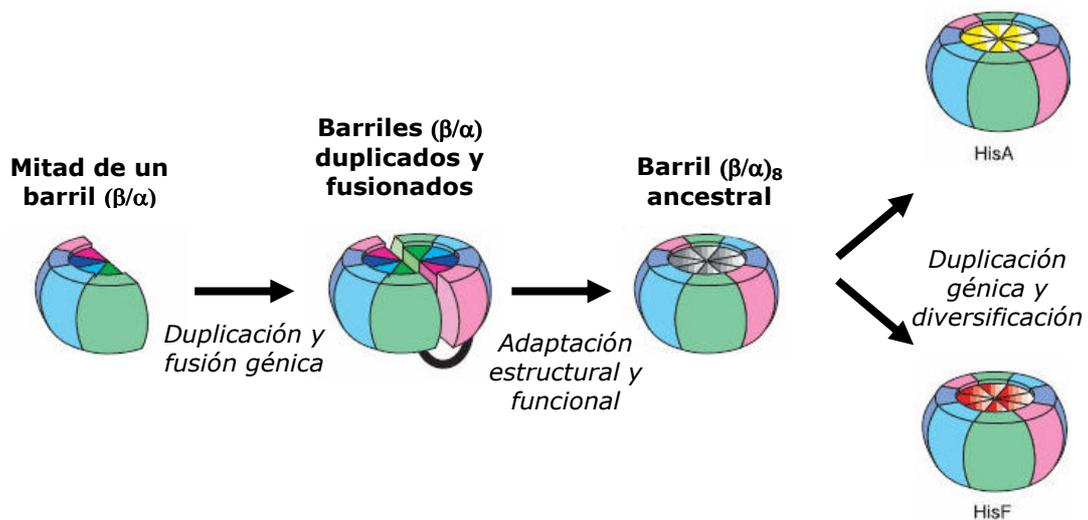


Fig. 8. Modelo evolutivo del surgimiento del plegamiento barril $(\beta\alpha)_8$ por duplicación génica.

Por otro lado y tomando como base que el estudio de los intermediarios en el plegamiento de una proteína nos pueden proporcionar información crucial de la relación entre la secuencia y la estructura de ésta, Zitzewitz y colaboradores, demostraron que en el caso de de la subunidad α de la triptofano sintasa de *E. coli*, una proteína perteneciente al plegamiento barril TIM, fragmentos mayores a 100 residuos de aminoácidos pueden comportarse como unidades que se pliegan independientemente. Asimismo, se sugiere que el plegamiento involucra un ensamblaje modular de elementos de estructura supersecundaria $\beta\alpha\beta$, en donde el rol de estos módulos locales de plegamiento es el de servir como armazones iniciales para que la proteína como tal adquiriera su estructura terciaria nativa. (Zitzewitz *et al.*, 1999).

2.2 FOSFORIBOSIL ANTRANILATO ISOMERASA

El barril TIM elegido para este proyecto es la enzima Fosforibosil Antranilato Isomerasa (PRAI por sus siglas en inglés) codificada por el gen *trpF*. Participa en el tercer paso en la vía de síntesis del triptofano, catalizando un rearrreglo de Amadori que resulta en la apertura de un anillo del azúcar ribosa (Fig. 9).

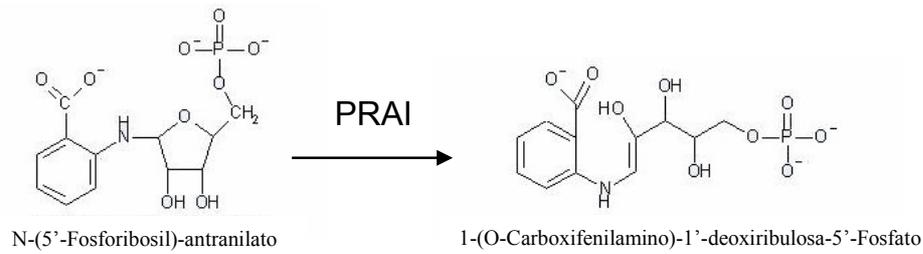


Fig. 9. Reacción enzimática catalizada por la PRAI

La PRAI de *E. coli* (ePRAI) se encuentra formando parte de una proteína bifuncional formada por dos dominios: PRAI-IGPS, en donde esta última cataliza la reacción subsiguiente en la vía. Nuestra proteína es una variante que ha sido separada para funcionar como una proteína monomérica. Tiene una extensión de 200 aminoácidos, es uno de los barriles TIM más pequeños que se conocen (Hommel *et al.*, 1995) y no tiene ningún elemento de estructura secundaria adicional (Fig. 10). La PRAI difiere de ser un barril TIM perfecto (ideal) en que presenta un giro helicoidal adicional en el asa que conecta la hebra β 8 con la hélice α 8; así como en la degeneración de la hélice α 5. (Jasanoff *et al.*, 1994)

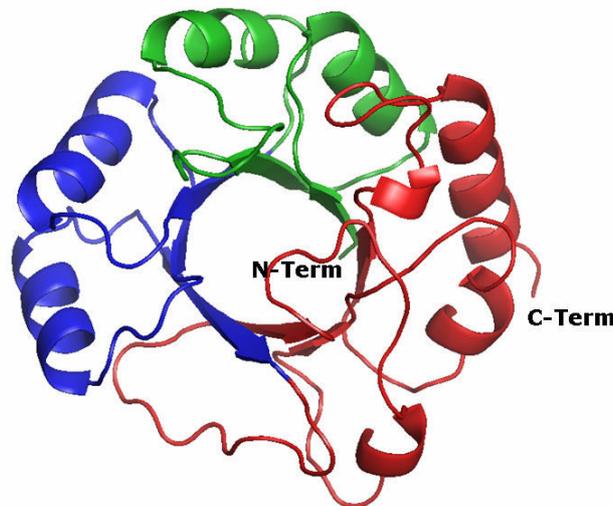


Fig. 10. Estructura terciaria de PRAI. Modificada de (Akanuma y Yamagishi, 2005)

En un estudio reciente se ha encontrado que ePRAI se puede dividir en tres porciones con respecto a su plegamiento. La porción C-terminal (en rojo) seguramente no es crucial para el plegamiento de la proteína, debido a que el fragmento $(\beta\alpha)_{1-4}\beta_5$ es suficientemente estable para retener su conformación nativa, siendo el segmento más pequeño capaz de plegarse de manera autónoma. Experimentos de permutación circular

indican que una porción N-terminal intacta (en verde) se requiere para la formación de la estructura global, pero no influye en la formación de estructuras secundarias; mientras que la porción central (en azul) correspondiente al segmento $(\beta\alpha)_{3-4}\beta_5$ contiene interacciones esenciales para los eventos tempranos en el plegamiento (sitio de nucleación) ya que cuando este fragmento es interrumpido la proteína no es capaz de plegarse. (Akanuma y Yamagishi, 2005). A pesar de lo encontrado en este trabajo, anteriormente se habían propuesto otros dos teorías para explicar el plegamiento de ePRAI: el primero en donde los primeros 6 módulos (β/α) se plegarían al principio seguido de los últimos dos módulos: modelo 6+2 (Eder y Kirschener, 1992), también encontrado en la subunidad α de la Triptofano sintasa de *E. coli* (Zitzewits y Matthews, 1999). Y el segundo, en donde se propone que las mitades del barril son capaces de plegarse autónomamente sugiriendo un modelo 4+4 (Soberón *et al.*, 2004). Otro trabajo que da soporte a este último modelo es el llevado a cabo con la proteína HisF de *T. maritima* (Höcker *et al.*, 2001)

Finalmente, es importante mencionar algunos estudios de ingeniería de proteínas con PRAI, para destacar las posibilidades de insertar un péptido ajeno en este barril $(\alpha\beta)_8$. Se han generado dos variantes de PRAI enzimáticamente activas con más de las 8 repeticiones $(\beta\alpha)$ que tiene originalmente. La primera de ellas partiendo de la PRAI de levadura cuenta con un duplicado de los últimos dos módulos $(\beta\alpha)$ carboxilo terminales convirtiéndolo en un barril $(\beta\alpha)_{10}$ (Luger *et al.*, 1990); y en el segundo caso se construyó una variante de ePRAI que contiene una duplicación en serie del quinto módulo $(\beta\alpha)$ (Urfer y Kirschner, 1992). Asimismo, La PRAI de *E. coli* ha sido permutada circularmente de dos diferentes maneras, y las enzimas resultantes han sido catalíticamente activas y prácticamente indistinguibles de la enzima silvestre por varios criterios estructurales y espectroscópicos (Luger *et al.*, 1989).

OBEJETIVO GENERAL

Obtener proteínas plegadas *de novo* por recombinación no homóloga utilizando como templado un miembro de la familia de los barriles ($\beta\alpha$)₈, la PRAI de *E. coli*.

OBJETIVOS PARTICULARES

- Sustituir con secuencias de DNA genómico un fragmento $\beta\alpha\beta$ de la PRAI de *E. coli* con el fin de obtener proteínas quiméricas plegadas.
- Caracterizar algunas de las clonas para validar el sistema de selección de las proteínas recombinantes plegadas.

MATERIALES Y MÉTODOS

PLÁSMIDOS

- pT4-PRAI-CAT (Resistencia a kanamicina, promotor pTRC) (Farias Rico, 2005)

CEPAS (*E. coli*)

- XLI-Blue: *F' ::Tn10 proA-B+lacI_qΔ(lacZ)M15/recA1 endA1 gyrA96 (Nal.) thi hsdR17 (r_K-m_{K+}) glnV44 relA1 lac*
- MC1061: *F' araD139 Δ(ara-leu)7696 galE15 galK16 Δ(lac)X74 rpsL (Str.) hsdR2 (r_K-m_{K+}) mcrA mcrB1*
- JM101: *F' traD36 proA⁺B⁺ lacI^q Δ(lacZ)M15/ Δ(lac-proAB) glnV thi*
- JM101 Δ: *F' traD36 proA⁺B⁺ lacI^q Δ(lacZ)M15/ Δ(lac-proAB) glnV thi trpF*

1. PERFIL DE INTERACCIONES Y ALINEAMIENTOS MÚLTIPLES

Se utilizó el programa SCHEMA (www.che.caltech.edu/groups/fha/) para generar un perfil de interacciones de PRAI (código pdb 1pii). Posteriormente se llevó a cabo un alineamiento estructural utilizando el programa MultiProt (<http://bioinfo3d.cs.tau.ac.il/MultiProt/>) de 12 proteínas homólogas a la PRAI según la base de datos CATH: 1glyA, 1d3hA, 1dqwA, 1ep3A, 1geqB, 1gpwA, 1bwkA, 1i4n, 1k3uA, 1rpxA, 1tpfA, 1nsj.

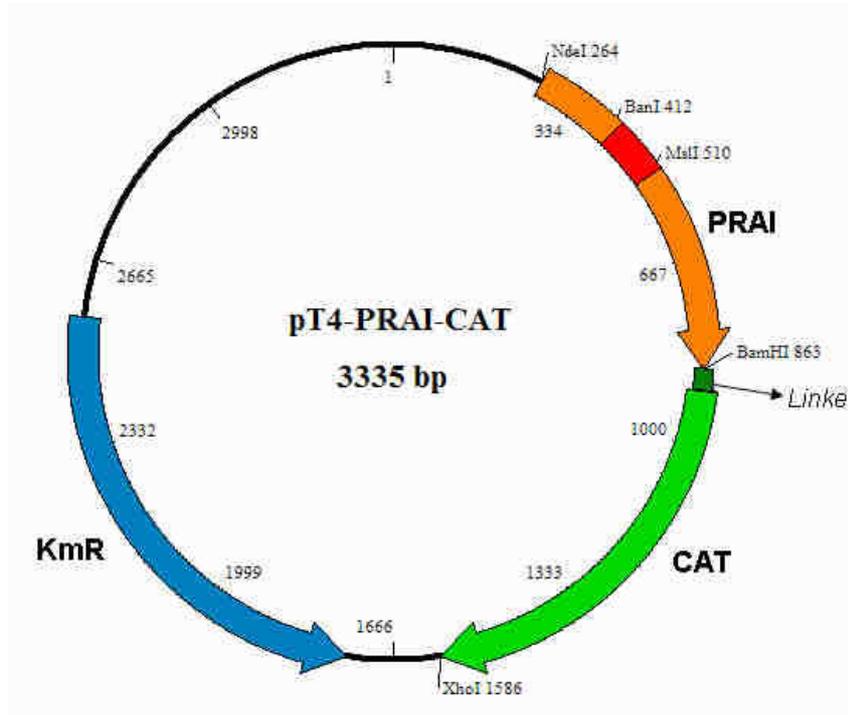
2. CONSTRUCCIÓN DEL BANCO

2.1 DETERMINACIÓN DE LA CMI DE LA PRAI Y PRAI-CAT SILVESTRE

Se consideró como la CMI (Concentración Mínima Inhibitoria) a aquella concentración de cloramfenicol en la que la mitad de las clonas ya no pueden crecer. Se determinó la CMI tanto de la fusión de PRAI silvestre con CAT (Cloramfenicol acetil transferasa) como la de la PRAI silvestre sin fusionar utilizando células recién electroporadas en cajas de LB sólido con concentraciones crecientes de cloramfenicol: 0, 10, 15, 20, 40, 80, 100, 150 y 200 µg/mL. Se inocularon gotas de 5 µL cada una del cultivo sin diluir y

de diluciones de 10^{-1} a 10^{-5} . Se incubaron las cajas a 30 °C por 24 h. Como control positivo se utilizó P+C y como control negativo se utilizó la cepa XL1-blue.

2.2 VECTOR



Purificación de Plásmido

Se inoculó una colonia del plásmido pT4-PRAI-CAT en 3 mL de LB con 3 μ L de kanamicina (1000x, 25 mg/mL) y se incubó a 37 °C durante 16 horas. Posteriormente se colectaron las células por centrifugación y se purificó el plásmido por lisis alcalina, con soluciones y columnas de Roche® (*High Pure Plasmid Isolation Kit*). Obteniéndose el plásmido en una concentración de 126 ng/ μ L.

Digestión del Plásmido con enzimas de restricción

4.4 microgramos del plásmido (35 μ L) se digirieron con 50 unidades de enzima BanI y 31.25 unidades de enzima MstII (New England Biolabs®) en la solución amortiguadora indicada por el proveedor (Buffer 2). El volumen de la reacción total se dividió en 5 tubos cada uno con 17.5 μ L y se incubó a 37 °C por 16 h.

La digestión se corrió en un gel de agarosa grado DNA (Bio-Rad®) al 1% en solución amortiguadora TAE (Sambrook, 1989). Se cortó la banda correspondiente a 3.2 kpb y se

hizo la extracción de DNA del gel de agarosa utilizando soluciones y columnas de Qiagen® (*Qiaquick Gel Extraction Kit*). El vector se recuperó en una concentración de 24 ng/μL.

Reparación del Vector

Con el fin de generar extremos romos en el vector se utilizaron las enzimas Klenow y T4 DNA polimerasa (New England Biolabs®) en presencia de dNTP's (10x, 2.5mM) y la solución amortiguadora indicada por el proveedor. La reacción se incubó de la siguiente manera: 30 min a 12 °C, seguido de otros 30 min a 25 °C y finalmente 10 min a 75 °C.

La reacción se limpió utilizando las soluciones y columnas de Roche® para purificación de productos de PCR (*High Pure PCR Product Purification Kit*) y se recuperaron 30 μL.

Defosfatación del Vector

Se defosfató el vector utilizando el siguiente protocolo. A 30 μL de vector reparado se le agregaron 5 unidades (0.5 μL) de Fosfatasa Alcalina de Carnero (New England Biolabs®) en la solución amortiguadora indicada por el proveedor (Buffer 3), la reacción se llevó a un volumen final de 50 μL. Posteriormente se incubó 30 min a 37 °C y 30 min a 65 °C. Se agregó 6.5 μL de EGTA (250 mM a pH 8) y se le adicionó agua tetradestilada hasta llegar a un volumen final de 80 μL. Finalmente se incubó por 30 min más a 65 °C.

La reacción se limpió utilizando las soluciones y columnas de Roche® para purificación de productos de PCR (*High Pure PCR Product Purification Kit*) y se resuspendió en un volumen de 30 μL. El vector se obtuvo con una concentración final de 24 ng/μL.

2.3 INSERTO

Purificación (Extracción) de DNA genómico

Se centrifugó a 13 000 rpm por 2 minutos un cultivo O/N de 4 mL de *E. coli* cepa XL1-blue, se descartó el sobrenadante y las pastillas se resuspendieron en 400 μL de buffer

amortiguador TE (10:1). Se volvió a centrifugar y resuspender en buffer TE y se agregaron 50 μL de SDS al 10%, 50 μL de Proteinasa K (10 mg/mL) y por último 20 μL de RNAasa (10 mg/mL). Se incubó por 1 h a 37 °C.

El lisado transparente resultante se pasó 3 veces por una jeringa de insulina y se pasó a un tubo limpio en donde se le adicionaron 200 μL de Fenol y 200 μL de Cloroformo, posteriormente la mezcla se agitó con vortex y se centrifugó 2 minutos a 13 000 rpm. Se hicieron visibles 2 fases de la cual se extrajo la fase acuosa y se volvió a repetir el procedimiento con el fenol-cloroformo. Enseguida se puso la muestra en un tubo limpio y se adicionaron 400 μL de cloroformo, se agitó con vortex y centrifugó por 2 minutos a 13 000 rpm. Se tomó nuevamente la fase acuosa y se le agregó 1 mL de etanol frío al 100%. Al mover el tubo apareció se hizo visible el DNA (pelusa blanca). Con ayuda de una pipeta pasteur sellada en forma de gancho se tomó el DNA y se enjuagó en un tubo *epENDORF* con etanol al 70%. Se escurrió bien y se resuspendió en 200 μL de agua tetradestilada.

Nebulización y Obtención de Fragmentos (Inserto)

Se debe calibrar el tanque de gas de nitrógeno ajustando la válvula a 20 psi de presión. Al volumen que se obtuvo de DNA genómico (entre 300 y 500 μL , a una concentración de 300 ng/ μL) se les agregó buffer (TE [50:20], glicerol 25 %) hasta alcanzar un volumen final de 1.5 mL. La mezcla se pasó a un nebulizador nuevo (infantil) y se conectó al tanque de gas sometiendo a la muestra a una presión de 20 psi por 1 minuto y finalmente se transfirió el volumen a un tubo *epENDORF* limpio.

Para la precipitación con acetato de sodio, se agregaron a la muestra 0.3 volúmenes de Acetato de Sodio 2.5M (por cada volumen de muestra) y 1.5 volúmenes de Etanol al 100%. Se dejó enfriar a -20 °C por 16 horas. Posteriormente se centrifugó 30 minutos a velocidad máxima y se hicieron dos lavados de la pastilla con 1 mL de etanol al 70%, se secó la pastilla en una centrífuga al vacío y calor (Savant); y por último se resuspende en 45 μL totales de agua tetradestilada.

Se repararon los fragmentos de la misma manera en que se hizo con el vector y se cargaron en un gel de agarosa de bajo punto de fusión al 1% para cortar y purificar de banda y así obtener los fragmentos del tamaño deseado, que en este caso fueron de 100 a 300 pb. La concentración final de los insertos fue de 120 ng/ μL .

2.4 LIGACIÓN

La ligación se realizó con la enzima T4-DNA ligasa (New England Biolabs®) en la solución amortiguadora indicada por el proveedor incubando a 16 °C durante 16 h. La reacción se llevó a cabo en una relación molar:vector inserto de 1:80, se utilizaron aproximadamente 430 ng de DNA por reacción. Para determinar la cantidad de vector ligado sobre sí mismo se pone un control en las mismas condiciones que el experimento pero sin agregar inserto.

Para concentrar los productos de la reacción y eliminar el alto contenido de sales de la solución amortiguadora se realizó una extracción con butanol (Sambrook, 1989), posteriormente se lavó con etanol al 100% y finalmente se resuspendió en 10 µL de agua tetradestilada.

2.5 TRANSFORMACIÓN DE CÉLULAS POR ELECTROPORACIÓN

Se mezclaron 60 µL de células competentes *E. coli* MC1061 (con una eficiencia mayor a 1×10^8 transformantes por µg de DNA) con 1 µL de ligación (alrededor de 43 ng). Se pasaron a una celda de 0.1 cm y se sometieron a un choque eléctrico de 1.8 V en un electroporador (Bio-Rad®) ajustado a 25µF y 200 Ω. Se recuperaron las células agitándose durante 1h a 30 °C en 1mL de medio SOC (Sambrook, 1989). El mismo procedimiento se siguió utilizando la misma cantidad de DNA de la reacción control de autoligación.

2.6 SELECCIÓN

Se plaquearon 100 µL de las células recién electroporadas en cada caja de LB sólido con kanamicina (35 µg/mL) y cloramfenicol (20 µg/mL) y se incubaron a 30 °C hasta encontrar colonias aisladas. Con el fin de cuantificar el banco se plaquearon 100 µL de las células electroporadas en una caja con kanamicina. A partir de éstas colonias se realizaron cultivos en LB líquido con antibiótico (Km y Cm₂₀) del que se purificó plásmido.

Finalmente se verificó la presencia de un inserto por digestión con las enzimas de restricción NdeI y BamHI y/o por una reacción de PCR con los oligonucleótidos NdePRAI y BamPRAIx.

NdePRAI	5' ATA TAT ACA TAT GCT GGA GAA TAA GGT ATG 3'
BamPRAIx	5' GCG GAT CCG TAT GCG CGC AGC GTC 3'

Los oligonucleótidos fueron sintetizados en la Unidad de Síntesis del Instituto de Biotecnología, UNAM

Las reacciones de PCR se llevaron a cabo en el equipo “iCycler” de Bio-Rad bajo el siguiente programa.

No. Ciclos	Temperatura	Tiempo
1	94° C	5'
25	94° C	40''
	55° C	40''
	72° C	1'
1	72° C	10'

2.7 SECUENCIACIÓN

Por último se mandaron a secuenciar las clonas que tenían inserto. La secuenciación de DNA de los plásmidos purificados se realizó con el apoyo de la Unidad de Secuenciación del Instituto de Biotecnología de la UNAM.

3. CARACTERIZACIÓN DE CLONAS

3.1 Caracterización Bioinformática

Una vez obtenidas las secuencias de las clonas de interés se verificó que las secuencias estuvieran en fase y se les hizo un BLAST, para posteriormente hacer una predicción de la estructura secundaria con el programa PSIPRED V2.4. Finalmente, se buscaron en el

PDB las secuencias de los fragmentos insertados para ver si tenían homólogos de estructura conocida (valor de $E < 0.01$)

3.2 CMI de las Clonas

Los plásmidos de las 18 clonas Cm^R fueron transformados en la cepa XLI Blue calcio competentes. Se determinó la CMI de 18 clonas con inserto Cm^R en cajas de LB sólido con concentraciones crecientes de cloramfenicol: 0, 20, 40, 60, 80, 100, 120 y 150 $\mu\text{g}/\text{mL}$. Se inocularon gotas de 5 μL cada una del cultivo sin diluir y de diluciones de 10^{-2} , 10^{-4} y 10^{-6} de un cultivo O/N crecido en LB con kanamicina. Se incubaron las cajas a 30 °C por 24 h. Como control positivo se utilizó P+C y como control negativo se utilizó la cepa XL1-blue.

4. EXTRACCIÓN DE PROTEÍNAS (TOTALES) Y WESTERN BLOT

Se llevó a cabo un Western blot con el fin de corroborar que la clona C5 se expresaba y para verificar que estuviera fusionada a CAT. Para la extracción e inducción de la proteína se inocularon 10 mL de medio LB con kanamicina (1000x) con 1 mL de un cultivo O/N (16 horas). Se creció a 30 °C hasta alcanzar una densidad óptica de 0.4 OD y se indujo agregando 1mM de IPTG. Se retiró el cultivo de la incubadora cuando éste llegó a una densidad óptica de 0.9 OD, posteriormente se centrifugó en frío a 4000 rpm por 5 min y se resuspendió el pellet en 1 mL de TBS 1x. La muestra se sonicó y se le agregó PMSF (100 μM). A continuación se centrifugó en frío a 4000 rpm durante 5 min para separar las proteínas solubles (sobrenadante) de las no solubles (pellet), éstas últimas se resuspendieron en Tris pH 8 con 1% de SDS. Por último se cargan y corren las muestras en un gel de acrilamida al 12.5%.

Para llevar a cabo el western blot primero se transfirieron las proteínas separadas por SDS-PAGE a una membrana de nitrocelulosa que posteriormente fue bloqueada con 5% de una solución de leche en polvo y TBST O/N a 4 °C. A continuación se lavó 3 veces con TBST por 10 min y se le agregó el anticuerpo ANTI-CAT en una relación 1:3000 en solución TBST con 0.1% de leche, se incubó durante 1 h y se lavó 3 veces con TBST. Se puso el segundo anticuerpo ANTI-DIG en una relación de 1:5000 incubando 1 h en TBST 0.1% de leche. Se lavó 2 veces con TBST. Finalmente se adicionaron a la

membrana 1mL de BCIP (Zymed®) y 1 mL de NTB (Zymed®) aforando con agua (mQ) a 10 mL.

5. PROTOCOLO PARA DESFUSIONAR CAT

Para remover la CAT de la PRAI se digirió el vector pT4-PRAI-CAT con las enzimas de restricción BamHI y BglII (New England Biolabs®) que cortan al principio y al final de la CAT respectivamente. Se incubó a 37 °C por 16 h.

La digestión se corrió en un gel de agarosa grado DNA (Bio-Rad®) al 1% en solución amortiguadora TAE (Sambrook, 1989). Se cortó la banda correspondiente a 2 958 pb y se hizo la extracción de DNA del gel de agarosa utilizando soluciones y columnas de Qiagen® (*Qiaquick Gel Extraction Kit*). Posteriormente el vector fue religado utilizando la enzima T4-DNA ligasa (New England Biolabs®). Para eliminar el alto contenido de sales de la solución amortiguadora se realizó una extracción con butanol (Sambrook, 1989), posteriormente se lavó con etanol al 100% y finalmente se resuspendió en 15 µL de agua tetradestilada. El vector se electroporó en células electrocompetentes JM101 y se plaquearon en cajas con kanamicina (35 µg/mL).

Finalmente se mandaron a secuenciar algunas clonas para corroborar que la proteína estuviera en la fase correcta y sin la fusión a CAT.

6. ACTIVIDAD IN VIVO DE PRAI

Se electroporó 1 µL de plásmido superenrollado de la clona C5 con 20 µL de células competentes JM101 $trpF^-$ y 30 µL de glicerol al 10%. Se incubaron las células recién electroporadas a 37 °C en medio SOC hasta que se llegó a una densidad óptica de 1 OD. Se centrifugó a 4000 rpm por 5 min y se resuspendió el pellet en 1 mL de PBS 1x cuidando de quitar por completo cualquier residuo de medio SOC. Se ajusta la densidad óptica a la misma OD que los controles (PRAI-CAT silvestre y cepa sola) y se hicieron 7 diluciones 1:10, se tomó 5 µL de cada dilución y se puso la gota en las cajas con medio mínimo Vogel-Bonner (VB) tanto con casaminoácidos como sin ellos. Por último se incubaron las cajas a 30 °C hasta que se vio crecimiento considerable (48-72 horas).

Preparación de Medio Mínimo VB (250 mL)

Reactivos	Volumen/Peso
Antibiótico (Km)	250 μ L
Glucosa 20%	5 mL
Casaminoácidos 20%	87 μ L
Sales VB 20x	12.5 mL
FeCl ₃ 5 mg/mL	250 μ L
Tiamina 10 mg/mL	62.5 μ L
Agar Noble	6 g
Agua mQ	200 mL (aprox)

Preparación Sales VB 20x (100 mL)

Reactivos	Peso
MgSO ₄ · 7H ₂ O	0.4 g
Ácido cítrico anhidro	3.94 g
K ₂ HPO ₄	20 g
Na ₂ HPO ₄ · 2H ₂ O	3 g

*Ajustar el pH a 7 con KOH

RESULTADOS Y DISCUSIÓN

1. Elección de un fragmento de la PRAI a sustituir

Con el fin de elegir el segmento de la PRAI a sustituir se utilizó el programa SCHEMA (www.che.caltech.edu/groups/fha/), un algoritmo computacional que identifica los fragmentos de una proteína que pueden ser recombinados sin alterar la integridad de su estructura tridimensional (Voigt, 2002).

SCHEMA calcula el número de interacciones entre aminoácidos y determina el número de interacciones que serían perturbadas al haber un evento de recombinación, lo que genera un perfil de schema “schema profile” (S); si S es grande para un residuo significa que ese residuo está involucrado en una estructura más compacta; y por el contrario si S es pequeño el residuo está involucrado en un menor número de interacciones. Se considera que dos aminoácidos interactúan cuando sus átomos (excluyendo los átomos de hidrógeno) se encuentran a una distancia igual o menor a 4.5 Å, lo que corresponde de entre 5-8 interacciones por residuo (Voigt, 2002).

Al generar el perfil para la PRAI (*Fig. 11*) encontramos 8 picos bien definidos que corresponden a las 8 subunidades que caracterizan al plegamiento $(\beta\alpha)_8$.

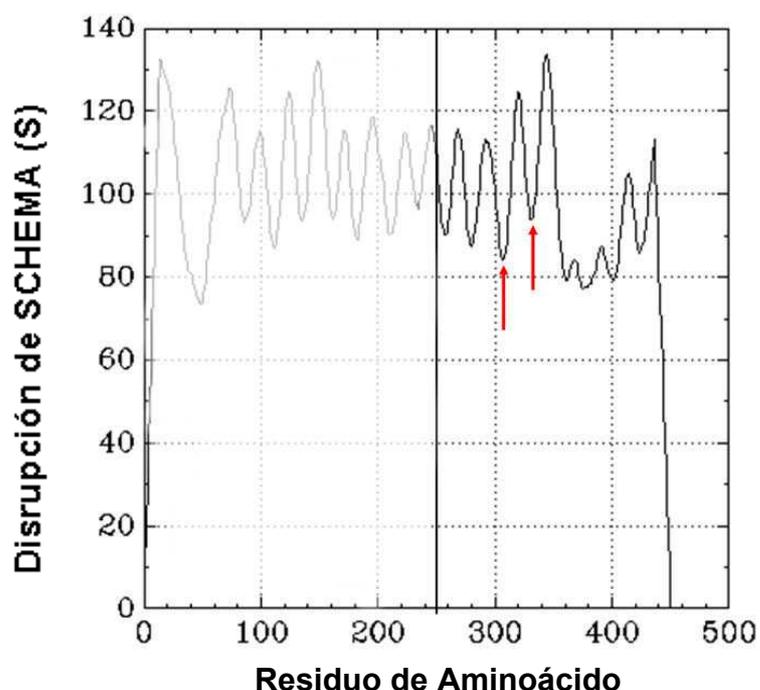


Fig. 11. Perfil de SCHEMA para IGPS-PRAI (pdb: 1pii). La PRAI comprende del residuo 256 al 452. Se muestran con flechas rojas el que será el segmento a sustituir.

Para elegir cual de las 8 subunidades (encontradas por SCHEMA) se iba a sustituir, se hizo un alineamiento estructural de 12 proteínas homólogas a la PRAI según la base de datos CATH (1glyA, 1d3hA, 1dqwA, 1ep3A, 1geqB, 1gpwA, 1bwkA, 1i4n, 1k3uA, 1rpxA, 1tpfA, 1nsj) utilizando el programa MultiProt; posteriormente, se comparó una de ellas, la PRAI de *T. maritima* (1nsj) con la PRAI de *E. coli*. Se encontró un fragmento conservado estructuralmente en todas las proteínas alineadas que correspondía a una unidad $\beta\alpha\beta$ y que también se relacionaba con los mínimos encontrados con el programa SCHEMA. Se eligió un segmento estructuralmente conservado con la finalidad de que el fragmento a sustituir fuera un elemento común en los barriles $(\beta\alpha)_8$.

En el caso de la PRAI de *E. coli* la región seleccionada comprendía del residuo 51 al residuo 83 (96 pares de bases) lo cual codifica para las estructuras $(\beta\alpha)_3\beta_4$ (Fig. 12). En este fragmento de la proteína se encuentran 3 residuos pertenecientes al sitio activo, pero ninguno de ellos es catalítico ni de unión a ligando. Asimismo el fragmento a sustituir se encuentra en la porción que se ha sugerido como un sitio de nucleación importante para los eventos tempranos en el plegamiento, por lo que en teoría sin este segmento la PRAI no podría plegarse sola (Akanuma y Yamagishi, 2005); y se necesitaría forzosamente de una secuencia que pudiera complementar el plegamiento.

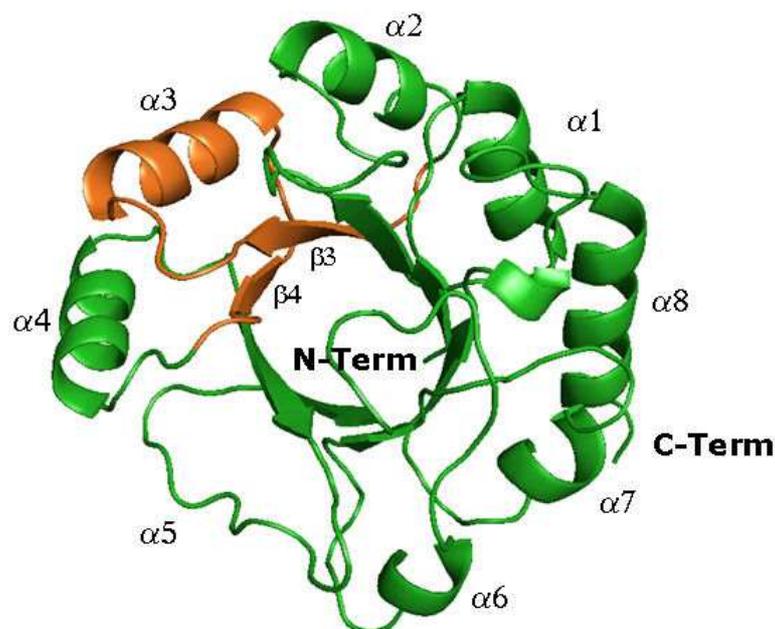


Fig. 12. Representación tridimensional de la PRAI en donde se muestra el fragmento $(\beta\alpha)_3\beta_4$ a ser sustituido (en naranja).

2. Obtención del banco de DNA genómico

Como primer paso para la construcción del banco se utilizaron las enzimas BanI y MslI para digerir el vector ya que se encontraban cercanos a los residuos encontrados por los programas SCHEMA y MultiProt para liberar el fragmento a sustituir en la PRAI (Fig. 13). Posteriormente se generaron extremos romos en el vector con el fin de tener extremos compatibles con los insertos y se defosfató para disminuir la probabilidad de eventos de autoligación. Si llegara a religarse algún plásmido, la segunda mitad de la PRAI así como la CAT quedarían fuera de fase por lo que no se podría traducir el producto.

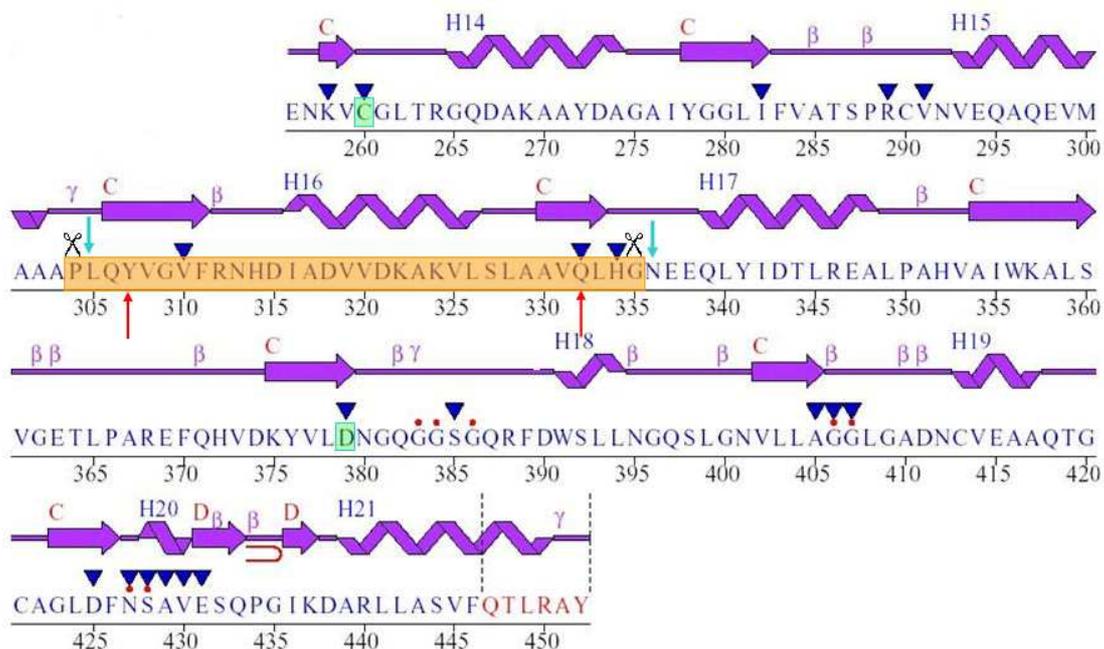


Fig. 13. Secuencia y estructura secundaria de PRAI en donde se muestra: aminoácidos pertenecientes al sitio activo (▼); sitios de corte con las enzimas de restricción BanI y MslI (✂); Mínimos encontrados por SCHEMA (↑); Alineamiento estructural (↓) y el fragmento a ser sustituido (■).

En el caso de los insertos, se fragmentó el genoma de *E. coli* totalmente al azar por nebulización con nitrógeno gaseoso (método físico) y al igual que con el vector, se repararon para generar extremos romos. De esta manera la ligación de los fragmentos de entre 100 y 300 pb con el vector sería un proceso que dependería únicamente del azar (Fig. 14 y 15).

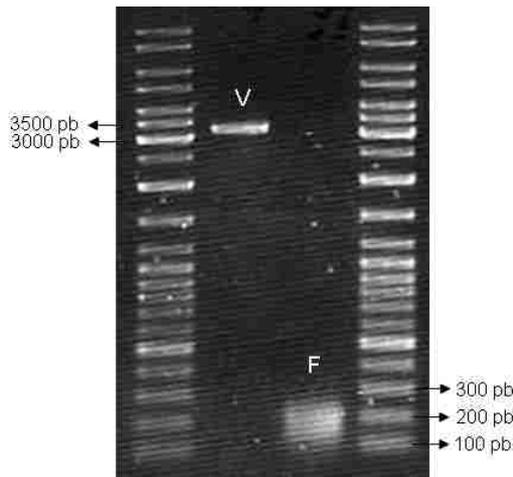


Fig. 14. Vector (V) de 3.2 kb y fragmentos (F) de 100 a 300 pb que serán ligados para la construcción del banco. Marcador GeneRuler™ DNA Ladder Mix (Fermentas®)

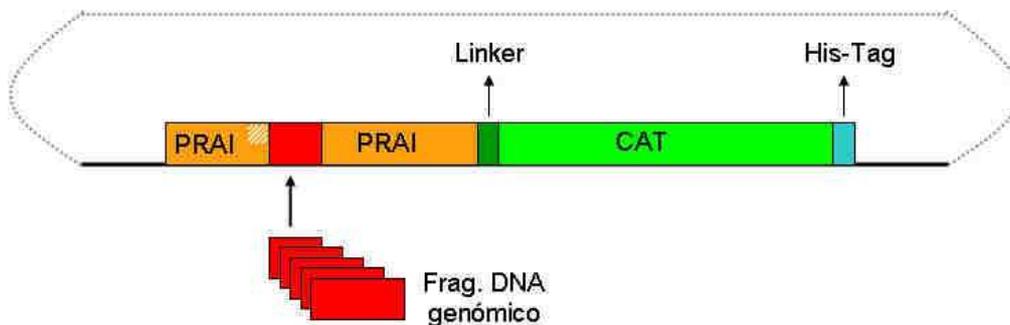


Fig. 15. Construcción de la fusión de PRAI con CAT en donde se clonaran los fragmentos de DNA Genómico.

El método que se utilizó para seleccionar las proteínas quiméricas que estuvieran plegadas fue la resistencia a cloramfenicol. Para este sistema (de selección *in vivo*) la cloramfenicol acetil transferasa (CAT), una enzima homotrimérica altamente soluble, sirve como reportero de plegamiento al estar fusionada a la proteína de interés, en este caso a PRAI (Fig. 16), y conferir así a la célula la capacidad de hidrolizar el antibiótico en cuestión y crecer en presencia de cloramfenicol (Maxwell, 1999).

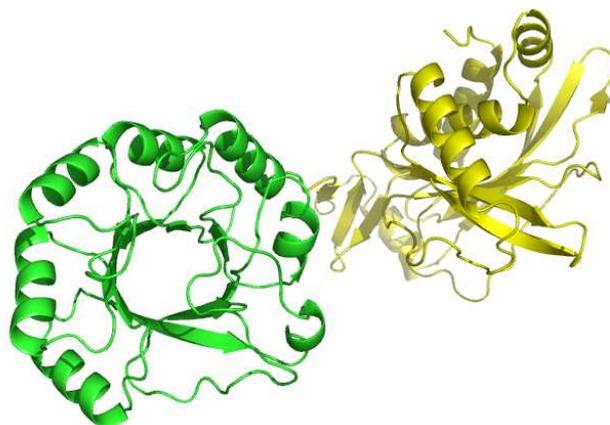


Fig. 16. Modelo de la fusión de la proteína de interés PRAI (verde) con la CAT (amarillo)

Se determinó que los bancos serían plaqueados en cajas de LB sólido con kanamicina (resistencia del vector) y con una concentración de 20 µg/mL de cloramfenicol, ya que al hacer los CMI's correspondientes utilizando células recién electroporadas la fusión de PRAI silvestre con CAT (P+C) resistió una concentración de cloramfenicol de 40 µg/mL mientras que la PRAI silvestre sin fusionar a CAT únicamente resistió 5 µg/mL; por lo que cualquier clona que lograra crecer por arriba de este valor tendría el fenotipo deseado.

En un primer banco se encontraron tan solo 2 clonas cloramfenicol resistentes (Cm^R) con inserto y en fase, las clonas C5 y C7, el resto del banco correspondía a colonias religadas y a clonas P+C que no habían sido bien digeridas.

Posteriormente se optimizaron los protocolos para la construcción del banco (utilizando una relación molar vector:inserto de 1:80), en donde la ligación se electroporó en células MC1061 con eficiencias mayores a 1×10^8 transformantes por microgramo de DNA y generándose una librería de 1.95×10^3 clonas de las cuales un 3% fue Cm^R . Con respecto al control de religadas ninguna clona se encontró creciendo en cajas con cloramfenicol.

Sin embargo, este porcentaje disminuyó debido a las falsas positivas que genera el sistema de selección utilizando CAT, aunado a que el cloramfenicol se degrada en las cajas de plaqueado conforme pasa el tiempo. Por lo que clonas que logren crecer posteriores a las 48-72 horas no son confiables en cuanto al fenotipo que presentan ya que al ser secuenciadas la mayoría de estas aunque poseen inserto tienen codones de paro distribuidos a lo largo de la secuencia.

3. Análisis de las clonas obtenidas

Del banco obtenido se tomaron al azar 33 clonas con el fenotipo deseado a las que se les digirió con las enzimas de restricción NdeI y BamHI que cortan al principio y al final de la PRAI para corroborar que tuvieran inserto y ser posteriormente secuenciadas (*Fig. 17*). De las 33 clonas, 18 contaban con inserto distinto a PRAI, 2 de ellas era la PRAI sin digerir y las 13 restantes fueron falsas positivas debido al sistema de selección con CAT y clonas con codones de paro que crecieron después de 48.

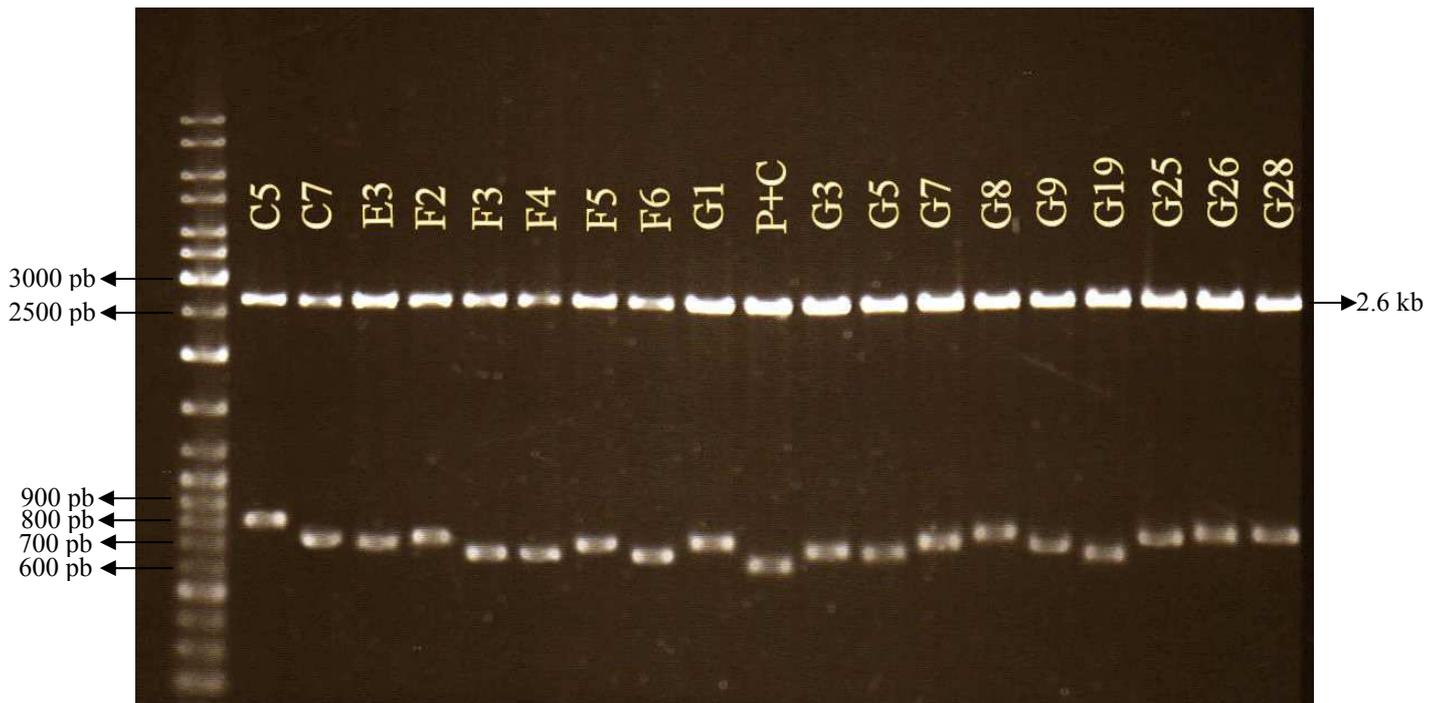
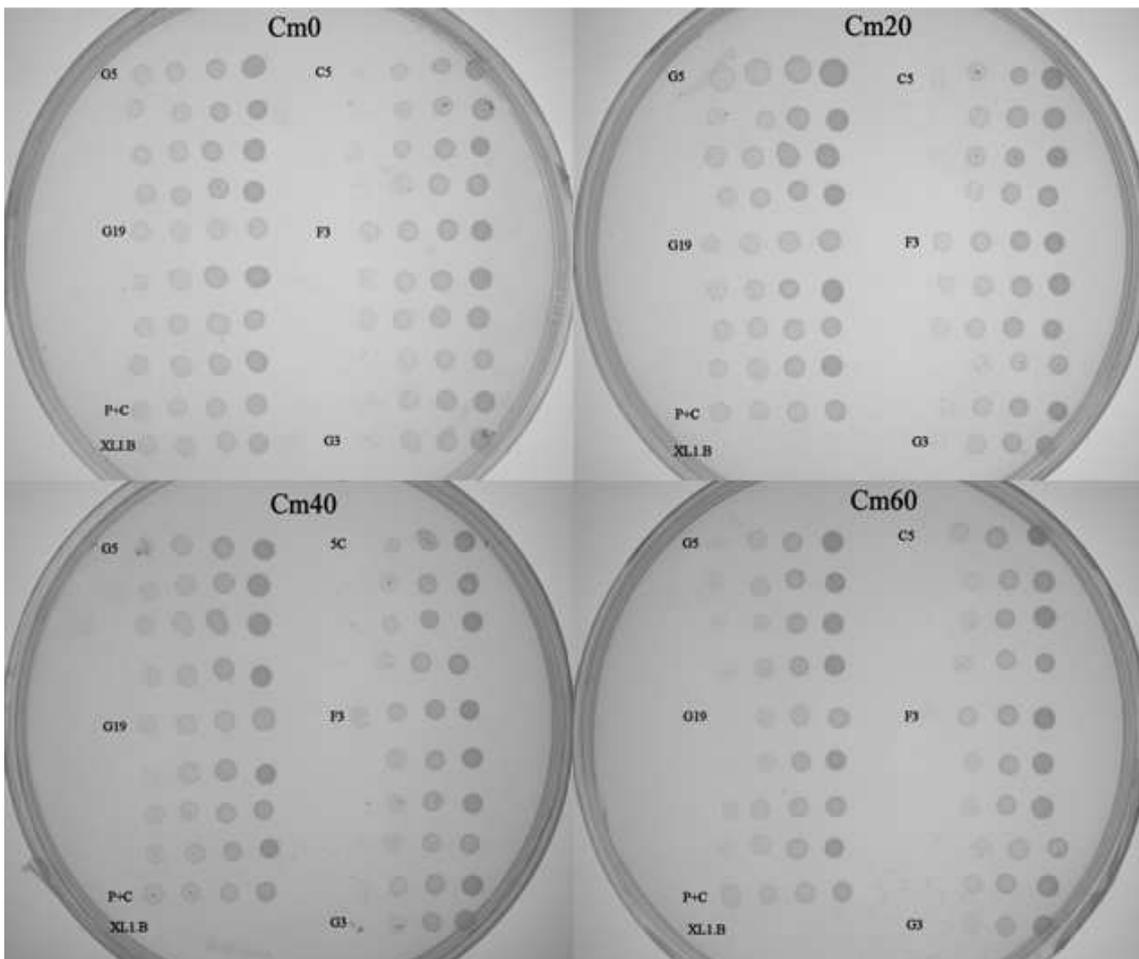
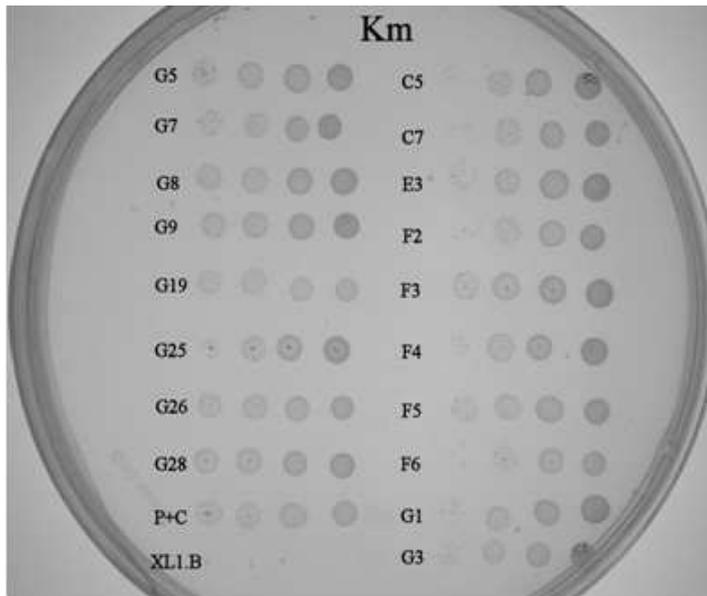


Fig. 17. Gel de agarosa (1%) en donde se muestran las digestiones de los 18 plásmidos con clonas Cm^R. Las digestiones se llevaron a cabo con las enzimas de restricción NdeI y BamHI, que reconocen una secuencia al principio y al final de la PRAI respectivamente; por lo que el segmento liberado corresponde al tamaño el gen quimérico en cada clona. La banda de 2.6 kb que se observa corresponde al tamaño del plásmido sin PRAI. Marcador GeneRuler™ DNA Ladder Mix (Fermentas®).

Para las 18 clonas se encontraron segmentos distintos los cuales fueron identificados dentro del genoma de *E. coli*. Para todas estas secuencias había un ORF (marco de lectura abierto) desde el principio de la PRAI hasta el final de la CAT.

Se observó un porcentaje alto de secuencias que codificaban para transportadores de membrana (21%) y reguladores (21%); siendo que estas dos clases de proteínas no son particularmente abundantes dentro del proteoma de *E. coli*, 6% y 9% respectivamente (Riley, 2006).

A las 18 clonas con inserto distinto a PRAI se les midió la CMI en concentraciones crecientes de cloramfenicol a partir de cultivos O/N para darnos una idea de que tan bien estaban plegadas las proteínas recombinantes encontradas (Fig. 18).



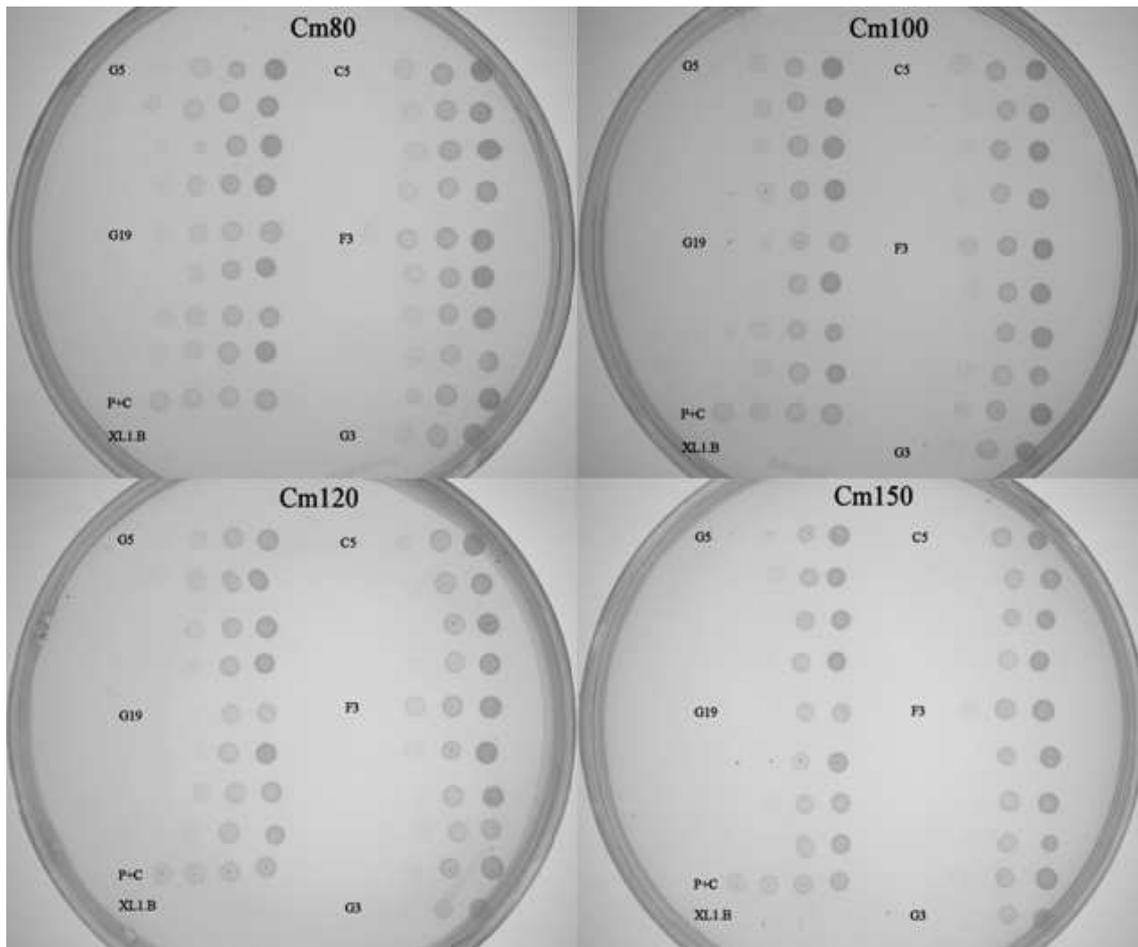


Fig. 18. Determinación de CMI de Cloramfenicol de las 18 clonas con inserto. Como control positivo se utilizó la fusión de PRAI+CAT (P+C) y como control negativo se utilizó la cepa XLI.Blue sin plásmido. En la parte superior de cada caja se indica la concentración del antibiótico en $\mu\text{g/mL}$.

Como se muestra en la TABLA 1 algunas clonas (la F3, F6 y G7 por ejemplo), resisten una alta concentración de cloramfenicol; estas proteínas cuentan con una mayor estabilidad dentro de la célula al encontrarse mejor plegadas (y más solubles) que aquellas cuyo valor de CMI se encuentra alrededor de los $40 \mu\text{g/mL}$, como es el caso de las clonas 7C, E3, F2, F5 y G3. Sin embargo, es importante recalcar que la fusión de la PRAI silvestre con la CAT tiene una alta estabilidad dentro del citoplasma al estar plegada correctamente, hecho que pudimos constatar ya que su CMI está por arriba de los $150 \mu\text{g/mL}$ Cm, valor por encima de cualquiera de la proteínas recombinantes encontradas hasta el momento.

Del análisis de las 18 clonas seleccionadas resalta que aproximadamente la mitad de los fragmentos que se insertaron para generar las proteínas recombinantes se encontraban

en el mismo marco de lectura que el del gen original del cual partieron; mientras que en la otra mitad de los segmentos hubo un cambio en el marco de lectura, por lo que la secuencia de aminoácidos es diferente al de la proteína propuesta para *E. coli*. Es importante hacer notar que de las clonas que resistieron una mayor concentración de cloramfenicol, todas a excepción de una, derivaron de segmentos expresados en el marco de lectura de la proteína original.

Tabla 1. Características de las 18 clonas encontradas.

<i>CLONA (tamaño del inserto)</i>	<i>ORIGEN GEN</i>	<i>PROTEÍNA DE ORIGEN</i>	<i>EN FASE CON EL GEN DE ORIGEN</i>	<i>PRED ESTRUC SEC</i>	<i>CMI Cm (µg/mL)</i>
C5 (107 aa)	fepA	Receptor de la membrana externa para el transporte del complejo hierro-sideróforo	✓	βββαβ	60–80
C7 (73 aa)	fiu	Posible receptor de la membrana externa para el transporte de hierro	✗	βαααβ	40
E3 (68 aa)	yagR	Posible oxidoreductasa con un dominio de unión a molibdeno	✓	αβαβ	40
F2 (71 aa)	yegE	Posible diguanilato ciclasa (“GGDEF domain signalling protein”)	✓		40
	+			ββαβ	
	ygbJ	Posible deshidrogenasa con un dominio Rossmann de unión a NAD(P)	✗		
F3 (52 aa)	ygcP	Posible proteína reguladora	✗	βαβαβ	100
F4 (50)	ydcS	Posible transportador de espermidina/putrescina, miembro de la superfamilia ABC de transportadores	✗	βαβαβ	40–60
F5 (65 aa)	rnt	Ribonucleasa T	✗	βααβα	40
F6 (49 aa)	ydiJ	Posible oxidoreductasa ligada a FAD	✓	βαβα	100–120
G1 (68 aa)	cbl	Activador transcripcional de unión a DNA de la biosíntesis de cisteína.	✓	ββαβ	80–100
G3 (50 aa)	fecA	Transportador de citrato férrico de la membrana externa	✓	βααα	40
G5 (44 aa)	ykgB	Posible proteína de la membrana interna	✗	βααα	80
G7 (59 aa)	frvR	Posible regulador del catabolismo de polisacáridos	✓	βαβαβ	100
G8 (70 aa)	ydjZ	Posible proteína de la membrana interna	✗	αβαβ	60
G9 (51 aa)		Posible S-transferasa	✗	βαβα	60

G19 (34 aa)	acnB	Aconitato hidratasa	✘	$\alpha\beta\alpha$	60
G25 (60 aa)	ypdE	Posible aminopeptidasa	✘	$\beta\alpha\beta\alpha$	60
G26 (68 aa)	ybgP	Posible chaperona (“assembly protein”)	✓	$\beta\beta\beta\alpha\beta\alpha\beta$	80–100
G28 (75 aa)	zwf	Glucosa-6-fosfato 1-deshidrogenasa	✘	$\beta\alpha\beta\alpha$	60–80
PRAI +CAT	-	-	—	-	> 150
XLI.B	-	-	—	-	< 20

Al enfocarnos en las secuencias de aminoácidos de los fragmentos insertados en cada una de las proteínas quiméricas encontradas, se observó que el porcentaje de aminoácidos hidrofóbicos en todas ellas era menor que en la región de PRAI removida (37%) y que el porcentaje de aminoácidos cargados (carga total) era diferente al de la proteína silvestre (Anexo 1). Esto nos indica que ni la cantidad de aminoácidos hidrofóbicos ni de aminoácidos cargados fue una restricción o condición para que los fragmentos insertados hayan complementado el plegamiento.

Por otro lado, al llevar a cabo la predicción de estructura secundaria (PSIPRED V2.4) de las 18 proteínas nuevas encontramos en promedio un 53.3% de hebras β y una 46.7% de hélices α , sugiriéndonos que las proteínas quiméricas se pliegan utilizando una combinación de estos dos tipos de estructuras secundarias, de la misma manera en que lo hace la PRAI, cuyo plegamiento es una mezcla de β plegadas y α hélices. Para algunas proteínas el segmento insertado se pliega utilizando casi exclusivamente hebras beta (ej. 5C), en otros predominan las alfa hélices (ej. G5) pero en la mayoría la combinación de estas dos estructuras es lo que prevalece (ej. E3, F6, G7). No obstante, debido a que los fragmentos insertados eran relativamente grandes (>50 residuos de aminoácidos) era difícil y poco preciso distinguir la presencia de estructuras supersecundarias bien definidas o de algún atractor en particular.

4. Caracterización de la Proteína Quimérica C5

Del banco obtenido se eligió una de las clonas, la C5, para ser caracterizada más a fondo. En este caso el fragmento de DNA genómico de *E. coli* que se insertó dentro del

gen de la PRAI tiene un tamaño de 312 pares de bases que corresponden a 104 aminoácidos, y en donde el segmento clonado se expresa en el mismo marco de lectura que el de la proteína original.

El gen de esta proteína (fepA) codifica para una porina de la membrana externa que actúa como un receptor para el transporte de la enterobactina férrica (un complejo hierro - sideróforo) al periplasma de la célula. En la proteína original, FepA, el fragmento insertado dentro del vector va del residuo 199 al 303 y corresponde a tres hojas beta y dos asas (*Fig. 20*), éstos últimos posiblemente involucrados en la unión al ligando. (Buchanan et al., 1999)

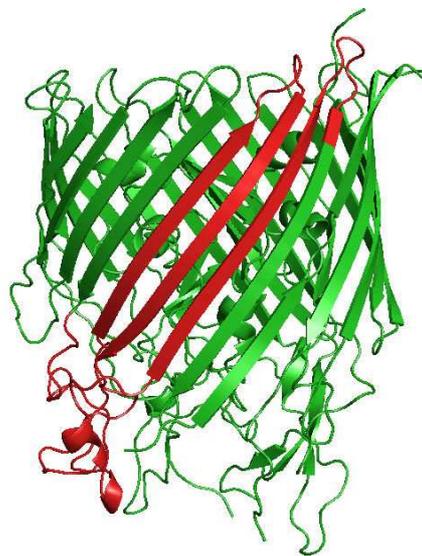


Fig. 20. Estructura tridimensional de la proteína FepA. En rojo se muestran las estructuras correspondientes al fragmento insertado dentro de PRAI.

Al llevar a cabo la predicción de la estructura secundaria (PSIPRED) para esta clona se encontró que el fragmento insertado correspondía a una serie de $\beta\beta\beta\alpha\beta$ (*Fig. 19*).

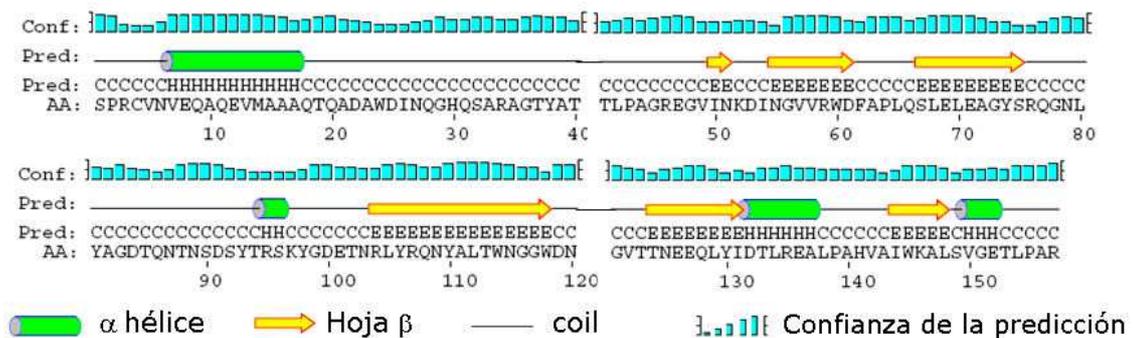


Fig. 19. Predicción de la estructura secundaria de la proteína quimérica C5

EXPRESIÓN Y WESTERN BLOT DE LA FUSIÓN Y LA NO FUSIÓN

Con el fin de confirmar si la proteína recombinante estaba unida al reportero de plegamiento CAT, se llevó a cabo un Western (anti-CAT) en donde se observó que la fusión de la C5 con la CAT existía, coincidiendo con el tamaño (peso) calculado teóricamente, 56 kD (*Fig. 21*).

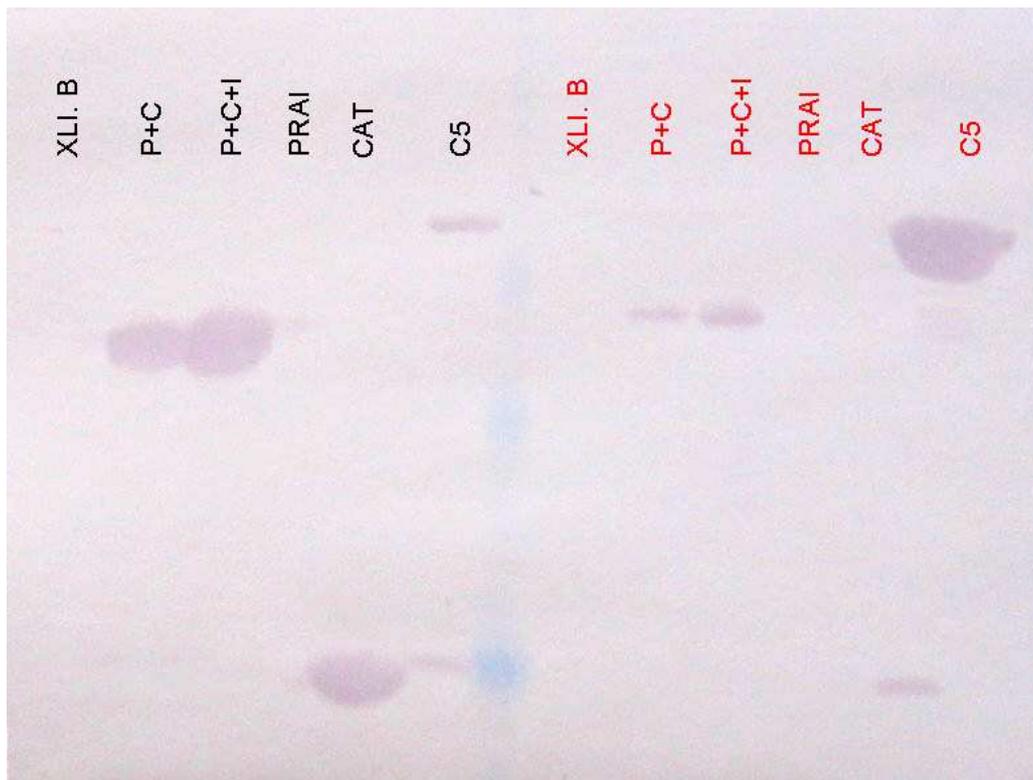


Fig. 21. Western Blot anti-CAT en donde se muestra la proteína C5 fusionada con CAT (57 kD). La parte de la izquierda corresponde a la fracción soluble de la célula (negro) y la parte de la derecha (rojo) corresponde a la fracción insoluble. Como controles positivos se utilizaron: Fusión P+C (48 kD), P+C Inducida, CAT sola (26 kD). Como controles negativos se utilizaron la cepa XLI.B sin plásmido y la PRAI sola (22 kD).

Se encontró, que aunque la mayoría de la proteína se concentraba en la fase insoluble, debido probablemente a la naturaleza del vector utilizado (influencia de un fuerte promotor *lac*), la fusión de la C5 con CAT estaba presente en la fracción soluble de la célula; dato importante ya que nos indica que efectivamente la proteína está plegada y estable dentro del citoplasma.

Así mismo, se removió la CAT de la C5 a nivel de DNA, y en un gel de acrilamida desnaturalizante se observó una banda de 30kD (275 aminoácidos), tanto en la fracción

horas de plaqueo (Fig. 23). En medio rico LB, la C5 y todos los controles crecieron en un periodo de 12 a 24 horas debido a que no hay limitación por falta de triptofano.

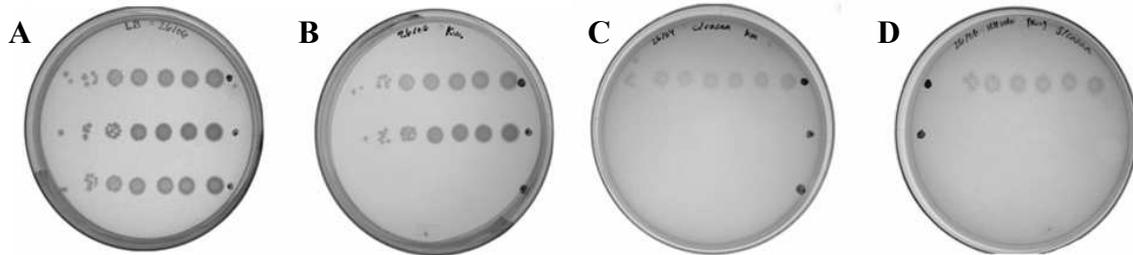


Fig. 23. Ensayo de conservación de la actividad de PRAI en la clona C5. De arriba hacia abajo: P+C, C5, cepa JM101 Δ trpF. Caja A: Medio LB, Caja B: Medio LB + Km, Caja C: Medio Mínimo con casaaminoácidos + Km, Caja D: Medio Mínimo sin casaaminoácidos + Km.

En este caso la inserción proveniente de la porina que sustituyó el fragmento $\beta\beta$ de la proteína silvestre, en donde se encontraban involucrados 3 residuos del sitio activo, no logró complementar para dar la función de ésta; por lo que la nueva proteína química C5 perdió la actividad de isomerización que tenía PRAI.

CONCLUSIONES

Hasta ahora, la mayoría de los trabajos de ingeniería de proteínas han estado dirigidos a la exploración local del espacio de secuencia de una proteína. En este proyecto se desarrolló una estrategia para generar nuevas proteínas plegadas por recombinación independientemente de la existencia de homología a nivel estructural o funcional. Se logró generar una librería de proteínas quiméricas por la fusión de un segmento que codifica para la PRAI de *E. coli*, una enzima representante de la amplia familia de los barriles ($\beta\alpha$)₈, con fragmentos del genoma de este mismo organismo de entre 100 y 300 pares de bases, con ello se logró una exploración global del espacio de secuencia, mucho más extensa que la que se logra con otras metodologías como las mutaciones puntuales y el “family shuffling”. El que se haya implementado con éxito este sistema de generación de variabilidad apoya con datos experimentales la teoría de que la recombinación no homóloga es una herramienta eficiente de búsqueda del espacio de adecuación global así como de innovación estructural, necesaria para la diversificación del proteoma.

En el caso particular de la PRAI, encontramos que existe una gran diversidad de segmentos polipéptidos que son capaces de sustituir al fragmento $\beta\alpha\beta$ removido del barril y generar nuevas proteínas quiméricas plegadas. Estas secuencias de residuos de aminoácidos provienen tanto de marcos de lectura de proteínas existentes como de secuencias que se encuentra en un ORF distinto; esto nos habla de sugiere que el genoma de *E. coli*, así como del resto de los genomas, tienen un gran potencial evolutivo para generar diversidad a nivel proteico; sin embargo, es importante hacer notar que en la naturaleza no se ven proteínas plegadas nuevas fácilmente. De 10^3 recombinaciones un 2% de ellas fueron exitosas dando lugar a un nuevo dominio plegado. Con el barril TIM elegido fue posible generar muchos más eventos de recombinación exitosos que los obtenidos en un trabajo publicado de la misma naturaleza, en parte debido a que el andamiaje utilizado tiene una alta estabilidad capaz de tolerar profundos cambios en su estructura. Los resultados obtenidos en este trabajo podrían ser extrapolados a otros miembros de la familia (β/α)₈ ya que el fragmento sustituido se encuentra conservado estructuralmente en ellos.

Sugerimos tres posibilidades en cuanto al plegamiento que pueden adoptar las nuevas proteínas generadas: 1) que el fragmento insertado complemente el plegamiento con una estructura $\beta\alpha\beta$ generándose de nuevo un barril $(\beta/\alpha)_8$; 2) que se obtenga un plegamiento distinto al andamiaje original pero ya conocido y 3) que la fusión de la PRAI con el segmento proveniente del genoma de *E. coli* haya generado un plegamiento totalmente nuevo. Estas tres posibilidades tendrán que ser evaluadas para entender de mejor manera el rol de cada uno de los segmentos del nuevo dominio generado, por lo que en un futuro se caracterizarán termodinámicamente las distintas proteínas recombinantes para determinar su estabilidad, presencia de estructuras secundarias y terciarias así como el mecanismo de plegamiento.

Además de las implicaciones que tiene este trabajo en el terreno de la evolución de proteínas al resaltar la importancia de la recombinación no homóloga como un proceso importante que ha moldeado la diversidad de proteínas con las que contamos hoy en día, los resultados obtenidos servirán como base para la generación *in vitro* de proteínas *de novo* con aplicaciones útiles en distintos sectores como el biotecnológico, el industrial y el médico.

ANEXO 1

Características de los segmentos insertados en las proteínas recombinantes

CLONA (Fragmento Insertado)	AA HIDRO- FÓBICOS (%)	CARGA TOTAL (%)	CARGA NETA (%)
C5	14	19.6	-2.8
C7	30.1	15.1	+1.4
E3	29.4	17.6	+11.8
F2	36.6	16.9	0
F3	23.1	7.7	+3.8
F4	26	14	+10
F5	35.4	18.5	+6.2
F6	36.7	26.5	-18.4
G1	33.8	27.9	-1.5
G3	22	12	0
G5	24.4	13.3	+8.9
G7	33.9	15.3	-1.7
G8	24.3	11.4	+8.6
G9	29.4	37.3	+17.6
G19	20.6	26.5	+2.9
G25	38.3	25	+11.7
G26	25	27.9	+1.5
G28	21.3	18.7	+8
PRAI+CAT	37.5	18.8	0

BIBLIOGRAFÍA

- Akanuma, S. y Yamagishi, A. (2005). Identification and characterization of key substructures involved in the early folding events of a $(\beta/\alpha)_8$ -barrel protein as studied by experimental and computational methods. *J. Mol. Biol.* 353: 1161–1170.
- Bogard, L. y Deem, M. (1999) A hierarchical approach to protein molecular evolution. *Proc. Natl. Acad. Sci. USA* 96: 2591–2595.
- Bono, S., Riechmann, L., Girard, E., Williams R. y Winter, G. (2005). A segment of cold shock protein directs the folding of a combinatorial protein. *Proc. Natl. Acad. Sci.* 102: 1396-1401.
- Buchanan, S., Smith, B., Venkatramani, B., Xia, D., Esser, L., Palnitkar, M., Chakraborty, R., Van der Helm, D. y Deisenhofer, J. (1999). Crystal structure of the outer membrane active transporter FepA from *Escherichia coli*. *Nature Structural Biology* 6: 59-63.
- Chothia C. (1992). One thousand families for the molecular biologist. *Nature* 357, 543–544.
- Copley, R. y Bork, P. (2000). Homology among $(\beta/\alpha)_8$ Barrels: Implications for the evolution of metabolic pathways. *J. Mol. Biol.* 303: 627-640.
- Cui, Y., Wong, W., Bornberg-Bauer, E. y Chan, H. (2002). Recombinatoric exploration of novel folded structures: A heteropolymer-based model of protein evolutionary landscapes. *Proc. Natl. Acad. Sci.* 99: 809-814.
- Eder, J. y Kirschner K (1992) Stable substructures of eightfold $\beta\alpha$ -barrel proteins: fragment complementation of phosphoribosylanthranilate isomerase. *Biochemistry* 31: 3617–3625.
- Farias Rico, JA. (2005). Generación de un andamiaje estable, diseño de un barril TIM consenso. Tesis Profesional. Facultad de Ciencias, UNAM, México D.F.
- Höcker, B., Jürgens, C., Wilmanns, M. y Sterner, R. (2001). Stability, catalytic versatility and evolution of the $(\beta/\alpha)_8$ -barrel fold. *Curr Opin Biotechnol* 12: 376–381.
- Hocker, B., Beismann-Driemeyer, S., Hettwer, S., Lustig, A. y Sterner R. (2001). Dissection of a $(\beta/\alpha)_8$ -barrel enzyme into two folded halves. *Nature Structural Biology* 8: 32-36.
- Holm, L. y Sander, C. (1996). Mapping the Protein Universe. *Science* 273: 595-602.
- Hommel, U., Eberhard, M. y Kirschner, K. (1995). Phosphoribosyl anthranilate isomerase catalyzes a reversible amadori reaction. *Biochemistry* 34: 5429-5434.
- Jasanoff, A., Davis, B. y Fersht, A. (1994). Detection of an intermediate in the folding of the $(\beta/\alpha)_8$ -barrel N-(5'-phosphoribosyl) anthranilate isomerase from *Escherichia coli*. *Biochemistry* 33: 6350–6355.

- Keller, M., Blochl, E., Wachtershauser, G. y Stetter K. (1994). Formation of amide bonds without a condensation agent and implications for origin of life. *Nature* 368: 836-838.
- Kinch, L. y Grishin, N. (2002). Evolution of protein structures and functions. *Current Opinion in Structural Biology* 12: 400-408.
- Kolkman, J. y Stemmer, W. (2001). Directed evolution of proteins by exon shuffling. *Nature Biotechnology* 19: 423-428.
- Koonin, E., Wolf, Y. y Karen, G. (2002). The structure of the protein universe and genome evolution. *Nature* 420: 218-223.
- Lang, D., Thoma, R., Henn-Sax, M., Sterner, R. y Wilmanns, M. (2000). Structural evidence for evolution of the β/α barrel scaffold by gene duplication and fusion. *Science* 289: 1546–1550.
- Luger, K., Hommel, U., Herold, M., Hofsteenge, J. y Kirschner, K. (1989) Correct folding of circularly permuted variants of a $\beta\alpha$ barrel enzyme *in vivo*. *Science* 243: 206–210.
- Luger, K., Szadkowski, H. y Kirschner, K. (1990). An 8-fold (β/α)-barrel protein with redundant folding possibilities. *Protein Eng* 3: 249–258.
- Maxwell, K., Mittermaier, A., Forman-Kay, J. y Davidson, A. (1999). A simple *in vivo* assay for increased protein solubility. *Protein Science* 8: 1908-1911.
- Nagano, N., Orengo, C. y Thornton, J. (2002). One Fold with Many Functions: The evolutionary relationships between TIM Barrel families based on their sequences, structures and functions. *J. Mol. Biol.* 321: 741–765.
- Patrick, W. y Blackburn, J. (2005). *In vitro* selection and characterization of a stable subdomain of phosphoribosylanthranilate isomerase. *FEBS Journal* 272: 3684–3697.
- Riechmann, L. y Winter, G. (2000). Novel folded protein domains generated by combinatorial shuffling of polypeptide segments. *Proc. Natl. Acad. Sci.* 97: 10068–10073.
- Riley, M., *et al.* (2006). Escherichia coli K-12: a cooperatively developed annotation snapshot—2005. *Nucleic Acids Research* 34: 1–9.
- Salem, G., Hutchinson, G., Orengo, C. and Thornton, J. (1999). Correlation of Observed Fold Frequency with the Occurrence of Local Structural Motifs. *J. Mol. Biol.* 287: 969-981.
- Shukla, A. y Guptasarma, P. (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 (β/α)₈-barrel. *Proteins* 55: 548-57.

Soberón, X., Fuentes-Gallego, P. y Saab-Rincón, G. (2004). In vivo fragment complementation of a $(\beta/\alpha)_8$ barrel protein: generation of variability by recombination. *FEBS Letters* 560: 167–172.

Söding, J. y Lupas, A. (2003). More than the sum of their parts: on the evolution of proteins from peptides. *BioEssays* 25: 837–846.

Sterner, R. y Höcker, B. (2005). Catalytic versatility, stability, and evolution of the $(\beta/\alpha)_8$ -barrel enzyme fold. *Chemical Reviews* 105: 4038-4055.

Urfer, R. y Kirschner, K. (1992). The importance of surface loops for stabilizing an eightfold beta alpha barrel protein. *Protein Science* 1: 31-45.

Vogel, C., Teichmann, S. y Pereira-Leal, J. (2005). The relationship between domain duplication and recombination. *J. Mol. Biol* 346: 355-365.

Voigt, C., Martinez, C., Wang, Z., Mayo, S. y Arnold, F. (2002). Protein building blocks preserved by recombination. *Nature Structural Biology* 9: 553-558.

Wierenga, R. (2001). The TIM-barrel fold: a versatile framework for efficient enzymes. *FEBS Letters* 492: 193-198.

Zitzewitz, J., Gualfetti, P., Perkons, I., Wasta, S. y Matthews, C. (1999). Identifying the structural boundaries of independent folding domains in the α subunit of tryptophan synthase, a β/α barrel protein. *Protein Science* 8: 1200–1209.

Zitzewitz, J y Matthews, C. (1999). Molecular dissection of the folding mechanism of the alpha subunit of tryptophan synthase: an amino-terminal autonomous folding unit controls several rate-limiting steps in the folding of a single domain protein. *Biochemistry* 38: 10205-10214.

