

**UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO**

FACULTAD DE CIENCIAS

**GENERACIÓN DE UN ANDAMIAJE ESTABLE
DISEÑO DE UN BARRIL TIM CONSENSO**

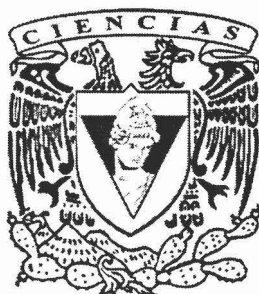
T E S I S
QUE PARA OBTENER EL TÍTULO DE :

B I Ó L O G O

P R E S E N T A:
JOSÉ ARCADIO FARÍAS RICO

DIRECTOR DE TESIS:
DR. LORENZO PATRICK SEGOVIA FORCELLA

2005



FACULTAD DE CIENCIAS
UNAM

m. 345027





UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

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

Comunicamos a usted que hemos revisado el trabajo escrito:

Generación de un andamiaje estable Diseño de un barril TIM consenso

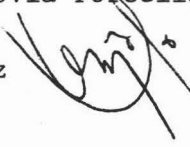
realizado por Farías Rico José Arcadio

con número de cuenta 09622447-0 , quien cubrió los créditos de la carrera de: Biología

Dicho trabajo cuenta con nuestro voto aprobatorio.


Atentamente

Director
Propietario Dr. Lorenzo Patrick Segovia Forcella 

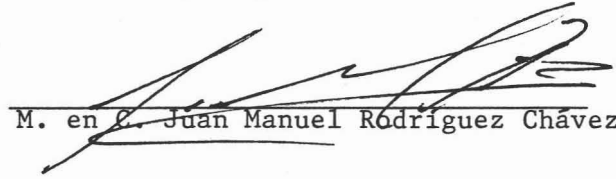
Propietario Dr. Enrique Merino Perez 

Propietario Dr. Joel Osuna Quintero 

Suplente Biol. Luis José Delaye Arredondo 

Suplente Q. Viviana Escobar Sánchez 

Consejo Departamental de Biología


M. en C. Juan Manuel Rodríguez Chávez

FACULTAD DE CIENCIAS



UNIDAD DE ENSEÑANZA
DE BIOLOGÍA

*Para Juan José y Guadalupe
Que son el principio*

*Para Víctor Hugo y León Fernando
Juntos somos la historia*

*Para Mariel
Con quien invento el mundo*

*Porque nada de esto es
tan importante
como todas las cosas que
he dejado de hacer*

*Where beauty is, then there is ugliness;
where right is, also there is wrong.
Knowledge and ignorance are interdependent;
delusion and enlightenment condition each other.
Since olden times it has been so.
How could it be otherwise?
Wanting to get rid of one and grab the other
is merely realizing a scene of stupidity.
Even if you speak of the wonder of it all,
how do you deal with each thing changing?
R.*

*he sat naked and drunk in a room of summer
night, running the blade of the knife
under his fingernails, smiling, thinking
of all the letters he had received
telling him that
the way he lived and wrote about
that--
it had kept them going when
all seemed
truly
hopeless.*

*putting the blade on the table, he
flicked it with a finger
and it whirled
in a flashing circle
under the light.*

*who the hell is going to save
me? he
thought.*

*as the knife stopped spinning
the answer came:
you're going to have to
save yourself.*

*still smiling,
a: he lit a
cigarette
b: he poured
another
drink
c: gave the blade
another
spin*

C.B.

Muchas gracias:

A Lorenzo, nada me gustaba tanto antes de tus clases.

A Viviana, me has enseñado todo.

A todo el grupo, juntos hacemos cosas chidas.

A mis amigos y a toda la gente que me ha enseñado algo.

INDICE

| | |
|--|-----------|
| I. RESUMEN | 1 |
| II. INTRODUCCION | |
| 1. Evolución | 2 |
| 2. Enzimas y Biocatálisis | 3 |
| 3. Evolución dirigida | 4 |
| 4. Diseño racional | 5 |
| III. ANTECEDENTES | |
| 1. Termoestabilidad | 6 |
| 2. El barril (β/α) ₈ | 9 |
| 3. Índice de adaptación de codones y el diseño de genes | 12 |
| 4. Purificación de proteínas | 13 |
| IV. OBJETIVOS | |
| General | 15 |
| Particulares | 15 |
| V. HIPOTESIS | 15 |
| VI. MATERIALES Y METODOS | |
| 1. Retrotraducción del gen | 16 |
| 2. Biología molecular | 16 |
| 3. Bioquímica | 24 |
| VII. RESULTADOS | |
| 1. Diseño de la proteína consenso | 27 |
| 2. Retrotraducción del gen | 32 |

| | |
|---|-----------|
| 3. Diseño de los oligonucleótidos y síntesis del nuevo gen | 35 |
| 4. Clonación del gen consenso | 36 |
| 5. Complementación de auxotrofia en medio sólido VB | 38 |
| 6. Complementación de auxotrofia en medio líquido VB | 39 |
| 7. Purificación de la proteína consenso | 40 |
| 8. Desnaturalización térmica | 42 |
| VII. DISCUSION | 44 |
| IX. CONCLUSIONES | 50 |
| X. PERSPECTIVAS | 50 |
| XI. REFERENCIAS | 51 |

I. RESUMEN

Las enzimas se encuentran adaptadas a su función particular en la célula. Muchas veces se necesitan variantes con características distintas a las naturales para diversos fines, tanto básicos como aplicados. La termoestabilidad es una característica muy importante que la ingeniería de proteínas intenta manipular. Es difícil predecir la contribución exacta de cada residuo a la estabilidad global del polipéptido, ya que esta propiedad es determinada por una multitud de interacciones, tanto locales como de rango amplio. El presente trabajo se vale de un acercamiento semiracional para obtener proteínas más estables, que consiste en la síntesis de un gen que codifica para la secuencia consenso de un grupo de proteínas homólogas. Se utilizó como modelo a la Fosforibosil antranilato isomerasa (PRAI), una enzima con plegamiento de barril TIM ya que la amplia distribución filogenética, su diversidad catalítica y tamaño hacen de este plegamiento un buen candidato para generar un andamiaje estable. Se diseñó una secuencia consenso utilizando 12 enzimas con actividad de PRAI homólogas. Se sintetizó el gen por ensamblaje de oligonucleótidos sintéticos y se purificó la nueva proteína. Este gen codifica para una nueva enzima con 49 cambios con respecto a la PRAI de *E. coli* y complementa a una cepa Δ PRAI. Resultados preliminares de experimentos de desnaturalización térmica indican un aumento en la T_m de 27°C de la nueva proteína consenso con respecto a PRAI de *E. coli*.

II. INTRODUCCION

*Nothing in Biology Makes Sense
Except in the Light of Evolution.
T.D.*

1. Evolución

Charles Darwin fue el fundador de las bases de la Biología evolutiva moderna al formular su teoría sobre el origen de las especies mediante la selección natural, (*On The Origin of Species by Means of Natural Selection*) en donde propone que la evolución ocurre cuando la selección natural actúa sobre la variación en las poblaciones. La evolución puede ser definida, de manera somera, como un cambio con herencia en función de las propiedades adaptativas de ese cambio. Desde el punto de vista poblacional evolución significa cambio en la poza genética (el grupo completo de genes de una especie o población) de una población a lo largo del tiempo (1). Darwin postuló su teoría sin saber el origen de la variabilidad en las poblaciones naturales.

Ahora sabemos que la variabilidad en las poblaciones es generada, entre otra cosas, por mutaciones espontáneas en los genes (el gen es la unidad hereditaria que puede ser transmitida de una generación a otra) de los individuos. Fue hasta finales de 1920 y principios de 1930 que se unificó la teoría de la herencia de Gregorio Mendel, (que funciona como la base de la genética moderna) con la evolución darwiniana. Los trabajos teóricos para la unificación de estas teorías fueron realizados independientemente por Ronald Fisher (1890-1962), John Haldane (1892-1964), y Sewell Wright (1889-1988) (1).

Esta unificación dio origen a la teoría sintética de la Evolución o Neo Darwinismo. En la síntesis moderna de la teoría evolutiva, como ya lo habíamos mencionado antes, evolución significa un cambio en las frecuencias génicas dentro de la poza genética de una población. Las principales fuerzas que operan en el proceso evolutivo son: mutación, selección natural, deriva génica y los cambios en la estructura de la población (2).

Aunque la evolución se ve reflejada al nivel de especie, las fuerzas evolutivas operan sobre los genes. La entidad que evoluciona y que podemos distinguir como tal es la especie; sin embargo, el material genético es el que moldea el fenotipo de la misma y los cambios que sufren los genes se ven reflejados en el organismo. De la fusión de la genética de poblaciones, (que proporciona las bases teóricas para el estudio de los procesos evolutivos) y la Biología Molecular (que proporciona los datos empíricos que soportan este tipo de estudios) surge la Evolución Molecular.

La evolución molecular tiene como objetos de estudio dos áreas bien diferenciadas:

- a) Los patrones y las tasas de cambio en el material genético, (así como los productos para los que codifica durante el tiempo evolutivo) y el mecanismo responsable de esos cambios (selección natural y acumulación de mutaciones por deriva génica)

- b) La historia evolutiva de los organismos y las macromoléculas inferida a través de datos moleculares. (*Molecular phylogenetics*)

El descubrir las causas y los efectos de los cambios evolutivos en las macromoléculas nos remite al estudio de la relación estructura y función de las proteínas. Las proteínas son el producto de los genes, y son los ladrillos constructores de toda la materia viva que existe sobre la tierra. El estudio de su estructura y cómo ésta se relaciona con las miles de funciones que hasta ahora les han sido asignadas, es uno de los campos de estudio más importantes de la Biología moderna.

2. Enzimas y Biocatálisis

Las enzimas son biomoléculas con propiedades sumamente interesantes. Aceleran la velocidad de las reacciones celulares, reduciendo la energía de activación y estabilizando los estados de transición de los intermediarios (3). Su especificidad les permite distinguir entre compuestos químicos tan solo por su estereoquímica (4). También pueden trabajar bajo condiciones ambientales extremas, haciendo posible la vida de organismos que se encuentran desde glaciales hasta ventilas hidrotermales en lo más profundo del océano (5). Por éstas y otras muchas cualidades, estas macromoléculas son el objeto de un estudio amplio para tratar de comprender el conjunto de características que las hacen extraordinarias.

Algunas de las propiedades de las proteínas se pueden entender en función de las características individuales de cada residuo, sin embargo aún no es posible establecer una relación clara entre la secuencia de aminoácidos y la estructura terciaria de los polipéptidos. Muchos grupos de investigación tratan descubrir como está encriptada la información en la secuencia de aminoácidos (6). Es difícil describir exactamente los patrones de aminoácidos que confieren la eficiencia catalítica, la afinidad por el sustrato, la estabilidad o la solubilidad, debido a que no tenemos aún la respuesta global al problema secuencia vs. estructura. Todo este conjunto de características se trata de modificar tanto con fines básicos, en el estudio de la relación estructura función, como aplicados al tratar de incorporar biocatalizadores en procesos industriales (7).

Probablemente uno de los problemas más grandes de la bioingeniería actual es establecer una relación clara entre la estructura y la función de los polipéptidos, debido a que las interacciones moleculares presentes en estas macromoléculas son demasiado complejas para ser descritas mediante los modelos que hasta el momento se han generado (8). El estudio de la relación estructura función es fundamental en el diseño y selección de nuevas variantes enzimáticas con características específicas que nos permitan acercarnos cada vez más a la meta final de la ingeniería de proteínas: el diseño *de novo* de enzimas con la función y propiedades requeridas para cualquier tipo de proceso (9). La biocatálisis es una de las áreas más prometedoras y astringentes en el campo de la ingeniería de proteínas, al tratar de definir los aminoácidos que confieren ciertas propiedades a las proteínas, por ejemplo termoestabilidad (10), estaremos más cerca de generar modelos que permitan diseñar biocatálisis *de novo*.

La ingeniería de proteínas se vale de varios acercamientos para modificar las propiedades de las enzimas, las dos principales vertientes son: a) evolución dirigida, que consiste en simular el proceso evolutivo de manera acelerada en el laboratorio y el b) diseño racional, que necesita un conocimiento más extenso de la relación estructura función, al realizar sustituciones específicas con base en el estudio de la estructura tridimensional. Ambas metodologías se discutirán más ampliamente a lo largo del texto. Últimamente ha surgido como una herramienta bastante poderosa la fusión de estos acercamientos obteniéndose grandes avances en el estudio de la relación estructura función(11).

3. Evolución dirigida

La evolución dirigida simula el proceso evolutivo en el laboratorio y en una escala de tiempo acelerada. Generalmente consiste en dos pasos principales: generar diversidad molecular e identificar las variantes mejoradas. Los métodos de generación de variabilidad más ampliamente utilizados son mutagénesis y recombinación *in vitro*. Una de las técnicas sumamente socorridas para generar mutagénesis al azar es el PCR mutagénico (12). Bajo las condiciones correctas (pH definido, presencia de cationes divalentes, mayor o menor concentración de DNA) la DNA polimerasa introduce mutaciones a una tasa controlada en cada ciclo de PCR para terminar con una población de variantes del gen de la proteína.

También existe la recombinación homóloga de genes (*gene shuffling*) (13). Mediante esta técnica es posible recombinar mutaciones positivas presentes en variantes múltiples de la proteína, en una reacción de PCR para generar diversidad, mientras que por otro lado es posible segregar mutaciones deletéreas al recombinar las variantes con el gen que codifica para la proteína silvestre (5). Las variantes generadas se muestrean con base en diversos marcadores o se seleccionan mediante filtros que permitan recuperar solamente las mutantes de interés.

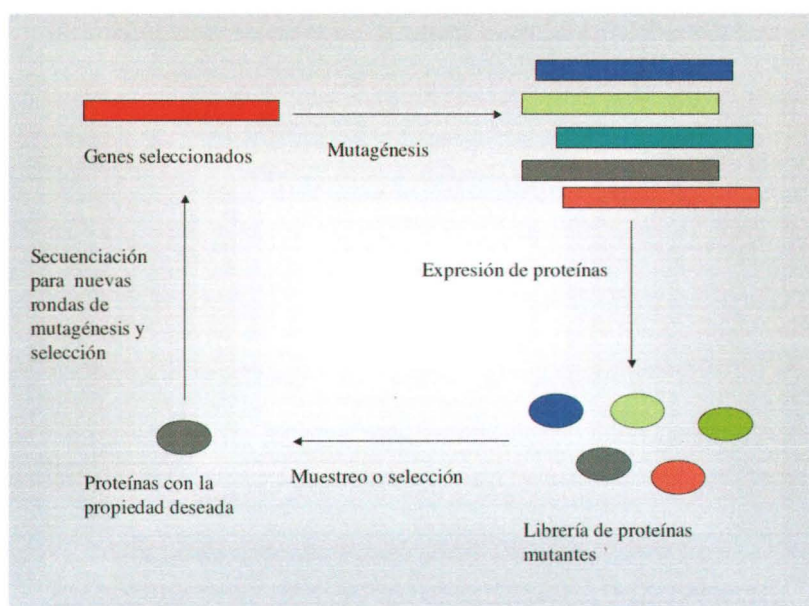


Figura 1. Evolución dirigida.

Los métodos utilizados en la selección y tamizaje de las mutantes de interés son variados: complementación de auxotrofia, resistencia a antibióticos, despliegue en fagos, despliegue en ribosomas, fusiones traduccionales, etc. En la metodología que utiliza la complementación de auxotrofia como sistema de selección, se establecen ciertas condiciones de cultivo en las cuales no pueden crecer cepas que han sido deletadas en el gen de interés. Posteriormente estas cepas se transforman con los bancos de variantes generados y se seleccionan las que hayan sido capaces de complementar la auxotrofia (14). Por otro lado, existen marcadores de muestreo que permiten distinguir clonas que cuenten con características específicas (plegamiento correcto, resistencia a pH alcalino) como son las fusiones traduccionales (8). El método más comúnmente utilizado es fusionar la enzima en cuestión a un reportero de plegamiento, como la proteína verde fluorescente; si la proteína que ha sido modificada por ingeniería, mantiene un plegamiento correcto, la GFP (*green fluorescent protein*) fluoresce y las bacterias que acarrean el plásmido con la construcción deseada pueden ser separadas del resto (15).

Es posible aplicar técnicas de evolución dirigida para lograr variantes más estables. Esto implica simular a una escala acelerada de tiempo el proceso evolutivo, generar diversidad en la secuencia y seleccionar la característica deseada. Hacer rondas de mutagénesis al azar, seleccionando la mejor mutante para volver a someterla a un nuevo proceso. En este tipo de acercamiento es necesario reducir eficientemente el número de mutaciones deletéreas o neutras, establecer claramente la diferencia entre los comportamientos enzimáticos dirigidos por los compromisos biológicos y las leyes de la fisicoquímica en el diseño de las técnicas de selección y búsqueda, así como el mejoramiento de las mismas. Todo esto para poder realizar caminatas más largas por el espacio de secuencia de la proteína en cuestión (8).

4. Diseño racional

El diseño racional es un acercamiento por demás diferente a la evolución dirigida. Es necesario contar con un cuerpo de conocimiento bastante amplio en referencia a las características que se pretenda modificar de la enzima en cuestión (15). Después de un análisis muy complejo de la estructura tridimensional de la enzima, se hacen sustituciones específicas que se cree modificarán sus parámetros cinéticos, por citar un ejemplo. El problema radica en determinar cuáles aminoácidos están involucrados en cierta característica y de qué manera lo hacen. Es necesario tomar en cuenta la modificación que la proteína sufrirá, tanto de manera local como global al realizar las sustituciones. Por ésta y otras tantas razones, el acercamiento racional requiere de algoritmos computacionales sumamente complejos.

Aunque el diseño racional se ha anotado algunos trabajos muy exitosos, no es posible hacer una generalización de su efectividad, ya que los modelos actuales que explican el comportamiento y los parámetros termodinámicos de una enzima, no son suficientes aún para discernir a ciencia cierta qué aminoácidos son determinantes para cierta función y que estos cambios a su vez, no alteren el plegamiento o la estabilidad de manera global de la enzima que se intenta modificar.

El diseño de proteínas nuevas ha evolucionado de la mano con nuestro entendimiento de la función y estructura biomacromolecular (9). Este crecimiento se da gracias al avance en las metodologías computacionales y de biología molecular que están disponibles hoy en día. Cada vez tenemos mejores modelos que explican los parámetros termodinámicos con gran calidad y de manera integradora. El problema de la relación estructura función es la barrera más grande que hay que salvar para poder generar enzimas con características nativas (actividad específica cercana a la enzimas naturales, o gran estabilidad, etc.) Con cada trabajo nos acercamos más a un modelo integrador de todo el conocimiento actual sobre las proteínas, ampliando nuestras perspectivas en la exploración del universo proteico.

III. ANTECEDENTES

1. Termoestabilidad

Dado que la función de una proteína depende de su estructura, y que ésta estructura nativa se adquiere cuando el péptido recién sintetizado viaja a través de un paisaje de energía para descender hasta un mínimo global único, podemos decir que la estabilidad es la medida métrica experimental de la información contenida en una secuencia polipeptídica (16). Por lo anterior, es de suma importancia entender los parámetros que están gobernando este tipo de fenómenos y no solo en el aspecto biotecnológico, sino en la más pura ciencia básica.

Debido a la potencial aplicación de las enzimas en procesos industriales a gran escala, la termoestabilidad se encuentra entre las características enzimáticas más importantes que se han intentado optimizar. En este tipo de procesos se requiere que las enzimas trabajen a temperaturas altas, ya sea para incrementar su número de recambio (17) aumentando la velocidad de reacción, o también para elevar la solubilidad de los reactantes, mientras que una ventaja de desarrollar procesos a altas temperaturas es que se evita la contaminación por microorganismos (3). Los biorreactores no son los únicos procesos beneficiados por la generación de enzimas termoestables; en procedimientos de limpieza a gran escala se utilizan detergentes adicionados con enzimas que deben soportar grandes temperaturas, así como en la industria alimenticia ciertos alimentos son complementados con enzimas que facilitan la degradación completa de compuestos que en otras condiciones no se aprovecharían del todo (18). Está claro que la obtención de enzimas termoestables es de suma importancia para la ingeniería de proteínas.

Es necesario generar modelos que permitan distinguir el tipo de compromiso al cual se halla sujeta una proteína, pudiendo ser termodinámico o biológico. Sin embargo, esta es una de las tareas más difíciles para la ciencia moderna. La causa de que una proteína en específico no haya visto incrementada su termoestabilidad a lo largo de la evolución, puede deberse a una combinación en las presiones de selección que sólo se estén enfocando en ciertos parámetros termodinámicos de la misma y no estén optimizando la termoestabilidad, dado que ésta se encuentra en el umbral necesario para realizar sus funciones (16).

El diseño de proteínas nuevas ha evolucionado de la mano con nuestro entendimiento de la función y estructura biomacromolecular (9). Este crecimiento se da gracias al avance en las metodologías computacionales y de biología molecular que están disponibles hoy en día. Cada vez tenemos mejores modelos que explican los parámetros termodinámicos con gran calidad y de manera integradora. El problema de la relación estructura función es la barrera más grande que hay que salvar para poder generar enzimas con características nativas (actividad específica cercana a la enzimas naturales, o gran estabilidad, etc.) Con cada trabajo nos acercamos más a un modelo integrador de todo el conocimiento actual sobre las proteínas, ampliando nuestras perspectivas en la exploración del universo proteico.

III. ANTECEDENTES

1. Termoestabilidad

Dado que la función de una proteína depende de su estructura, y que ésta estructura nativa se adquiere cuando el péptido recién sintetizado viaja a través de un paisaje de energía para descender hasta un mínimo global único, podemos decir que la estabilidad es la medida métrica experimental de la información contenida en una secuencia polipeptídica (16). Por lo anterior, es de suma importancia entender los parámetros que están gobernando este tipo de fenómenos y no solo en el aspecto biotecnológico, sino en la más pura ciencia básica.

Debido a la potencial aplicación de las enzimas en procesos industriales a gran escala, la termoestabilidad se encuentra entre las características enzimáticas más importantes que se han intentado optimizar. En este tipo de procesos se requiere que las enzimas trabajen a temperaturas altas, ya sea para incrementar su número de recambio (17) aumentando la velocidad de reacción, o también para elevar la solubilidad de los reactantes, mientras que una ventaja de desarrollar procesos a altas temperaturas es que se evita la contaminación por microorganismos (3). Los biorreactores no son los únicos procesos beneficiados por la generación de enzimas termoestables; en procedimientos de limpieza a gran escala se utilizan detergentes adicionados con enzimas que deben soportar grandes temperaturas, así como en la industria alimenticia ciertos alimentos son complementados con enzimas que facilitan la degradación completa de compuestos que en otras condiciones no se aprovecharían del todo (18). Está claro que la obtención de enzimas termoestables es de suma importancia para la ingeniería de proteínas.

Es necesario generar modelos que permitan distinguir el tipo de compromiso al cual se halla sujeta una proteína, pudiendo ser termodinámico o biológico. Sin embargo, esta es una de las tareas más difíciles para la ciencia moderna. La causa de que una proteína en específico no haya visto incrementada su termoestabilidad a lo largo de la evolución, puede deberse a una combinación en las presiones de selección que sólo se estén enfocando en ciertos parámetros termodinámicos de la misma y no estén optimizando la termoestabilidad, dado que ésta se encuentra en el umbral necesario para realizar sus funciones (16).

Los ensayos de evolución dirigida en el laboratorio nos dan ciertos indicios acerca de los determinantes de estabilidad en las proteínas (19). Después de varios ciclos de evolución a los que se ve sometido un polipéptido, generalmente se favorece la aparición de residuos aromáticos para mejorar el empacamiento del núcleo hidrofóbico en la parte interna de la proteína, así como un incremento en los puentes salinos (5). En el análisis cristalográfico comparativo entre proteínas mesófilas y termófilas, generalmente se tratan de descubrir las diferencias entre ambas proteínas homólogas y se procede a sustituir los determinantes de estabilidad inferidos en la proteína mesófila. La complejidad que involucra realizar sustituciones puntuales basados únicamente en estudios sobre enzimas termófilas es muy grande. No se sabe con exactitud cuál será el efecto global de la presencia de un aminoácido en específico en un contexto distinto al original.

Es muy claro que las enzimas se encuentran adaptadas a su función particular en la célula; una enzima de un organismo mesófilo no ha sido seleccionada a lo largo de la historia evolutiva para trabajar bajo condiciones de temperatura extrema. La selección natural únicamente ha operado para que sea funcional a la temperatura a la que se desarrolla el organismo que la posee o para desarrollar la función bajo ciertas condiciones específicas (16).

Los acercamientos racionales para incrementar la termoestabilidad de las proteínas son sumamente difíciles, debido a la complejidad que implica calcular la contribución exacta de cada aminoácido a la estabilidad global del polipéptido. Es bien sabido que la termoestabilidad no se puede incrementar mediante las metodologías que funcionan para cambiar una especificidad de sustrato, o para lograr una mayor eficiencia catalítica. Estos acercamientos tratan de cambiar propiedades enzimáticas que se infieren contenidas por unos cuantos aminoácidos más o menos bien identificados, sin embargo este tipo de modificaciones siguen siendo una de las pruebas más astringentes del conocimiento proteico actual. La estabilidad de una proteína está determinada por una multitud de interacciones tanto locales como de distancias más largas (20). Esto significa que para lograr generar una variante más termoestable, es necesario cambiar muchas posiciones del polipéptido, o saber exactamente cuáles son los aminoácidos clave que le confieren un aumento significativo en la estabilidad.

En un estudio (21) en el cual se realizó la comparación de las proteínas de choque frío entre el termófilo *Bacillus caldolyticus* y el mesófilo *Bacillus subtilis*, se encontró que estas proteínas diferían solo en 12 de 67 aminoácidos mostrando una diferencia muy grande en estabilidad (15.8 kJ/mol) a 70° C. La posterior mutagénesis de los 12 residuos en la proteína de *Bacillus Caldolyticus* mostró que en sólo dos sustituciones Glu3Arg y Glu66Leu en la superficie de la proteína, estaba sostenida la diferencia en termoestabilidad. Lo que nos muestra la dificultad en la predicción de sustituciones estabilizantes.

En los últimos años ha surgido un acercamiento en el problema de la estabilidad que es del tipo semirracional. Podemos contar varios ejemplos en la aplicación de esta metodología que han resultado exitosos, como fue la estabilización de la proteína P53 en el cual se realizó un alineamiento de 23 secuencias homólogas de este factor de transcripción y se cambiaron 20 aminoácidos por residuos consenso en el sitio de unión a DNA (22). Después de analizar las mutaciones de manera independiente, se pudo definir que cuatro mutaciones eran en realidad estabilizantes (Met133Leu, Val203Ala, Asn239Tyr, Asn268Asp), logrando un incremento en termoestabilidad de 2.65 kcal/mol y un incremento en la T_m de 5.6° C.

La mejor explicación teórica de este concepto hasta el momento, está dada por el grupo de Boris Steipe. Ellos logran predecir mediante estadística termodinámica 6 mutaciones estabilizantes de un total de 10 en el dominio de unión de inmunoglobulinas (Ig) a través de un acercamiento denominado aproximación de la secuencia canónica (16). Para entender este concepto es necesario recordar que la selección juega un papel muy importante en la evolución de los dominios de unión de Ig. La adecuación que se observa en estas estructuras está compuesta por múltiples factores, entre los cuales hallamos afinidad por el antígeno, ensamblaje de las cadenas ligeras y pesadas, resistencia a las proteasas y competencia en la secreción (24). Cada factor contribuirá de manera diferencial dependiendo de cada posición que se estudie en el dominio.

La hipótesis de este grupo de trabajo es que el repertorio de inmunoglobulinas se aproxima a un *conjunto canónico de secuencias*, cada una generada a partir de un grupo de secuencias germinales, que son seleccionadas en un proceso de mutaciones al azar, independientes para ser compatibles con cada aspecto de la función de los anticuerpos (24). Se esperaría que este conjunto cuente con dos propiedades fundamentales 1) el nivel promedio de estabilidad del dominio es marginal y 2) el conjunto se encuentra en un estado de equilibrio con respecto a los cambios en la secuencia que afectan la estabilidad.

En otras palabras, mientras que las mutaciones desestabilizantes no afecten la proteína más allá de un nivel en el cual se vea comprometida su estructura global, serán selectivamente neutrales. Las mutaciones estabilizantes al azar son altamente improbables, debido a que no existe una presión de selección por abajo de cierto umbral. A este conjunto de secuencias canónicas se le puede aplicar la ecuación de distribución de Boltzman. Esta ecuación describe cuantitativamente como la energía de un sistema es distribuida entre los estados de sus componentes. Los estados de menor energía son más frecuentes que los estados de mayor energía.

En esta misma línea hacemos referencia a un conjunto de dominios de inmunoglobulinas, ya que si se superimpone un número de dominios variables de estas proteínas a una baja resolución, éstos se verán similares mimetizándose como un solo plegamiento. Podemos imaginar esta súper imposición como un medio ambiente promedio en el cual se evalúan los cambios en los aminoácidos que serían al azar e independientes. La adecuación de cada dominio se puede asumir casi constante debido a que no existen mutaciones que afecten de manera fatal la estabilidad de la proteína. En la base de datos utilizada no hay anticuerpos no funcionales.

Entonces podríamos decir que la más probable distribución de frecuencias de aminoácidos es una distribución de Boltzman y el residuo consenso es el más adecuado (más estable).

Todo esto adquiere sentido si tomamos un grupo de secuencias que tienen un mismo plegamiento. En el marco de la termodinámica estadística, este plegamiento puede ser visto como un sistema. Cada una de las posiciones de los residuos individuales serían componentes o partículas de este sistema. Cada componente podría ocurrir en 21 microestados distintos, que son los 20 aminoácidos que forman las proteínas o un gap. El grupo de todas las secuencias específicas o macroestados que son compatibles con este plegamiento sería un conjunto canónico. La secuencia específica que mejor represente ese conjunto es la secuencia consenso (16).

Revisando varios trabajos que se han realizado hasta el momento encontramos que al generar un alineamiento de secuencias de 130 miembros de la familia de la chaperonina 60, se diseñaron 34 mutaciones puntuales y se logró la generación de minichaperonas estables (23). La predicción de una sustitución en particular estabilizante A233T resultaba prácticamente, imposible si el estudio se hubiese basado en un análisis estructural, sin embargo el acercamiento por homología permitió descubrirlo.

En otro estudio realizado por Lehmann se diseñó una fitasa consenso comparando 13 secuencias de proteínas de hongos, para obtener una enzima consenso final con un incremento de 22° C en la Tm (18). Este es uno de los ejemplos más impactantes en este campo, debido a que de manera muy simple y rápida se logró un aumento bastante significativo en la termoestabilidad. En un estudio posterior de este grupo (20), se logró obtener mediante la adición de más secuencias y el análisis individual de cada una de las mutaciones, un aumento final en la estabilidad, llevando la fitasa consenso original desde una Tm de 78° C hasta una secuencia consenso que presentaba una Tm de 90.4° C .

En resumen, podemos decir que el concepto del consenso se basa en la hipótesis de que en una posición dada de un alineamiento de secuencias de proteínas homólogas, el respectivo aminoácido consenso contribuye en mayor medida a la estabilidad promedio de la proteína que el aminoácido no consenso.

2. El Barril (β/α)₈

El plegamiento de barril TIM es el más común en las bases de datos de estructuras de proteínas. También es visto en muchas diferentes familias de proteínas catalizando reacciones totalmente disimilares (25). El plegamiento de barril (β/α)₈ está constituido por 8 hojas β plegadas rodeadas por 8 α hélices. En un barril (β/α)₈ cerrado se detecta un anillo completo de puentes de hidrógeno, cada hoja beta esta formando puentes de hidrogeno con sus dos vecinas, todas están conectadas, y en general se arreglan de una manera antiparalela. Una excepción es el plegamiento de la TIM (Triósa Fosfato Isomerasa) que se arregla de manera paralela, con 8 hojas beta que están cubiertas por 8 alfa hélices en el exterior (figura 1).

El número de hojas beta varía, pueden ser desde cuatro hasta veinte; estas hojas pueden estar orientadas de forma antiparalela, paralela o mezclada. El clásico barril TIM consta de ocho repeticiones de unidades beta/alfa, en el cual cada hoja beta en el interior está cubierta por una hélice alfa en el exterior (25). En este caso las hojas beta se encuentran orientadas de forma paralela. Este plegamiento se descubrió en la Triósa Fosfato Isomerasa, la enzima encargada de convertir gliceraldehido 3 fosfato en dihidroxiacetona, dentro de la vía glucolítica.

Las hojas beta y las alfa hélices están conectadas por asas que nombradas en función de su posición, pueden ser asas β/α o asas α/β . Estudios tanto de tipo estructural como a nivel 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 (28), indicando que la más pequeña unidad evolutiva, dentro de este sistema es la mitad del barril.

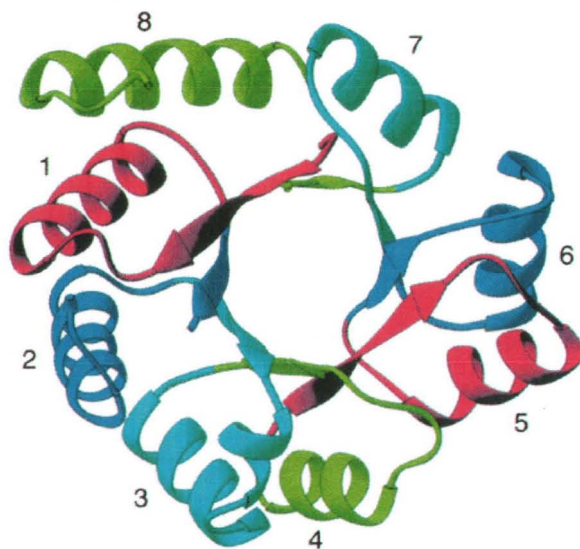


Figura 2. Plegamiento $(\alpha/\beta)_8$ (28).

Los sitios activos de todas las enzimas con el plegamiento de barril TIM se encuentran en los extremos carboxilo terminales de las hojas β , implicando que la geometría de los sitios activos está determinada por las ocho asas que siguen después de las hojas β . Las asas β/α pueden ser de tamaño variable pero tienden a ser más largas que las α/β , como en la Triósa Fosfato Isomerasa (27). Mediante modelación y el diseño de asas en el extremo carboxilo terminal se han obtenido evidencias de los papeles que juegan ambos tipos de asas, tanto las β/α como las α/β . Se ha demostrado que las asas que no están implicadas en la catálisis están jugando un papel fundamental al mantener la estabilidad estructural (29). Esto permite intercambiar las asas del carboxilo terminal sin comprometer la estabilidad del plegamiento al realizar experimentos de evolución dirigida.

Las proteínas que presentan el plegamiento de barril TIM son comúnmente enzimas o están relacionadas con ellas. Las actividades de estas proteínas se presentan ligadas con 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.

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 plegamiento es la de ligasa (figura 2).

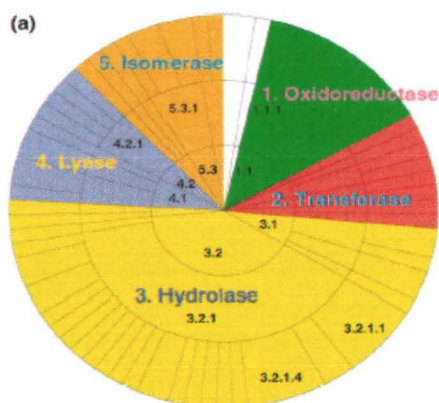


Figura 3. Actividades encontradas en las enzimas que presentan el plegamiento de barril TIM. Se indican los números enzimáticos y la frecuencia relativa en que se presentan las funciones (27).

Alineamientos por pares entre secuencias de familias no homólogas que presentan el plegamiento de barril TIM, muestran un porcentaje de identidad indetectable. Sin embargo, muchos estudios de tipo estructural y de motivos conservados en residuos comunes a muchas enzimas con este plegamiento, han generado evidencia para argumentar un posible origen monofilético en este plegamiento. Mediante la generación de perfiles en búsquedas utilizando *psi blast* y con el uso de alineamientos estructurales para buscar patrones en la aparición de residuos conservados en muchas de las familias, se han logrado unir 21 superfamilias de proteínas con plegamiento de barril TIM (27). El conjunto de propiedades que se presenta en el plegamiento de barril $(\beta/\alpha)_8$ lo convierten en un modelo interesante, al tratar de entender la relación estructura función de las proteínas.

La Fosforibosil Antranilato Isomerasa o PRAI es una enzima que presenta el plegamiento clásico de barril TIM, en *E. coli* se encuentra formando una enzima bifuncional que consta de dos dominios (30), implicados en la vía de síntesis de triptofano. Uno de los dominios cataliza una reacción intramolecular redox, conocida también como rearreglo amadori. En ésta la enzima toma como sustrato el N-(5' fosforibosil) antranilato (PRA) y lo transforma en 1-[(2- carboxifenil)amino]-1 deoxiribulosa 5-fosfato (CdRP). Este compuesto funciona a su vez como sustrato putativo para el otro dominio de la enzima (IGPS), que lo convierte en Indol glicerol fosfato (30).

Se han postulado varias hipótesis que expliquen la fusión de estos dos dominios que, mediante técnicas de ingeniería recombinante, han sido separados en enzimas monofuncionales con la misma actividad que su contraparte bifuncional (31). Uno de los principales argumentos para explicar la fusión de dominios, es la estabilización del plegamiento por oligomerización. Sin embargo este tipo de estrategias se ha observado más frecuentemente en enzimas de organismos extremófilos (32) y *E. coli* es un mesófilo clásico. Tal vez la explicación más plausible radica en la naturaleza de los sustratos para ambas enzimas, PRA es un compuesto muy inestable en condiciones fisiológicas al igual que CdRP, por lo que una canalización eficiente de los intermediarios optimiza en gran medida la vía biosintética.

3. Índice de adaptación de codones y el diseño de genes

El CAI (Codon Adaptation Index) es una medida efectiva de las tendencias en el uso de codones (33). Se genera a partir de un grupo de genes con alto nivel de expresión para asignar un mérito relativo a cada codon, y calcular una calificación para un gen en específico a partir de la frecuencia de uso de todos los codones en ese gen. El índice refleja la fuerza de la selección al modular el patrón del uso de codones.

En ese sentido, resulta útil para predecir el nivel de expresión de un gen, para evaluar la adaptación de los genes virales a su hospedero y para hacer comparaciones del uso de codones en diferentes organismos (34). El índice puede ofrecer una indicación aproximada del éxito que podría tener la expresión de genes heterólogos en sistemas específicos.

El avance en la determinación de la secuencia en muchos genes de una variedad amplia de especies, ha demostrado que los codones no se utilizan al azar. Las tendencias en el uso de codones son especie específicas y más aún, gen específicas. Otro descubrimiento muy importante es que existe una correlación sumamente marcada entre la tendencia en el uso de codones y el nivel de expresión (34). Existe evidencia experimental de que el uso de codones del organismo hospedero puede afectar el nivel de expresión de un gen heterólogo (34). Mimetizando el uso de codones para un gen exógeno se puede optimizar la expresión del mismo. Este aspecto es una herramienta sumamente útil en las aplicaciones biotecnológicas, ya que determinará en gran medida la correcta expresión del gen.

El uso de codones ha sido identificado como el factor más importante en la expresión de genes exógenos en procariontes, debido a que los codones preferidos correlacionan de manera expresa con la existencia de los tRNAs cognados (34). Genes que presentan codones raros en la mayoría de la secuencia se transcriben ineficientemente debido a que la poza de tRNAs cognados es somera, en consecuencia, la traducción del mensajero se ve pausada hasta que el tRNA para el codon raro encuentre su triplete complementario. En muchos trabajos que expresan genes en sistemas eucariontes se da una optimización del mismo utilizando codones de uso muy frecuente en el organismo hospedero. Algunas veces esta aproximación ocasiona que la expresión sea anómala al eliminar pausas durante la traducción.

Existe una hipótesis que establece una relación entre las pausas en la traducción mediadas por la inclusión de codones raros y el plegamiento de los polipéptidos (8). El uso de codones raros en la secuencia de una proteína está determinando pausas en el plegamiento cotraduccional, estas pausas serían necesarias para que la estructura tridimensional de la proteína pudiera ser alcanzada de manera óptima. En este sentido, el mimetizar las curvas del CAI del gen heterólogo con las curvas presentadas por los genes del organismo hospedero, aseguraría una expresión adecuada de la proteína exógena.

En el proceso de diseño de genes existen varios pasos fundamentales: ajustar el uso de codones para mimetizar el CAI de los genes del hospedero, eliminar secuencias repetitivas, eliminar estructuras desfavorables en el mRNA, adicionar o remover sitios de restricción, finalmente el diseño de oligonucleótidos para el ensamblaje del gen con características específicas que permitan construir la secuencia de manera óptima (34). Todos los pasos anteriores pueden ser resueltos con herramientas bioinformáticas. Posteriormente se procede a ensamblar la secuencia, clonarla en el vector adecuado verificando que sea la secuencia diseñada mediante secuenciación, finalizando el proceso al transformarla en el sistema de expresión elegido.

4. Purificación de proteínas

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 (36). 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 mismos en función de sus propiedades fisicoquímicas individuales, con su posterior separación.

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 contienen poros en un rango relativo de tamaños. Si una solución acuosa es pasada por una columna que contiene estos poros de diferentes tamaños, los solutos presentarán una elusión diferencial (36).

La nuevas técnicas de DNA recombinante permiten la construcción de proteínas fusionadas a marcas de afinidad específicas; estas marcas facilitan la purificación de la secuencia polipeptídica de interés utilizando columnas cromatográficas de afinidad. La marca más común que se agrega a las proteínas que se pretenden purificar es una serie de 6 residuos de histidinas. Esta marca tiene una 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 se quedan unidas al metal níquel, mientras que los demás polipéptidos que no están fusionados pasan por la columna separándose de la proteína problema. La proteína se recupera con un gradiente creciente de imidazol al eluir las diferentes fracciones.

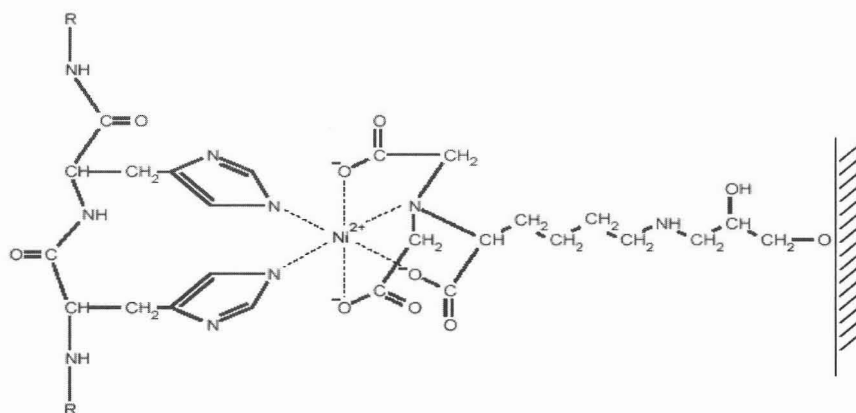


Figura 4. Interacción entre residuos adyacentes de Histidinas con la matriz de níquel (45).

IV. OBJETIVOS

General

Queremos diseñar y construir un barril TIM termoestable utilizando el concepto de consenso de secuencias homólogas. Este tipo de andamiaje puede ser utilizado posteriormente en experimentos de evolución dirigida y diseño racional al ofrecer mayor estabilidad desde el inicio.

Particulares

Definir una secuencia consenso de aminoácidos a partir de alineamientos de secuencias homólogas.

Definir la secuencia nucleotídica del gen consenso respetando las preferencias de uso de codones del gen silvestre de *E. coli*.

Sintetizar el gen consenso por ensamblaje de oligonucleótidos sintéticos.

Identificación de clonas que expresen una proteína plegada.

Caracterizar genética- y biofísicamente la nueva proteína consenso.

V. HIPOTESIS

Es posible diseñar barriles $(\beta/\alpha)_8$ termoestables mediante la generación de una secuencia consenso de proteínas homólogas

VI. MATERIALES Y METODOS

1. Retrotraducción del gen

Anterior a la síntesis del gen consenso, la secuencia de aminoácidos de la nueva proteína se utilizó como templado para obtener la secuencia de DNA valiéndonos del uso de codones de *E. coli* para asegurar la expresión correcta de la nueva proteína en la bacteria hospedera. La metodología utilizada fue la siguiente:

Evaluar la frecuencia de utilización de un codon en específico y elegir el codon que codificaba para el aminoácido de la proteína consenso, tratando de utilizar la frecuencia más similar tomando como referencia la base de datos de uso de codones de *E. coli*. La metodología se explica con más detalle en la sección de resultados.

Escherichia coli K12 [gbcbt]: 5054 CDS's (1603901 codons)

fields: [triplet] [amino acid] [fraction] [frequency: per thousand] ([number])

| | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|
| UUU F 0.57 22.4 (35930) | UCU S 0.15 8.5 (13633) | UAU Y 0.57 16.3 (26180) | UGU C 0.45 5.2 (8306) |
| UUC F 0.43 16.6 (26609) | UCC S 0.15 8.6 (13783) | UAC Y 0.43 12.3 (19675) | UGC C 0.55 6.4 (10330) |
| UUA L 0.13 13.9 (22279) | UCA S 0.12 7.1 (11438) | UAA * 0.64 2.0 (3244) | UGA * 0.29 0.9 (1448) |
| UUG L 0.13 13.7 (22000) | UCG S 0.15 8.9 (14305) | UAG * 0.07 0.2 (365) | UGG W 1.00 15.3 (24553) |
| CUU L 0.10 11.0 (17707) | CCU P 0.16 7.0 (11291) | CAU H 0.57 12.9 (20686) | CGU R 0.38 21.0 (33711) |
| CUC L 0.10 11.0 (17715) | CCC P 0.12 5.5 (8861) | CAC H 0.43 9.7 (15595) | CGC R 0.40 22.0 (35311) |
| CUA L 0.04 3.9 (6182) | CCA P 0.19 8.5 (13664) | CAA Q 0.35 15.5 (24787) | CGA R 0.06 3.5 (5668) |
| CUG L 0.50 52.8 (84714) | CCG P 0.52 23.3 (37316) | CAG Q 0.65 28.8 (46256) | CGG R 0.10 5.4 (8631) |
| AUU I 0.51 30.4 (48766) | ACU T 0.17 8.9 (14303) | AAU N 0.45 17.6 (28256) | AGU S 0.15 8.7 (13976) |
| AUC I 0.42 25.0 (40097) | ACC T 0.44 23.4 (37495) | AAC N 0.55 21.7 (34752) | AGC S 0.28 16.0 (25716) |
| AUA I 0.07 4.3 (6866) | ACA T 0.13 7.0 (11267) | AAA K 0.77 33.6 (53920) | AGA R 0.04 2.1 (3291) |
| AUG M 1.00 27.8 (44539) | ACG T 0.27 14.4 (23056) | AAG K 0.23 10.2 (16370) | AGG R 0.02 1.2 (1949) |
| GUU V 0.26 18.4 (29487) | GCU A 0.16 15.3 (24609) | GAU D 0.63 32.2 (51670) | GGU G 0.34 24.9 (39862) |
| GUC V 0.22 15.2 (24406) | GCC A 0.27 25.5 (40914) | GAC D 0.37 19.1 (30559) | GGC G 0.40 29.4 (47212) |
| GUA V 0.15 10.9 (17443) | GCA A 0.21 20.3 (32529) | GAA E 0.69 39.6 (63484) | GGA G 0.11 7.9 (12696) |
| GUG V 0.37 26.2 (42097) | GCG A 0.36 33.7 (53984) | GAG E 0.31 17.8 (28529) | GGG G 0.15 11.0 (17628) |

Figura 4. Base de datos de uso de codones para *E. coli* (40).

2. Biología molecular

Cepas

Se utilizaron las siguientes cepas de *E. coli*:

XL1blue – MRF'

JM101

JMB9 F' trpΔ

Se prepararon células electrocompetentes de todas las cepas (Tung 1995)

Medios de cultivo

Luria Bertani (37)

Vogel Bonner, medio mínimo carente de triptofano (38)

Preparación 250 ml

5ml Glucosa 20%

5ml Casaminoácidos 20% (esterilizados por autoclave y no filtrados)

12.5 ml Sales VB 20x

250 μ l FeCl₃ 5 mg/ml

62.5 μ l Tiamina 10 mg/ml

Para preparar placas agregar 6 gramos de Agar Noble (*Sigma*)

Sales VB 20x 100ml (38)

MgSO₄ * 7H₂O 0.4 g

ácido cítrico anhidro 3.94 g

K₂HPO₄ 20 g

Na₂HPO₄ * 2H₂O 3 g

Ajustar a pH 7 con KOH

Antibióticos soluciones stock (1000x)

Kanamicina 25 mg/ml

Tetraciclina 12.5 mg/ml,

Cloramfenicol 10, 15, 25 y 35 mg/ml

Ampicilina 200 mg/ml,

Plásmidos

pUC 18 Gilbert, W. (1991).

Vector de clonación de 2686 pb con marcador de resistencia a ampicilina. Marcador de Beta galactosidasa. Plásmido de alto número de copias.

pT4 (construido en el lab. del Dr. Hilvert , ETH , Suiza. Donado por el Dr. Joel Osuna Vector de expresion en *E. coli* de 1845 pb. Conteniendo el marcador de resistencia a kanamicina y origen de replicación pUC. Plásmido de alto número de copias. Contiene promotor ptrc.

Electroforesis

Se utilizaron geles de agarosa (*Biorad*) al 1% , 1.5%,y 2% en solución amortiguadora TAE grado DNA (*Biorad*).

Enzimas

Para clonar el gen silvestre en el vector de trabajo utilizamos las enzimas *BamHI* y *NdeI* (*New England Biolabs*). En la revisión de la inserción de plásmido utilizamos las enzimas *XhoI* y *EcoRv* (*New England Biolabs*). En los experimentos de ligación se utilizó *T4* DNA ligasa (*Roche*).

Purificación de plásmido

Se inoculó una colonia aislada en 3 ml de medio mínimo LB con kanamicina 35mg/ml y se incubó durante 16 horas. Posteriormente se colectaron células por centrifugación y se purificó el plásmido mediante lisis alcalina con soluciones y columnas del High Pure Plasmid Isolation Kit de Roche. En los experimentos de purificación de plásmido se utilizó el kit de Roche (High Pure Plasmid Isolation Kit). Para extraer plásmido de géles de agarosa se utilizó el kit de Qiagen (QIAquick Gel Extraction Kit).

Purificación de productos de PCR

Los productos de PCR se purificaron utilizando soluciones del kit High Pure PCR Product Purification de Roche.

Síntesis de oligonucleótidos y secuenciación

La fabricación de los 16 oligonucleótidos utilizados en este trabajo y la secuenciación de las clonas, se realizó en la unidad de síntesis y secuencia del Instituto de Biotecnología de la UNAM. Se diseñaron y sintetizaron los siguientes oligonucleótidos para la clonación de los genes consenso y silvestre:

| | |
|----------|---|
| NdePRAI | ATA TAT ACA TAT GCT GGA GAA TAA GGT ATG |
| BamPRAIx | GCG GAT CCG TAT GCG CGC AGC GTC |

Se diseñaron los siguientes oligonucleótidos para la síntesis del gen consenso, mediante el programa PRIMO de *Chang Bioscience* (39).

| | |
|---------|--|
| PRAIx1 | ATG CTG GAG AAT AAG GTA TGT GGC CTG ACG CGT GGG CAA GAT GCT AAA GCA GTA TAT AAG |
| PRAIx2 | TGA TGA CTT AAC AAA AAT CAA CCC ACC GTA AAT CGC GCC CGC CTT ATATAC TGC TTT AGC |
| PRAIx3 | GAT TTT TGT TAA GTC ATC AAA ACG TTA TGT CAG CGT TAA ACA GGC GCAGGA AAT CAT GGC |
| PRAIx4 | TAT CGT GAT TTT GGA ACA CGC CAA CAA AGC GCA ACG GTG CCG CAG CCATGA TTT CCT GCG |
| PRAIx5 | TGT TCC AAA ATC ACG ATA TTG CCG ATA TCG TGG ACA TTG CTA AGA AGT TAT CGC TGT ACG |
| PRAIx6 | TGA TCG ATA TAC TCC TGA TCT TCA TTA CCA TGC AGT TGC ACT GCG TACAGC GAT AAC TTC |
| PRAIx7 | CAG GAG TAT ATC GAT CAA CTG CGT AAA GCT CTG CCA AAG CAT ATC GCCATC TGG AAA GCA |
| PRAIx8 | TTA TTA TCA AAA TTC AGG GCG GGC AGG GTG CTA CCG ACG CTG ATT GCTTTC CAG ATG GCG |
| PRAIx9 | CCC TGA ATT TTG ATA ATA ATG ATA AAT ATA TCT TCG ACA GCC AAC AGGGTG GAA GCG GGA |
| PRAIx10 | ACG TTG CCA AGG ATT TGA CCA TTT AAT AGT GAC CAG TCA AAG CTCTTC CCG CTT CCA CCC |
| PRAIx11 | CAA ATC CTT GGC AAC GTT ATT CTG GCG GGG GGC TTA AAT GCA GAT AACTGC GTG CAG GCG |
| PRAIx12 | ACT CTA CCC CAG AAT TAA AAT CAA GTC CCA GGC AGC CCA GCT TTG CCGCCT GCA CGC AGT |
| PRAIx13 | TAA TTC TGG GGT AGA GTC GAG CCC GGG CAT CAA AGA CGC AAA GAA GATCAA GTC GGT TTT |
| PRAIx14 | GCG GAT CCG TAT GCG CGC AGC TTC TGG AAA ACC GAC TTG ATC TTC |

Reacciones de PCR

Todos los oligos utilizados en las reacciones de PCR que no incluyen el ensamble de la construcción, se encuentran a una concentración de 10 pmol/ μ l. Los dNTPs utilizados en las reacciones de PCR están a 2.5 mM. Los oligos utilizados en el ensamble están a 4 pmol/ μ l

Se utilizaron dos polimerasas: *Taq* Polimerasa y *PWO* una enzima de alta fidelidad (*Biolabs*).

Clonación del gen de PRAI de *E. coli* (control positivo) desde pUC 18 hasta pT4

5 μ l Buffer Taq Pol I
 5 μ l dntps
 0.5 μ l pUC 18
 4 μ l NdePrai
 4 μ l BamPraix
 2 μ l Taq Pol
 29.5 μ l H₂O

Ciclados

1 ciclo (94° 5 minutos)

25 ciclos (94° 40 segundos, 52° 40 segundos, 72° 1 minuto 10 segundos)

1 ciclo (72° 10 minutos)

Las digestiones se realizaron en condiciones estándar (37), utilizando las soluciones amortiguadoras recomendados por el proveedor. El vector pT4 se digirió con las mismas enzimas que el producto de PCR y se ligó en condiciones normales (*NdeI* y *BamHI*). El gen de PRAI de *E. coli* se clonó río arriba del reportero transcripcional CAT, ambas proteínas se unieron mediante una secuencia de 8 aminoácidos o *linker*. La construcción se transformó en células electrocompetentes Δ PRAI para evaluar complementación de auxotrofia en medio mínimo VB. Las bacterias se plaquearon en concentraciones crecientes de cloramfenicol para validar el sistema de CAT como reportero de plegamiento en 10 mg/ml, 15 mg/ml, 25 mg/ml, 35 mg/ml. Se obtuvieron clones resistentes a cloramfenicol (35 mg/ml) y que complementaban auxotrofia en 24 horas.

Purificación de los oligonucleótidos

Antes de ensamblar la construcción consenso fue necesaria la purificación de los oligonucleótidos ya que las reacciones de ensamble son de alta dificultad y cualquier subproducto de síntesis generaría construcciones incompletas o con secuencias distintas a la diseñada. La purificación de oligonucleótidos se realizó mediante PAGE 8 M de Urea con la siguiente metodología:

Preparación de **100 mL** de Solución Stock de Acrilamida al 15% y 7M de Urea

1. Mezclar en un vaso de precipitados de 250mL:
 - 14.25g de Acrilamida
 - 0.75g de Bisacrilamida
 - 42g de Urea
 - 10mL de Buffer TBE10X
 - 46mL de Agua mQ
2. Disolver con agitación magnética
3. Filtrar la solución para eliminar partículas insolubles
1. Preparar un gel de acrilamida mezclando 15mL de la solución stock, 120uL de Persulfato de Amonio al 10% y 15uL de TEMED en una cámara de electroforesis tipo "Tall mighty" con separadores de 1.5mm y peine de 5 pozos del mismo grosor. Cada pozo debe tener un ancho y altura de 13 mm. El proceso de gelificación toma aproximadamente 20 min.
2. Precorrer el gel durante 15min a 300V, apagar la fuente de poder y lavar los pozos con el mismo buffer de corrida (TBE 1X) utilizando una jeringa de 5mL. Cargar 50ug de material (oligo crudo) por fila. NOTA: El volumen de muestra variará de acuerdo con la concentración del oligo, para cargar el gel se necesitan aprox. 50ul de oligo y 10ul de azul de bromofenol por fila.
3. Correr el gel a 350V hasta que el azul de bromofenol empiece a salir por el borde inferior del gel (aprox. 1.5h).
4. Visualizar el gel sobre una placa de silica fluorescente (de las utilizadas en cromatografía en capa fina) con luz UV de onda corta (254nm) y cortar la banda correspondiente a la longitud del oligo (generalmente la banda de mayor concentración y de menor migración).
5. Colocar la banda en un tubo eppendorf de 1.5ml y agregar 1ml de Agua mQ estéril. Incubar a 37 grados durante 12h de preferencia con agitación rotatoria.
6. Retirar el sobrenadante y concentrar la solución hasta 300ul aprox. en un Savant (2hs aprox).
7. Agregar 1.0ml de nButanol y "vortexear" fuertemente durante 30seg. Centrifugar a 13000rpm durante 1min. Eliminar la fase superior y repetir una vez mas la extracción con nButanol. En este punto solo debe quedar un pellet transparente.
8. Agregar 1ml de Etanol absoluto, agitar en Vortex y centrifugar (3min). Eliminar el sobrenadante por decantación.

9. Secar el pellet en Savant.
10. Disolver el pellet en 115ul de agua mQ estéril y medir absorbancia a 260nm. (Tomar como referencia el agua mQ que se usó para disolver. La dilución para medir absorbancia es de 5ul de oligo +495 de agua). El valor leído multiplicarlo por 3.3 y se obtiene la concentración en ug/ul.

Los oligonucleótidos obtenidos se analizaron en gel de archilamida 12% 8 M de Urea observándose una sola banda por UV.

Ensamble del gen.

La construcción se ensambló en tres partes mediante reacciones de PCR

PCR Parte 1

20µl Buffer
20µl dntps
10 µl oligo Praix 1
10µl oligo Praix 8
1µl oligos Praix 2 al 7
1µl PWO
133µl H₂O
200µl final. Dividir en 4 tubos

PCR Parte 2

20µl Buffer
20µl dNTPs
10µl oligo Praix 7
10µl oligo Praix 14
1µl oligos Praix 8 al 13
1µl PWO
133 µl de H₂O
200 µl final. Dividir en 4 tubos

PCR de ensamblaje final

10µl Buffer
 10µl dNTPs
 3µl Parte 1
 1µl Parte 2
 2µl NdePrai
 5µl Praix14
 .5µl PWO
 68.5 µl de H₂O
 100 µl final dividir en 2 tubos

Ciclado de las dos mitades

1 ciclo (94° 2 minutos)
 25 ciclos (94° 30 segundos, 57° 30 segundos, 72° 1 minuto 10 segundos)
 1 ciclo (72° 10 minutos)

Ciclado del ensamblaje final

1 ciclo (94° 5 minutos)
 25 ciclos (94° 40 segundos, 57° 40 segundos, 72° 1 minuto 10 segundos)
 1 ciclo (72° 10 minutos)

El producto se PCR final se analizó en un gel de agarosa al 1.5% (*Biorad* grado DNA) observándose una sola banda de 600pb .El producto de PCR se clonó en el vector pT4 de la misma manera que PRAI de *E. coli* entre los sitios *NdeI* y *BamHI* (37) quedando fusionado a CAT, para evaluar la viabilidad del plegamiento de la secuencia nueva. Se evaluaron niveles de resistencia a cloramfenicol en placas de medio LB. Posteriormente se secuenciaron 20 construcciones resistentes a cloramfenicol (35 mg/ml) en la unidad de secuencia del IBT. Se obtuvieron 4 clonas que contenían la secuencia diseñada como consenso. Se evaluó complementación de auxotrofia en cepa ΔPRAI de *E. coli* en medio sólido VB mediante diluciones decrecientes y goteo. Se pudo constatar complementación en 48 horas en placas de Cm Km (35mg/ml) incubando a 30°.

Se picaron (37) las siguientes células:

Colonias de la cepa ΔPRAI transformadas con la construcción consenso fusionada a CAT
 Colonias de la cepa ΔPRAI transformadas con la construcción consenso sin fusión.
 Colonias de la cepa ΔPRAI transformadas con PRAI de *E. coli* fusionada.
 Colonias de la cepa ΔPRAI transformadas con PRAI de *E. coli* sin fusionar
 Colonias a partir de células electrocompetentes almacenadas en glicerol de la cepa ΔPRAI (control negativo).
 Colonias a partir de células electrocompetentes almacenadas en glicerol de la cepa JM101 (control positivo)

Se inocularon tubos de medio mínimo VB estandarizados a .033 O.D. de todas las colonias picadas y se siguió una curva de crecimiento tomando muestras cada hora, durante trece horas consecutivas hasta que la curva del control positivo mostró un comportamiento lineal.

No existieron diferencias significativas en las curvas de crecimiento de cada una de las colonias, únicamente la cepa utilizada como control negativo no aumento en su densidad óptica. Posteriormente se procedió a separar la proteína reportero mediante digestión en sitios *BamHI* y *BglIII* (37), incorporando una marca de 6xHis en el extremo amino terminal de la proteína consenso y la proteína PRAI de *E. coli*.

3. Bioquímica

Electroforesis en gel de acrilamida desnaturalizante (SDS-PAGE)

Durante el análisis de la expresión y purificación tanto de la proteína silvestre como de la consenso se utilizaron geles de acrilamida (*Biorad*) al 15%.

Bradford (*Biorad*). Concentración de proteína.

Bradford Stock

100ml 95% etanol

200ml 88% ácido fosfórico

350 mg Serva Blue G

Para la curva patrón (595nm)

| Proteína | Solución estándar 1 mg/ml de BSA | Buffer experimental (µl) | Bradford (µl) |
|----------|----------------------------------|--------------------------|---------------|
| 0 | 0 | 800 | 200 |
| 2.5 | 2.5 | 797.5 | 200 |
| 5 | 5 | 795 | 200 |
| 10 | 10 | 790 | 200 |
| 15 | 15 | 785 | 200 |
| 20 | 20 | 780 | 200 |

Tabla 1. Preparación de las muestras para obtener la curva patrón en el ensayo de Bradford (*Biorad*).

Expresión y purificación de la proteína consenso y de PRAI de *E. coli*

El medio de cultivo utilizado para la expresión fue LB

El inductor utilizado fue IPTG (*Roche*)

Las células utilizadas fueron *E. coli* JM101 transformadas con los vectores correspondientes.

Inhibidor de proteasas PMSF de (*Sigma*)

Soluciones de purificación

1. NaH_2PO_4 50 mM
NaCl 600 mM
Imidazol 10 mM
Glicerol 10 %
2. NaH_2PO_4 50 mM
NaCl 600 mM
Imidazol 20 mM
Glicerol 20%
3. NaH_2PO_4 50 mM
NaCl 600 mM
Imidazol 40 mM
4. NaH_2PO_4 50 mM
NaCl 600 mM
Imidazol 50 mM
5. NaH_2PO_4 50 mM
NaCl 600 mM
Imidazol 100 mM
6. NaH_2PO_4 50 mM
NaCl 600 mM
Imidazol 200 mM
7. NaH_2PO_4 50 mM
NaCl 600 mM
Imidazol 300 mM

Método de purificación columna Ni-NTA

Montar un cultivo O/N en condiciones estandarizadas de 1/10 de la cantidad de cultivo total que se pretenda purificar. Inocular un litro de cultivo. Monitorear el cultivo hasta alcanzar 0.4 O.D. a 30° posteriormente inducir con 1 mM de IPTG. Detener el cultivo cuando alcance una concentración de 1 O.D. Centrifugar en refrigeración 20 minutos a 4,000 rpm. Agregar 5 ml de solución amortiguadora 1 y guardar O/N a - 70° . Posteriormente descongelar y sonicar (pulsos de 90 a una fuerza de 4 y siempre en frío) agregar inhibidor de proteasas. Centrifugar 10 minutos a 9,000 rpm. Posterior a estos pasos de expresión y lisado se obtiene la fracción soluble de las células lista para ser pasada por la columna de purificación.

Empacar 3 ml de resina Ni-NTA y equilibrarla con solución amortiguadora 1 (10 volúmenes de columna), agregar el lisado clarificado. Pasar por la columna lentamente. Lavar con 10 volúmenes de solución 2. Eluir la proteína con tres volúmenes de columna en 4 fracciones con las soluciones 3,4, 5 y 6. Posteriormente lavar la columna con 10 volúmenes de solución 7. La columna tiene que lavarse finalmente con 10 volúmenes de ácido acético 200 mM, glicerol 30 % y H₂O en ese orden. Se almacena en etanol 30%.

Las fracciones colectadas se analizan en un SDS PAGE para evaluar la eficiencia de la separación. La purificación tanto de la proteína silvestre como de la consenso se realizó mediante el Kit de *Qiagen* The QIAexpressionist. Este kit es especial para la purificación de proteínas marcadas con cola de afinidad de 6xhis.

Purificación mediante permeación en gel

Posterior a la purificación mediante marca de histidinas fue necesario un segundo método para alcanzar la pureza necesaria para el experimento de desnaturalización térmica. Se montó una cromatografía de permeación en gel (Sephacryl S 100 HR). Utilizando una columna: HiPrep 26/60 (Amersham Biosciences) con un eluente tampón Tris 25 mM, pH 7.5, NaCl 150 mM. Este gel está recomendado para la separación de proteínas de 1 a 100 kDa. PRAI consenso tiene un MW de 22kDa. La proteína se recupero en 6 fracciones terminales.

Experimentos de desnaturalización térmica

Los experimentos de desnaturalización térmica se realizaron en un espectrofotómetro Olis Cary 14 de Online Instrument Systems, acoplado a un baño con temperatura controlada y fue operado por el paquete Oliswork. Las muestras de proteína pura se dializaron utilizando membranas Spectra/Por de spectrum de MWCO: 12 – 14000 por cuatuplicado en frío, utilizando 1 litro de la siguiente solución amortiguadora filtrada y degasificada por cada diálisis:

Fosfatos 20 mM

NaCl 100 mM

pH 7.5

β-mercaptoetanol a 70 μl por litro de solución y únicamente en la primera solución de diálisis

La concentración de proteína de determino mediante Bradford (1976) de *Biorad*. Ambas proteínas se estandarizaron a una concentración de 0.1 mg/ml. Los experimentos de desnaturalización térmica tanto para la proteína consenso como para PRAI de *E. coli* se realizaron en cubetas de cuarzo gemelas conteniendo 1 ml de la proteína como el problema y 1 ml del buffer de diálisis como referencia. Los experimentos de desnaturalización térmica se siguieron aumentando la temperatura gradualmente, en intervalos de 0.5 grados centigrados cada 5 minutos (tiempo de estabilización de la muestra) hasta alcanzar una desnaturalización completa de la proteína problema. Los datos obtenidos se analizaron utilizando la ecuación de Van't Hoff mediante el método de determinación experimental de temperatura promedio de desnaturalización (41).

VII. RESULTADOS

1. Diseño de la proteína consenso

Se realizó una búsqueda en la base de datos *Swiss-Prot* de *ExpASY Proteomics Server* utilizando como nombre clave PRAI. Se obtuvieron 127 secuencias en *Swiss-Prot*, correspondiendo a las proteínas que ya se encuentran caracterizadas bioquímicamente. Mientras que en el *TrEMBL* se encontraron anotadas 123 secuencias de proteínas, que eran *ORFs (Open Reading Frames)* contenidos en el *EMBL* y traducidos como posibles proteínas.

Utilizamos únicamente las proteínas que se encontraban en el servidor *Swiss-Prot* debido a que esta base de datos se considera curada y contiene proteínas que están caracterizadas molecular o bioquímicamente. Decidimos no utilizar proteínas cuya actividad no estaba confirmada bioquímicamente, como sucede con el *TrEMBL* en donde existen anotaciones de proteínas que se han definido como tales únicamente utilizando *BLAST* en varios marcos de lectura.

Después de seleccionar las secuencias que serían objeto de este análisis, se realizó un primer alineamiento múltiple utilizando las 126 secuencias de PRAI en *Clustal X* con los parámetros *default* del programa.

El primer alineamiento múltiple que se realizó con las 126 secuencias de PRAI mostró que existían por lo menos dos dominios distintos entre las proteínas alineadas. Se sabe que PRAI se encuentra fusionada con la siguiente enzima en la vía de síntesis de triptofano IGPS en varios organismos, entre ellos las Arqueas y *E. coli*.

Nuestro modelo de trabajo consistía en la PRAI monofuncional de *E. coli* (31) que consta de un solo dominio de barril TIM, fue separado del otro dominio presente en la proteína bifuncional de *E. coli* PRAI-IGPS. También se sabe que la proteína PRAI se encuentra formando complejos multiproteicos en varios organismos.

Era necesario alinear únicamente el dominio de PRAI en todas las secuencias encontradas. Para ello, se utilizó la base de datos *HSSP* (46), donde se pueden encontrar alineamientos estructurales de proteínas homólogas hechos con base en alguna proteína de la que se tenga resuelta la estructura. Se utilizó la estructura que se tiene resuelta de PRAI de *E. coli* como semilla para obtener el alineamiento estructural, debido a que serviría a lo largo de todo el experimento como parámetro comparativo (proteína silvestre).

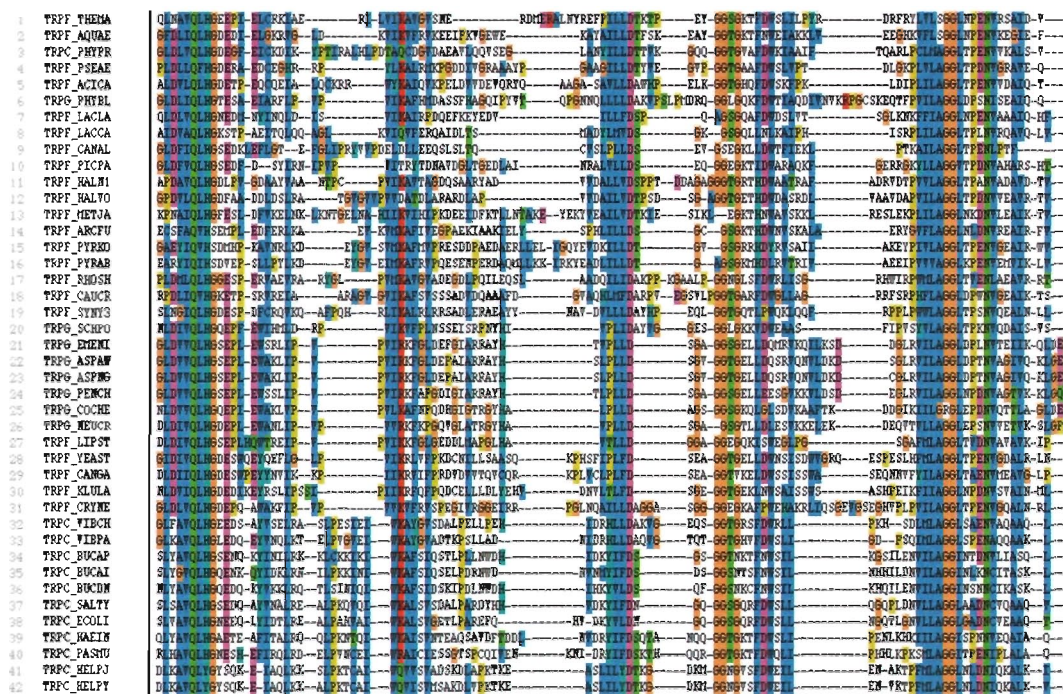


Figura 5. Alineamiento estructural tomado del HSSP (46) de 46 secuencias de PRAI.

El alineamiento múltiple se cargó en el programa *Clustal X* generándose un árbol de distancia UPGMA (figura 7). Este tipo de árboles únicamente reflejan la similitud general entre las secuencias. Después de generar un árbol con las 46 secuencias de PRAI basados en el alineamiento estructural obtenido del HSSP (46), pudimos observar que era posible subdividir las secuencias en tres grupos en función del % de identidad.

Como se aprecia en el árbol, existen tres grupos claramente diferenciados en tres ramas del árbol. Subdividiendo el árbol en tres grupos principales iniciamos el análisis de las secuencias para definir el consenso. Es claro que entre más secuencias conformaran el consenso más difícil sería tener un consenso bien resuelto. Si el grupo de secuencias era demasiado amplio era posible que el consenso resultara ambiguo.

También era necesario definir el porcentaje de identidad que implicara tener un consenso simple de mayoría. Como el consenso se definió con el programa *Jalview* era necesario especificar con qué nivel de identidad se diseñaría el consenso. Generamos varios consensos con tres grupos distintos a varios porcentajes de identidad. Entre más estricto fuera el porcentaje de identidad más difícil era definir una secuencia consenso.

Posteriormente se indican las secuencias contenidas en el alineamiento estructural en el HSSP utilizando como semilla la estructura tridimensional de PRAI de *E. coli*. (PDB Id. 1PII). Se muestra el identificador de *Swiss-prot* de la proteína PRAI junto al nombre del organismo:

1. TRPF_THEMEA : *Thermotoga Maritima* (Bacteria)
2. TRPF_AQUAE: *Aquifex aeolicus* (Bacteria)
3. TRPC_PHYPR: *Phytophthora parasitica* (Hongo)
4. TRPF_PSEAE: *Pseudomonas aeruginosa* (Proteobacteria)
5. TRPF_ACICA: *Acinetobacter calcoaceticus* (Proteobacteria)
6. TRPG_PHYBL: *Phycomyces blakesleeanus* (Hongo)
7. TRPF_LACLA: *Lactococcus lactis* (Bacteria Firmicutes)
8. TRPF_LACCA: *Lactobacillus casei* (Bacteria Firmicutes)
9. TRPF_CANAL: *Candida albicans* (Hongo)
10. TRPF_PICPA: *Pichia pastoris* (Hongo)
11. TRPF_HALN1: *Halobacterium sp* (Archaea)
12. TRPF_HALVO: *Halobacterium volcanii* (Archaea Euryarchaeota)
13. TRPF_METJA: *Methanococcus jannaschii* (Archaea Euryarchaeota)
14. TRPF_ARCF: *Archaeoglobus fulgidus* (Archaea Euryarchaeota)
15. TRPF_PYRKO: *Pyrococcus kodakaraensis* (Euryarchaeota)
16. TRPF_PYRAB: *Pyrococcus abyssi* (Euryarchaeota)
17. TRPF_RHOSH: *Rhodobacter sphaeroides* (Proteobacteria)
18. TRPF_CAUCR: *Caulobacter crescentus* (Proteobacteria)
19. TRPF_SYNY3: *Synechocystis sp* (Cyanobacteria)
20. TRPG_SCHPO: *Schizosaccharomyces pombe* (Fission yeast Hongo)
21. TRPG_EMENI: *Emericella nidulans* (*Aspergillus nidulans*) (Hongo)
22. TRPG_ASPAW: *Aspergillus awamori* (Hongo)
23. TRPG_ASPNG: *Aspergillus niger* (Hongo)
24. TRPG_PENCH: *Penicillium chrysogenum* (Hongo) Ascomycota
25. TRPG_COCHE: *Cochliobolus heterostrophus* (*Drechslera maydis*) (Hongo)
26. TRPG_NEUCR: *Neurospora crassa* (Hongo)
27. TRPF_LIPST: *Lipomyces starkeyi* (Hongo)
28. TRPF_YEAST: *Saccharomyces cerevisiae* (Hongo)
29. TRPF_CANGA: *Candida glabrata* (Yeast)
30. TRPF_KLULA: *Kluyveromyces lactis* (Yeast)
31. TRPF_CRYNE: *Cryptococcus neoformans* (*Filobasidiella neoformans*) (Hongo)
32. TRPC_VIBCH: *Vibrio cholerae* (Gamma proteobacteria)
33. TRPC_VIBPA: *Vibrio parahaemolyticus* (gamma proteobacteria)
34. TRPC_BUCAP: *Buchnera aphidicola* (subsp. *Schizaphis graminum*) Gamma proteobacteria
35. TRPC_BUCAI: *Buchnera aphidicola* (subsp. *Acyrtosiphon pisum*) Gamma proteobacteria
36. TRPC_BUCDN: *Buchnera aphidicola* (subsp. *Diuraphis noxia*) Gamma proteobacteria
37. TRPC_SALTY: *Salmonella typhimurium* Gamma proteobacteria
38. TRPC_ECOLI: *Escherichia coli* Gammaproteobacteria
39. TRPC_HAEIN: *Haemophilus influenzae* Gammaproteobacteria
40. TRPC_PASMU: *Pasteurella multocida* Gammaproteobacteria
41. TRPC_HELPJ: *Helicobacter pylori* J99 Epsilon proteobacteria
42. TRPC_HELPY: *Helicobacter pylori* Epsilon proteobacteria
43. TRPC_BUCSC: *Buchnera aphidicola* (subsp. *Schlechtendalia chinensis*) Gammaproteobacteria
44. TRPF_METTH: *Methanobacterium thermoautotrophicum* (Archaea Euryarchaeota)
45. TRPF_CHLTR: *Chlamydia trachomatis* (Bacteria Chlamydiae)
46. TRPF_CHLMU: *Chlamydia muridarum* (Bacteria Chlamydiae)

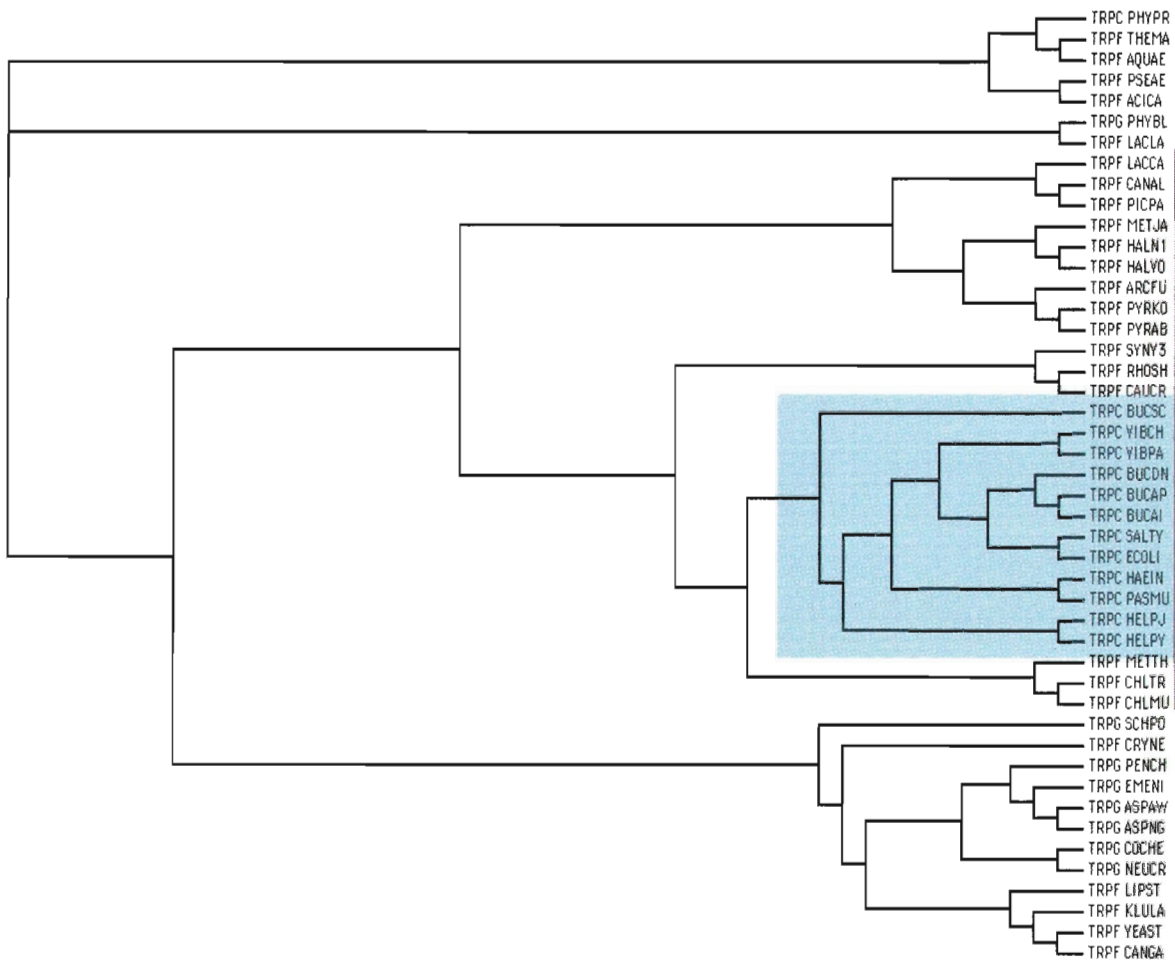


Figura 7. Árbol de 46 secuencias de PRAI generado con UPGMA. Las secuencias se identifican con a clave de *Swiss-prot*. Los tres grupos distintos se indican de la siguiente manera: a) grupo 1 en azul, b) grupo 2 barra gris y c) grupo 3 barra negra.

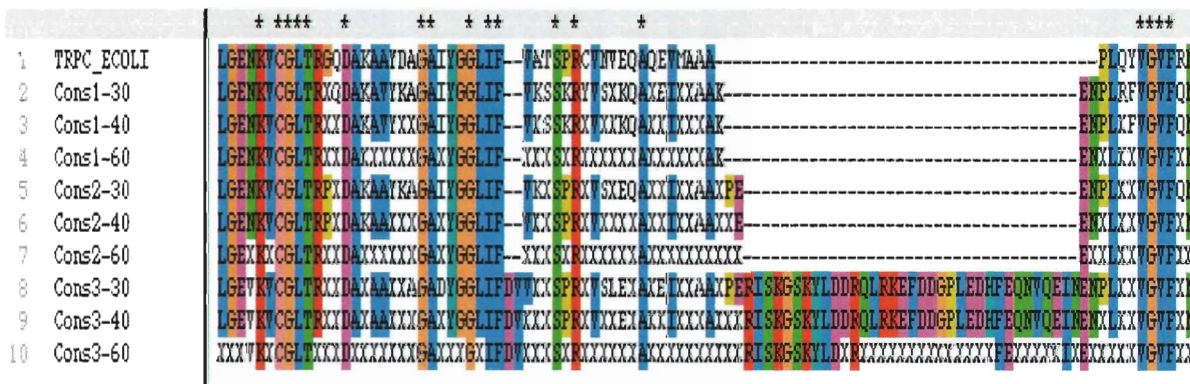


Figura 8. Diferentes consensos obtenidos a diferentes porcentajes de identidad. Se muestran también los consensos generados con tres diferentes grupos de secuencias.

El grupo 3 englobaba 46 secuencias, mientras que el grupo 2 incluía 24 y el último grupo contenía 12. La selección de secuencias se realizó con base en un estudio exhaustivo de las 46 iniciales. El grupo 3 incluía secuencias de arqueobacterias; generalmente estas bacterias adoptan estrategias de oligomerización para obtener una mayor estabilidad en sus proteínas (32).

TESIS CON FALLA DE ORIGEN

Como se decidió construir el consenso alrededor de la proteína de *E. coli*, incluir estas secuencias en el consenso resultaba difícil ya que generaban demasiado ruido. El alineamiento múltiple contenía demasiados y muy largos *gaps*, debido a las inserciones tan grandes que introducían las secuencias de las arqueas y de eucariontes. Al rasurar únicamente las regiones de las inserciones en el alineamiento, podríamos contar con la información de más secuencias; sin embargo, era difícil definir hasta dónde se debían rasurar las inserciones sin afectar la proteína, además de que la inclusión de un número muy grande de secuencias generaba un consenso demasiado ambiguo (figura 8).

El problema principal a resolver era seleccionar suficientes secuencias para tener un consenso consistente. Sin embargo no era posible utilizar demasiadas, debido a que la dificultad para definir un aminoácido consenso en cada sitio de la proteína, crecía conforme el número de secuencias incluidas en el consenso crecía también (figura 9). Era necesario eliminar secuencias pero sin perder información. También existía el problema de escoger el porcentaje de identidad del consenso. El objetivo era generar un consenso de mayoría simple. Esto quiere decir, que el aminoácido consenso se define como el residuo que aparece más veces que todos los demás en el conjunto seleccionado de secuencias. Se utilizaron tres cortes en tres niveles: 30, 40 y 60 % de identidad.

Los aminoácidos consenso se definían en menos sitios de la proteína entre más estricto resultaba el consenso. No era posible tener uno muy estricto, pero tampoco demasiado laxo. Haciendo una revisión bibliográfica, encontramos que en el trabajo de Alan Fersht en el que se generaron minichaperonas más estables, el porcentaje de identidad que mejor proporción de mutaciones estabilizantes / desestabilizantes presentaba, era en el corte en un 30 % de identidad. Después de varios análisis con los distintos consensos y al ver que no era fácil revolver la gran mayoría de aminoácidos consenso en los cortes de 40 y 60 % de identidad, se decidió utilizar el corte en un 30 % de identidad.

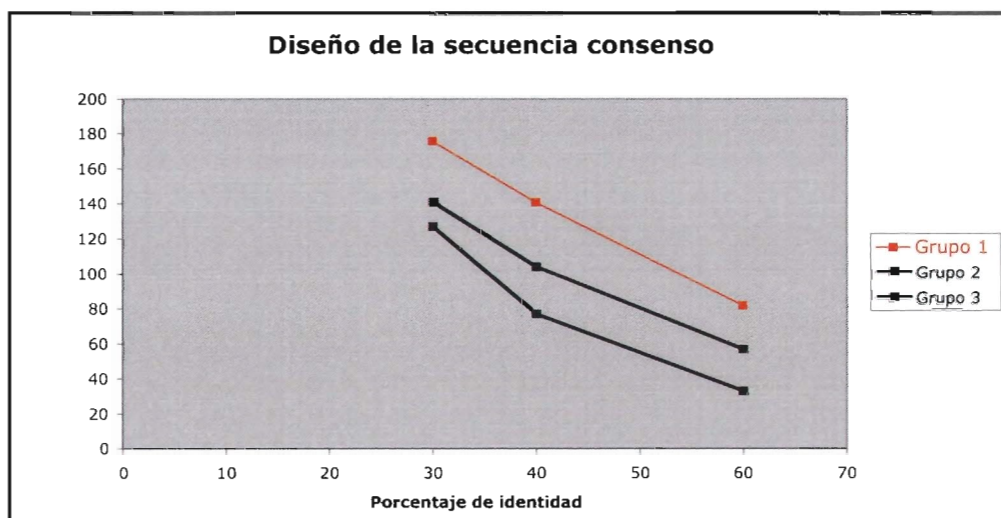


Figura 9. Elección del grupo de secuencias que integra el consenso. Se muestra el número de residuos consenso que se resuelven en cada grupo, en función del % de identidad. En rojo se indica el grupo de secuencias elegido para integrar la nueva secuencia.

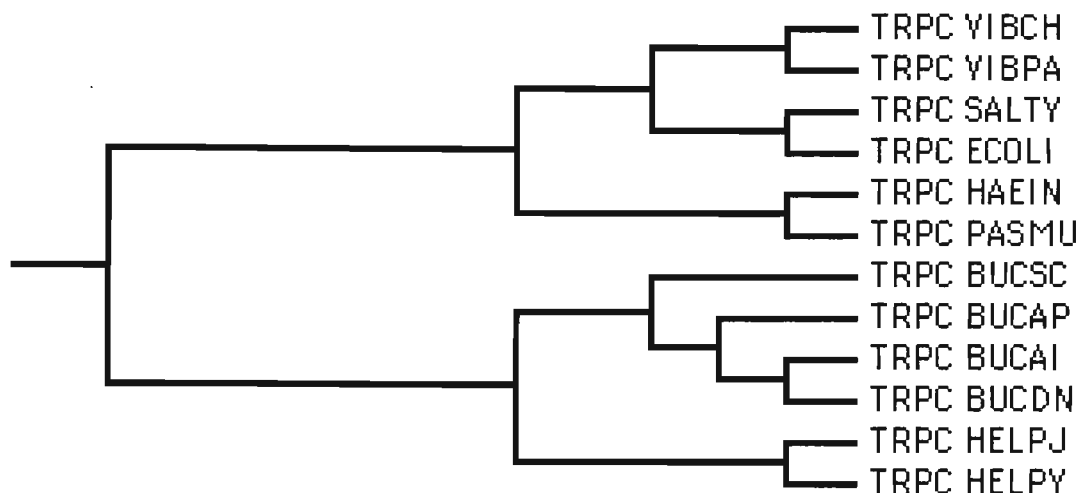


Figura 10. Se muestran las secuencias finales para la generación del consenso etiquetadas con los identificadores de *Swiss Prot*.

El consenso final se definió en un corte al 30 % de identidad y con el grupo 1 de secuencias. En el momento de obtener la secuencia consenso se utilizaba el aminoácido de *E. coli*, cuando no era posible definir un residuo consenso en la secuencia final. Esta decisión se debe a varios factores: a) en primer lugar por simplicidad, debido a que utilizaríamos este organismo como sistema de expresión de proteínas heterólogas, b) la selección final de secuencias estaba determinada por la cercanía de las proteínas a la enzima de *E. coli* en el árbol generado mediante UPGMA y c) que PRAI de *E. coli* sería utilizada como control positivo. El porcentaje de identidad entre la PRAI de *E. coli* y la nueva proteína consenso es de 75 %. Cambiaron 49 aminoácidos con respecto al gen de *E. coli*.

2. Retrotraducción del gen

A partir de la secuencia de aminoácidos de la nueva proteína consenso, se procedió a retrotraducir la secuencia de DNA. Se ha demostrado que el diseño de la secuencia de nucleótidos en la construcción de un gen sintético es un paso crucial para asegurar la expresión correcta del polipéptido (35). La retrotraducción de la secuencia de aminoácidos se realizó tomando en cuenta el uso de codones de *E. coli*, ya que este organismo fue utilizado como sistema de expresión. Generalmente las estrategias de optimización involucran elegir el codon de mayor frecuencia de utilización para diseñar la secuencia de nucleótidos; sin embargo, en este experimento se mimetizó la frecuencia de utilización de los codones de *E. coli* que presentaba el gen de PRAI de este organismo.


```

GCU A 0.16 15.4 ( 24591)
GCC A 0.27 25.5 ( 40827)
GCA A 0.21 20.3 ( 32501)
GCG A 0.35 33.6 ( 53883)

```

Figura 11. Frecuencia relativa en el uso de codones para el aminoácido Alanina en *Escherichia coli* (40).

Para describir de una manera más clara la metodología utilizada, se mostrará cómo se definió el codon utilizado en la sustitución de la Valina 17 por Alanina (figuras 11 y 12). El codon que utilizaba *E. coli* en la proteína es GCU con una frecuencia de 0.16; éste fue sustituido por un codon codificando para Valina GUA con una frecuencia de uso de 0.15. Al mimetizar el uso de codones de la proteína de *E. coli* en la nueva proteína consenso, se trataba de garantizar la expresión correcta de la proteína consenso utilizando *E. coli* como sistema de expresión.

```

GUU V 0.26 18.4 ( 29453)
GUC V 0.22 15.2 ( 24380)
GUA V 0.15 10.9 ( 17424)
GUG V 0.37 26.2 ( 42031)

```

Figura. 12 Frecuencia relativa del uso de codones para el aminoácido Valina en *Escherichia coli* (40).

Existen numerosos estudios que demuestran que tratar de expresar genes heterólogos en bacterias utilizando codones raros para el organismo en cuestión resulta en una pobre expresión del mismo.

Tampoco resulta conveniente utilizar los codones de máxima frecuencia debido al posible secuestro de los aminoacil tRNAs cargados óptimos inhibiendo el crecimiento normal de la bacteria. Es posible que se puedan perder posibles pausas en el plegamiento cotraduccional, con la utilización única de codones óptimos. Esto a su vez reduciría la fracción de proteína soluble.

Las curvas del uso de codones para PRAI de *E. coli* y PRAI consenso se sobrelapan bastante bien, así como lo valores de CAI de ambos genes resultaron cercanos. Al observar que la desviación en el uso de codones de ambos genes es baja podemos pensar que no existirán problemas de expresión en este sentido.

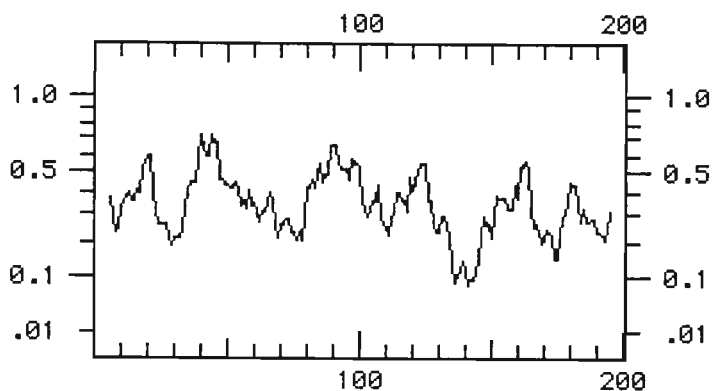


Figura 15. Curva del *Codon Adaptation Index* de PRAI de *E. coli*.

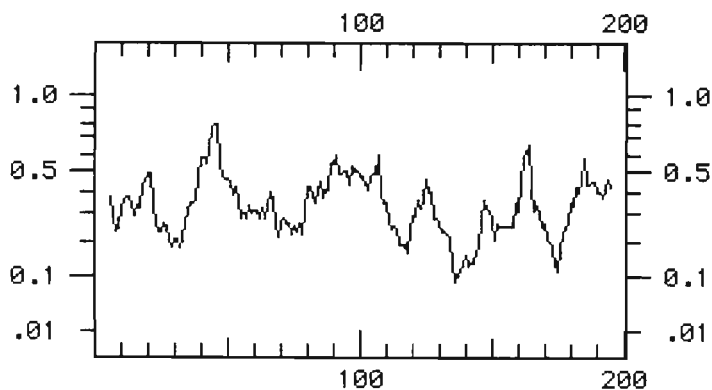


Figura 16. Curva del *Codon Adaptation Index* de PRAI Consenso.

Posteriormente se revisó la secuencia para constatar que no presentara regiones repetitivas y que no se hubieran generado estructuras secundarias en el mRNA que resultaran desfavorables para la correcta expresión del gen. Se introdujeron sitios de restricción en el gen nuevo para clonación en el vector de expresión y se constató que no existieran sitios de corte indebidos en la secuencia, para evitar que el gen se digiriera de manera incorrecta.

PRAI consenso comparte un 75% de identidad en la secuencia de aminoácidos con PRAI de *E. coli*. Mientras que a nivel de nucleótidos comparten un 80% de identidad en las secuencias. Los cambios que presenta la nueva proteína con respecto a la proteína de *E. coli* se distribuyen a lo largo de todo el polipéptido; en consecuencia, se decidió sintetizar el gen completo en vez de cambiar los aminoácidos nuevos mediante oligonucleótidos mutagénicos sintéticos sobre el gen de PRAI de *Escherichia coli*.

3. Diseño de los oligonucleótidos y síntesis del nuevo gen.

El diseño de los oligonucleótidos se realizó con el programa PRIMO de Chang Bioscience. En este programa se especifica el diseño de oligonucleótidos para síntesis de un gen completo en un hospedero en específico. También se seleccionan parámetros de temperatura de hibridación (T_m), el número de bases que aparea y la longitud del oligonucleótido. El programa está optimizado para que los oligonucleótidos no formen estructuras secundarias que impidan el ensamblado correcto de la construcción.

Después de haber alimentado el programa con la secuencia del gen sintético se obtuvieron 13 oligonucleótidos de 60 nt y uno de 45 nt en el que se introdujo un sitio de restricción para BamHI. Para la posterior clonación del gene consenso se introdujo un sitio NdeI en el oligo NdePRAI. Todos los oligonucleótidos presentaron una T_m de entre 48 y 50° C. Al utilizar una T_m similar para todos los oligonucleótidos, se pueden establecer condiciones similares en las reacciones de PCR para ensamblar las partes del gen. El método de ensamblaje de oligos para la síntesis de genes consiste en utilizar los oligonucleótidos de manera alternada. El oligonucleótido número uno aparea en 10 pb con el oligonucleótido número dos, en la región 3'. Los oligos corren en dirección antiparalela. Posteriormente al apareamiento de las bases complementarias en ambos oligonucleótidos, la polimerasa sintetiza la cadena que hace falta en las regiones en donde no hibridaron los oligos, terminando con un segmento del gen que también aparea en 10 pares de bases con el siguiente y así sucesivamente.

La síntesis del gen no es una reacción de PCR propiamente dicha, el gen no se amplifica únicamente se completan los segmentos de cadena sencilla que no sobrelapan en los oligonucleótidos alternados. La síntesis del gen se planeó sintetizando dos mitades por separado y posteriormente uniéndolas en un PCR final en el cual ambas mitades se unen por complementariedad en muchos pares de bases. Para obtener el producto final amplificado se utilizaron concentraciones elevadas de los oligonucleótidos externos, asegurando obtener únicamente la construcción de interés.

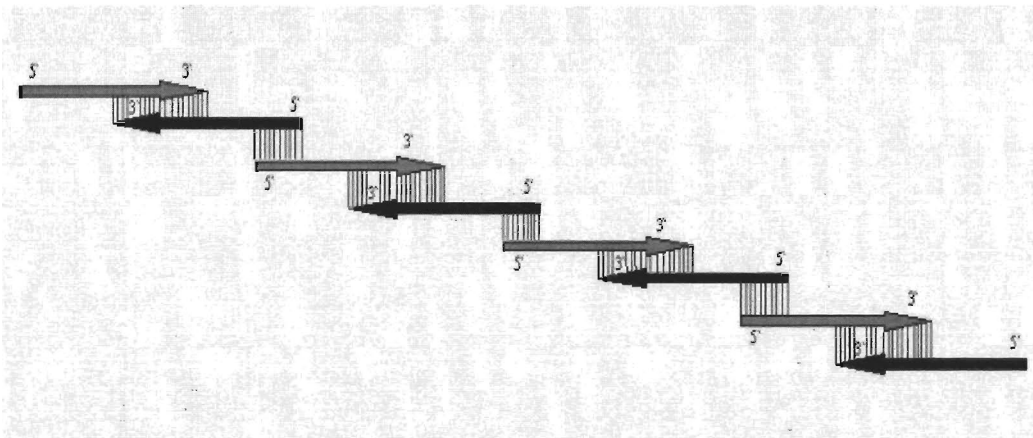


Figura 17. Ensamblaje de primera parte del gen. Construcción formada por los oligos 1-8. Los oligos sobrelapan en 10 pares de bases. La segunda parte del gen se ensambló de la misma manera pero con los oligos 8-14.

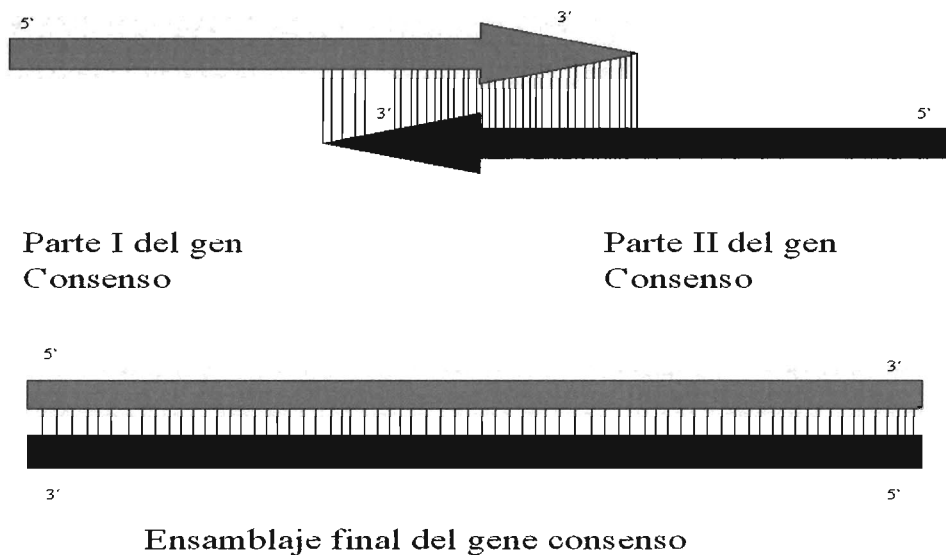


Figura 18. Unión de las dos mitades del gen consenso y ensamblaje final de la construcción. Para obtener la construcción completa se utilizaron concentraciones altas de oligos externos que pegan en el gen y equivalentes más bajas para las dos mitades.

El gen completo tiene 600 pares de bases, después de la reacción final de PCR se obtuvo una banda por electroforesis en gel de agarosa 1.5% que correspondía al tamaño final del gen. Esta banda se clonó en un vector pT4 fusionado a un reportero de plegamiento y con un marcador de resistencia a kanamicina.

4. Clonación del gen consenso.

El Vector pT4 contiene un reportero de plegamiento llamado CAT (cloramfenicol acetil transferasa). Esta enzima confiere a la bacteria resistencia al cloramfenicol si y sólo si la proteína se encuentra plegada correctamente. El nuevo gen se clonó en una posición río arriba en relación con el reportero unido con un *linker* de 8 residuos. Si el gen nuevo presentaba un plegamiento correcto, el reportero lo hacía también, entonces es posible recuperar proteínas plegadas correctamente. La proteína PRAI de *E. coli* se clonó en un vector pT4 como control positivo y se le dio el mismo tratamiento que al gen nuevo.

Posteriormente a la clonación en el vector pT4 se procedió a transformar la construcción en células electrocompetentes XL1Blue. Las clonas transformadas se seleccionaron en concentraciones crecientes de cloramfenicol (CM) 10, 15, 20 y 35 mg/ml.

Para estimar las colonias que contenían el inserto se practicó un PCR de colonia con los oligonucleótidos Praix9 y Praix14 (que son los únicos oligos sintetizados específicos para la construcción consenso en este trabajo). También se mantuvo un control de autoligación mostrando una proporción de 3:1 de clonas ligadas / autoligadas. Se recuperó plásmido de 20 clonas y se mandó a secuenciar a la unidad de síntesis del IBT. Se obtuvieron cuatro clonas con la secuencia diseñada originalmente.

Las clonas seleccionadas se sometieron a digestión con enzimas de restricción para remover el reportero de plegamiento. Con las enzimas *BamHI* y *BglII* (37) se liberó el gen de CAT en el pT4 y se marcó con una bandera de Hisx6 para su posterior purificación por medio de una columna de níquel (Quiagen). Cinco clonas seleccionadas de la autoligación del vector se mandaron a secuenciar recuperándolas sin ninguna mutación o deleción. El gen de PRAI de *E. coli* también se liberó del gen reportero de la misma manera que el gen consenso. Todas las clonas que se obtuvieron al final fueron secuenciadas para asegurar que tanto la proteína control como la proteína consenso se expresaran a partir de las construcciones diseñadas.

5. Complementación de auxotrofia en medio sólido VB.

Posterior a la obtención de la construcción completa del gen consenso se comprobó que se mantenía la actividad de PRAI al complementar auxotrofia en la cepa Δ PRAI de *E. coli*.

La proteína PRAI de *E. coli* clonada en el vector pT4, se transformó en la cepa Δ PRAI para comprobar que la complementación funcionaba en el sistema diseñado. Posteriormente se transformó la nueva proteína consenso para evaluar si presentaba actividad suficiente como para permitir el crecimiento de la cepa auxótrofa. La complementación en medio sólido VB (Boggel Vonner) de auxotrofia de la cepa Δ PRAI por el gen Silvestre (PRAI de *E. Coli*) se observó en 24 horas incubando las bacterias a 30° C. La incubación de las bacterias con la construcción consenso se observó en 48 horas a niveles equiparables de los obtenidos con la construcción silvestre y a la misma temperatura.

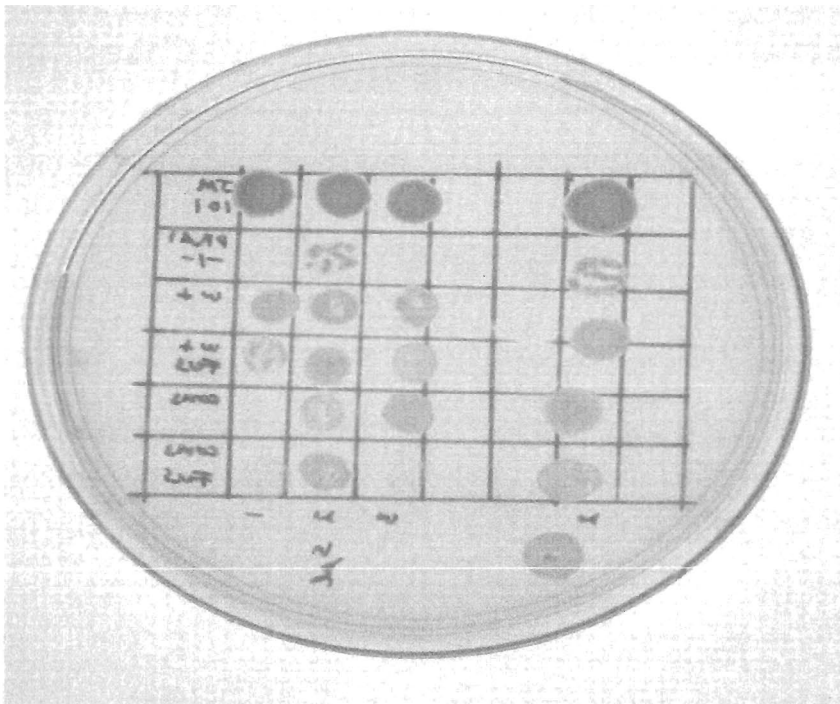


Figura 19. Complementación de auxotrofia con las construcciones silvestre y consenso.

Fila 1. JM101 (Control positivo).

Fila 2. Cepas Δ PRAI (Control negativo)

Fila 3. Complementación con PRAI de *E. coli*

Fila 4. Complementación con PRAI de *E. coli* fusionada a CAT

Fila 5. Complementación con PRAI Consenso

Fila 6. Complementación con PRAI Consenso fusionada a CAT

Primeras 3 columnas: gotas de 5 μ l. Última columna: 1 gota de 10 μ l

No hemos logrado atribuir a algún factor la diferencia en tiempos de complementación entre ambas proteínas, ya que la complementación en medio líquido VB no mostró diferencias significativas en ninguna de las proteínas que fueron evaluadas en la misma cepa.

6. Complementación de auxotrofia en medio líquido VB.

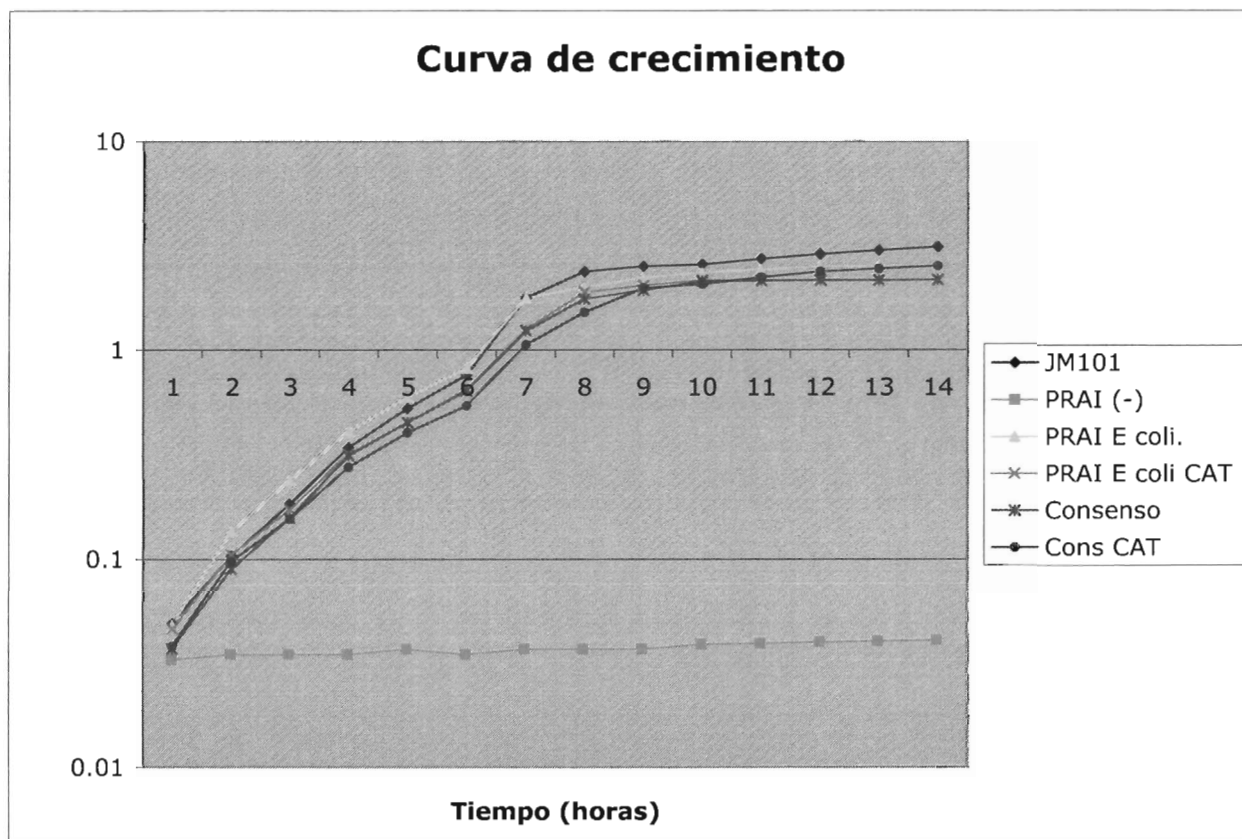


Figura. 20 Curva de complementación. Densidad óptica en función del tiempo. El experimento se realizó con dos repeticiones. Los resultados mostrados aquí son el promedio de ambas repeticiones.

Como podemos apreciar en las curvas de complementación, no existen diferencias significativas en los niveles de densidad óptica en las clonas examinadas. Se permitió el crecimiento del cultivo hasta que éste entró en fase estacionaria. El control negativo demuestra que el medio VB está funcionando eficientemente como método de selección.

7. Purificación de la proteína consenso.

Posterior a la comprobación de la actividad de la proteína consenso, se procedió a purificarla para medir sus parámetros de termoestabilidad. La expresión de la proteína se realizó en la cepa de *E. coli* JM101. En un análisis de SDS -PAGE se observó que la mayor concentración de proteína se encontró en la fracción insoluble, mientras que cerca de un 10% estaba presente en la fracción soluble. Era necesario utilizar únicamente la proteína contenida en la fracción soluble, ya que no sabemos si la desnaturalización es completamente reversible, dados los cambios contenidos en la nueva mutante. Es necesario recordar que la proteína sufrió un cambio en la polarización, la proporción de residuos negativos / positivos se invirtió en la nueva enzima.

Era importante contar con proteína pura totalmente nativa ya que la estimación experimental de la T_m podría resultar errónea si el experimento se iniciaba con proteína parcialmente plegada. La purificación se realizó con una columna de níquel (*Qiagen*) ya que la proteína fue marcada con una bandera de polihistidinas. La marca de seis histidinas presenta afinidad por el metal níquel que se encuentra montado en una resina de Ni-NTA de agarosa.

La proteína fue recuperada mediante eluciones en concentraciones crecientes de imidazol, (la cadena lateral de la histidina); este compuesto libera la proteína de la unión con el níquel ya que la afinidad de este compuesto es mayor que la de las histidinas, debido a que la molaridad del grupo imidazol es mayor en una solución de este compuesto puro que en la solución que contiene a la proteína con la marca de polihistidinas. La proteína no se pudo recuperar en forma pura después de someterla a la columna de níquel. En todas las fracciones recuperadas se copurificaba una banda de 27 kDa. Fue necesario utilizar otra columna para obtener la proteína con una pureza aceptable.

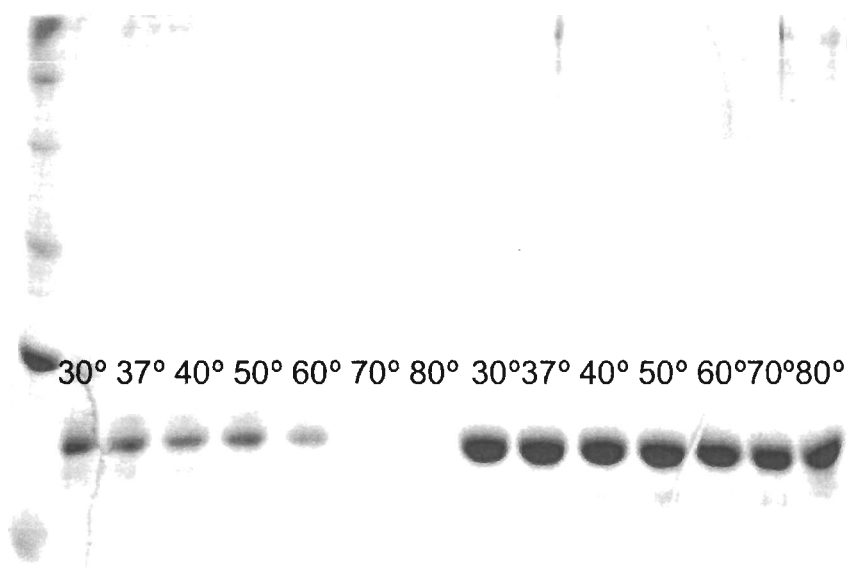


Figura 21. Indicios de termoestabilidad en la proteína consenso. En el análisis de SDS-PAGE se aprecian distintas muestras incubadas en temperaturas crecientes mostrando la concentración relativa residual de la proteína consenso y de PRAI de *E. coli*.

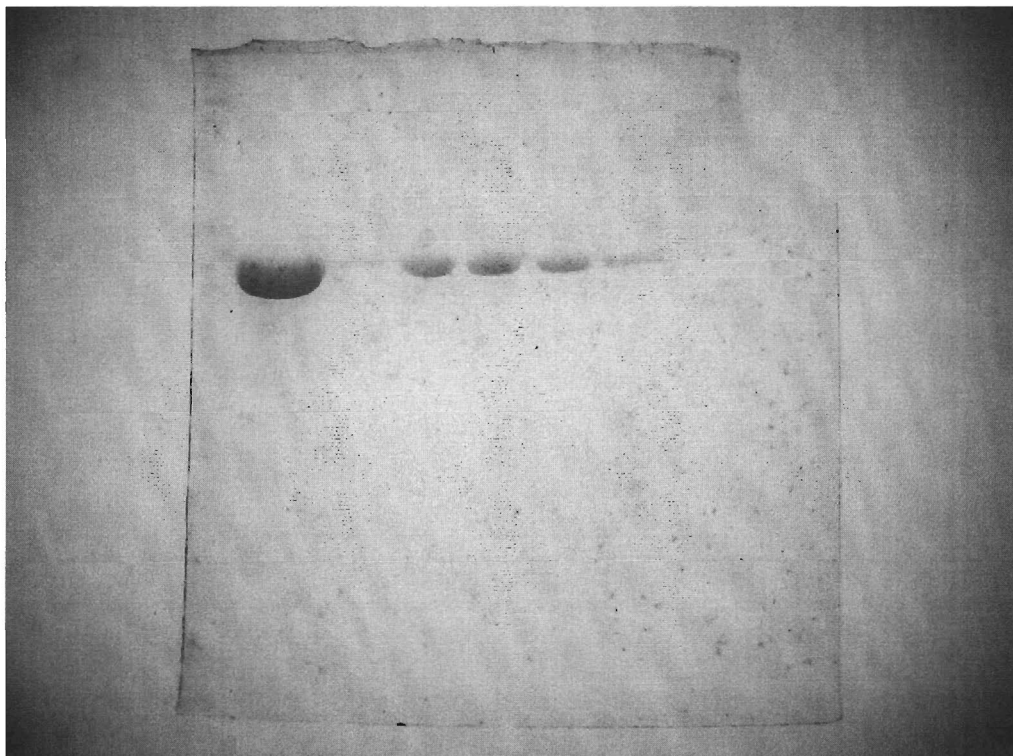


Figura 22. Análisis purificación de proteína consenso mediante exclusión molecular.

En el segundo paso de purificación se utilizó una columna de exclusión molecular. Al recuperar las fracciones que mostraban una mayor concentración en el cromatograma se obtuvo una sola banda analizada mediante SDS-PAGE.

8. Desnaturalización Térmica

Para evaluar la termoestabilidad de la nueva proteína se practicaron ensayos de desnaturalización térmica monitoreada por absorbancia diferencial. La técnica se basa en la exposición diferencial de residuos aromáticos, principalmente triptofano, en función de la temperatura. Generalmente los residuos aromáticos se encuentran en el núcleo de la proteína en donde están protegidos del solvente. Este tipo de conformación tridimensional maximiza la estabilidad global del plegamiento.

La proteína al encontrarse plegada correctamente muestra un espectro característico; con la exposición a distintas temperaturas el espectro se modifica en función de la exposición diferencial de los triptofanos.

Los resultados preliminares de los experimentos de desnaturalización térmica muestran una diferencia bastante significativa en las temperaturas de desnaturalización promedio (T_m) (figuras 23 y 24). Se realizaron experimentos comparativos con las proteínas purificadas PRAI consenso y PRAI de *E. coli*. Las concentraciones de ambas proteínas se estandarizaron a .1mg/ml en amortiguador de fosfatos 20mM y NaCl 300mM pH 7.5 en condiciones reductoras.

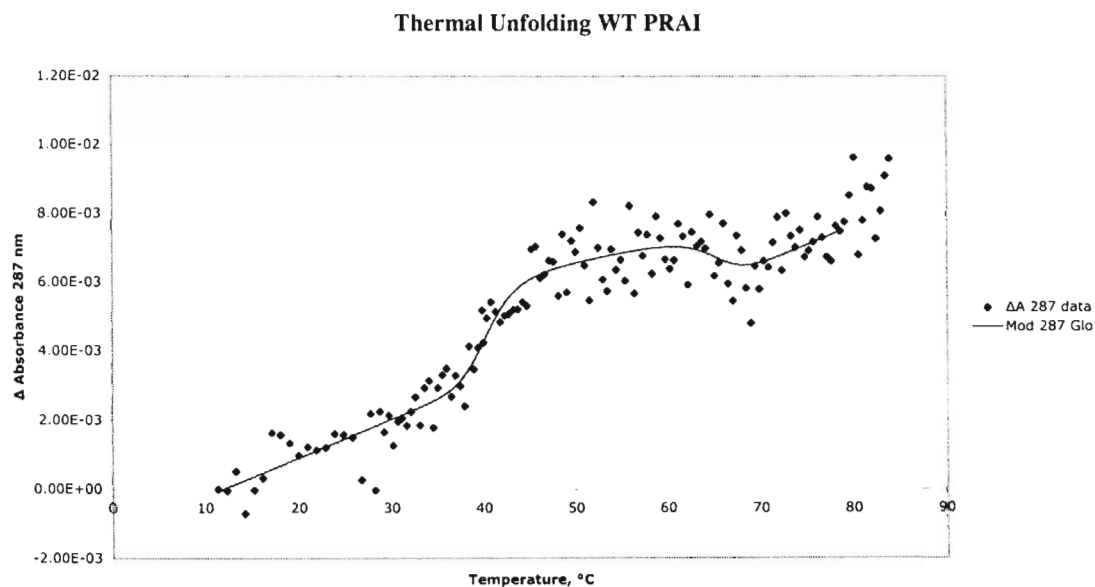


Figura 23. La temperatura promedio observada de desnaturalización para PRAI de *E. coli* fue de 40.71° Celsius con un error de $\pm .71$ y un ΔH de 119.28 kcal/mol con un error de ± 8.91 . Esta temperatura fue obtenida mediante el método experimental de obtención de T_m a través de la ecuación de Van't Hoff..

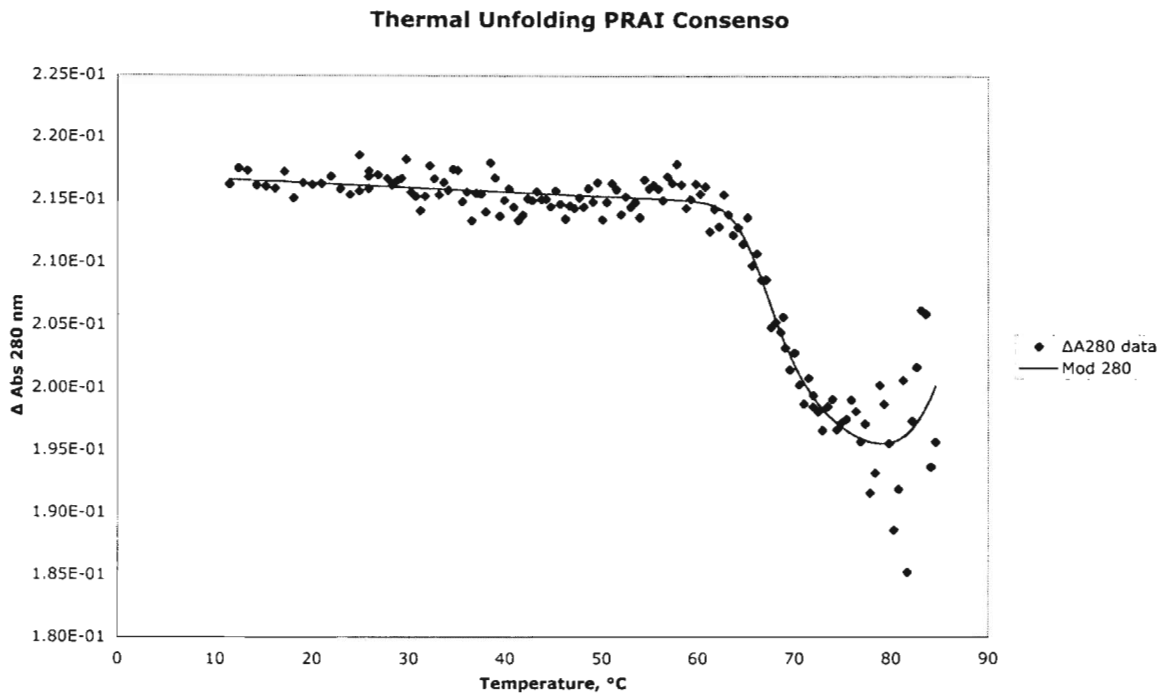


Figura 24. La temperatura promedio obtenida para la proteína consenso fue de 67.89° Celsius con un error experimental de +/- .48° Celsius, el ΔH fue de 98.5 kcal/mol con un error de +/- 1.457.

La diferencia en ambas T_m es de 27.18° C reflejando un aumento claro en la estabilidad de la proteína consenso con respecto a su contraparte de *E. coli*. Los resultados aquí mostrados son preliminares debido a aún no se logrado estandarizar el procedimiento de la desnaturalización térmica. Es necesario obtener réplicas precisas en los experimentos para estimar exactamente la temperatura de desnaturalización de la proteína consenso y poder compararla con la temperatura de desnaturalización de la proteína de *E. coli*. No es posible utilizar los datos termodinámicos reportados en la literatura sobre PRAI de *E. coli* de manera confiable, ya que las condiciones de purificación y de análisis no serían similares.

VIII. DISCUSION

En este trabajo se desarrolló un barril $(\beta/\alpha)_8$ termoestable utilizando una estrategia de diseño semiracional. Un alineamiento múltiple de 12 secuencias de proteínas homologas con actividad de PRAI funcionó como guía para generar una secuencia consenso que presento 49 cambios en la secuencia primaria de aminoácidos y una diferencia de temperatura media de desnaturalización de casi 30° C con respecto a la secuencia de PRAI de *E. coli*, proteína que fue utilizada como una de las secuencias parentales y como control positivo en el experimento. La actividad de la proteína consenso se vio reducida al evaluar el tiempo de complementación comparativamente en medio sólido VB. Sin embargo en medio líquido las cinéticas de crecimiento no mostraron una diferencia significativa.

Durante el análisis bioinformático inicial se habían elegido 126 secuencias de enzimas con actividad de PRAI contenidas en las base de datos *Swiss-prot*. No era posible incluir secuencias que no estuvieran determinadas como enzimas, es decir utilizar ORF's hipotéticos para diseñar el consenso introduciría demasiado ruido. Sin embargo el diseño final de la proteína consenso se baso en un alineamiento múltiple de secuencias obtenido del *HSSP* (46), ya que era fundamental trabajar sobre un alineamiento confiable.

Sólo la sobreposición en la estructura tridimensional exacta de múltiples secuencias con el mismo plegamiento, aseguran obtener una verdadera secuencia consenso, mientras que la elección de un barril TIM canónico y pequeño permitió trabajar el plegamiento como una unidad funcional bien delimitada excluyendo dominios que pudieran haber sido objeto de variación mas pronunciada.

No existen aun demasiados ejemplos de trabajos que permitan definir, tanto, el número adecuado de secuencias finales a incluir en el consenso, como el corte en el % de identidad al resolver la nueva secuencia. Es necesario realizar estudios más detallados con respecto a la información particular que cada secuencia pudiera aportar a la definición del consenso. Si el porcentaje de identidad entre las secuencias es demasiado bajo, la definición del consenso es sumamente difícil y la elección del aminoácido consenso obedece más a un procedimiento arbitrario. Sin embargo la cantidad de información aportada por cada secuencia parental resulta bastante significativa.

La elección del grupo final de secuencias se realizó de manera pragmática, analizando el balance entre la cantidad de residuos que cambiaban al incluir una nueva secuencia y los residuos que no era posible definir (figura 7), también se utilizó el árbol de distancia como una guía para definir los distintos grupos. El porcentaje de identidad se definió de forma arbitraria al analizar el número de residuos que se resolvían como consenso en cada uno de los distintos cortes.

Realizando un estudio comparativo de las propiedades fisicoquímicas de la composición de aminoácidos tanto de la proteína consenso como del control positivo (PRAI de *E. coli*) pudimos observar un cambio en la polaridad de la misma ya que el punto isoeléctrico de PRAI de *E. coli* es 5.74 mientras que el de PRAI consenso es de 9.15, lo que significa que a pH fisiológico presentan diferente carga, la relación de aminoácidos cargados negativamente / positivamente cambio en la nueva proteína, mientras que en PRAI de *E. coli* era de argininas + lisinas = 15 y aspárticos + glutámicos = 22 en la nueva proteína resultó argininas + lisinas = 24 y aspárticos + glutámicos = 18. Es difícil establecer una relación clara entre el cambio en la polaridad de la proteína y la diferencia en temperaturas medias de desnaturalización. Sin embargo esta diferencia se ve reflejada en la agregación diferencial de ambas proteínas en la misma solución amortiguadora.

La sustituciones realizadas con respecto a la proteína control en el consenso se distribuyeron a lo largo de todo el plegamiento, es necesario realizar estudios detallados de las nuevas interacciones no covalentes que se pudieran haber establecido en la nueva proteína. Preliminarmente podemos concluir que posiciones que están relacionadas con la unión al ligando y el sitio activo de la proteína no se vieron afectadas, con excepción de la sustitución A176G que está involucrada en la unión al ligando. Es necesario revertir este cambio en estudios posteriores para averiguar si la proteína es capaz de complementar en medio sólido en el mismo intervalo de tiempo que la silvestre. No todos los sitios de las proteínas se encuentran sometidos a las mismas presiones de selección, sin embargo al observar el consenso podemos más o menos predecir que los sitios definidos mediante este acercamiento se están comportando de manera similar.

Al mimetizar el uso de codones de PRAI de *E. coli* en la retrotranscripción del gen consenso aseguramos niveles de expresión similares a los genes del hospedero. En muchos casos este resulta uno de los pasos limitantes en el experimento, para este estudio se observaron niveles de expresión similares de la proteína control y la consenso al analizar fracciones solubles de ambas proteínas mediante SDS-PAGE. La mayor parte de la proteína expresada en los cultivos bacterianos se encontraba en la fracción insoluble, al sobre expresar ambas proteínas se observaba el mismo patrón, sin embargo la fracción de las dos proteínas que se recuperaba de la fracción soluble resultaba similar.

En la síntesis del gen consenso se utilizaron oligonucleótidos de 60 nucleótidos diseñados por computadora asegurando la exclusión de secuencias que pudiesen formar estructuras secundarias o cualquier otra dificultad en el ensamblado del gen. La rapidez en el ensamblado del gen sintético se atribuye a la purificación minuciosa de los oligonucleótidos, ya que posterior a la síntesis de los mismos se observaban una gran cantidad de subproductos que generarían construcciones incompletas o con secuencias distintas a la diseñada y al buen diseño experimental.

Otro aspecto fundamental del éxito en la síntesis del gen consenso fue la planeación del ensamblado en dos partes y el ensamblaje final a una temperatura tal que permitiera una máxima especificidad en la unión de los segmentos sintetizados inicialmente. Utilizar una concentración 10 veces mayor de los oligonucleótidos externos que de los internos aseguró la obtención final del producto deseado. La búsqueda mediante un gradiente de temperaturas permitió elegir la temperatura óptima en la cual la hibridación de los oligonucleótidos es específica. La conjunción de todos estos factores permitió la síntesis del nuevo gen en un tiempo relativamente corto.

La proteína consenso mantuvo actividad suficiente como para complementar auxotrofia en medio sólido en un lapso de 48 horas, mientras que la proteína de *E. coli* lo hizo en 24 horas. En medio líquido ambas proteínas complementaron de manera similar la auxotrofia de la cepa Δ PRAI observándose un comportamiento similar de las curvas de crecimiento. Aun no podemos atribuir a alguna razón en específico este tipo de observaciones, es posible que la cepa Δ PRAI se comporte de manera distinta posterior a la transformación y plaqueo en medio sólido, que cuando se parte de una concentración igual de bacterias y se inocula medio mínimo líquido.

Por otro lado, el intervalo de tiempo más largo observado en la complementación de auxotrofia por la proteína consenso podría deberse a un aumento en la rigidez de la proteína, implicando una actividad reducida a la temperatura óptima de la proteína mesófila, aunque tal vez se deba a la sustitución A176G. El paradigma flexibilidad / actividad es una explicación recurrente y bastante verosímil desde el punto de vista cinético, las proteínas que presentan un andamiaje más resistente a temperaturas altas se mueven menos al realizar la catálisis, es por eso que su actividad se ve reducida. Sin embargo es necesario realizar experimentos cinéticos a distintas temperaturas para estimar la actividad real de la nueva proteína termófila y poder establecer parámetros comparativos válidos para ambos casos, tanto para la proteína mesófila como para la nueva proteína consenso (42). La única manera exacta de comprobar si los parámetros cinéticos de la nueva enzima son distintos a los de la proteína nativa, es montando una reacción acoplada en la que se incluya la enzima anterior a PRAI en la vía de síntesis de triptofano, ya que se ha demostrado (30) que el sustrato de PRAI es altamente inestable y la única manera de montar una cinética es con su producción *in situ*.

Los experimentos de complementación y de sobre expresión de la proteína consenso se realizaron a 30° C, ya que al parecer la proteína al encontrarse clonada en un vector de expresión fuerte resultaba tóxica para la célula impidiéndole crecer apropiadamente. Fue necesario utilizar una segunda cepa de *E. coli* en el protocolo de purificación ya que la cepa Δ PRAI no resultó óptima para la sobre expresión, JM101 funciona con bastante eficacia en el procedimiento de purificación de ambas proteínas, tanto la consenso como PRAI de *E. coli*. Sin embargo, la utilización de una cepa que contiene el gen silvestre aumenta el riesgo de contaminación y la posible obtención de falsos positivos.

Durante el proceso de purificación se utilizaron dos métodos cromatográficos distintos. No fue posible utilizar las condiciones estándar del proveedor ya que bajo estos protocolos de concentraciones de sal y pH indicados por los manuales del Kit (Qiagen) la proteína no se obtenía con suficiente pureza. Durante la diálisis fue necesario elevar la concentración de NaCl hasta 600 mM para evitar la agregación de la proteína consenso. Al parecer PRAI de *E. coli* no mostró ninguna diferencia en solubilidad al ser dializada contra un tampón 600 mM de NaCl. La utilización de un solo método de purificación no fue suficiente ya que la mayoría de las veces se co-purificaba una proteína contaminante, es por esto que fue necesaria la implementación de un segundo método para lograr una pureza aceptable.

Los experimentos de desnaturalización térmica se realizaron bajo condiciones idénticas en lo que respecta a soluciones amortiguadoras e intervalos de estabilización en el aumento de temperatura. La temperatura media de desnaturalización para PRAI de *E. coli* resultó de 40.71° C con un error de +/- .71° C mientras que para PRAI consenso fue de 67.89° C con un error de +/- .48° C. Los datos fueron normalizados para establecer la línea de base correcta. El espectro observado de la proteína de *E. coli* se comporta de manera esperada y aumenta su absorbancia conforme el grado de exposición de los residuos aromáticos se hace más pronunciado.

Sin embargo, el experimento no se ha realizado de manera tal que permita predecir si el proceso de desnaturalización de esta proteína es reversible o no. El espectro observado del consenso muestra una reducción en la absorbancia conforme se aumenta la temperatura, a diferencia del espectro de la proteína de *E. coli*. Esto sugiere que PRAI consenso se agrega conforme va perdiendo su conformación nativa. Este fenómeno se ve reflejado en una disminución de la absorbancia relativa que presentan los residuos aromáticos de la proteína. Aun así es posible determinar una T_m ya que lo importante es el comportamiento diferencial de los espectros tomados a diferentes temperaturas. Es necesario decir que bajo condiciones de irreversibilidad en el proceso de desnaturalización térmica no es posible calcular un ΔG° , pero sí una T_m mediante el método gráfico (ecuación de Van Hoof).

Existe un argumento que se podría esgrimir en contra de la efectividad del consenso: es posible que cada proteína haya evolucionado hasta adquirir un set único de interacciones cooperativas favorables entre sus residuos y al generar el consenso se modifica esta red generando, una desestabilización disruptiva ya que las substituciones en el consenso estarían fuera de contexto (16). La clave radica en asumir evolución independiente para cada residuo, y que cada secuencia está siendo seleccionada para mantenerse en un umbral de estabilidad únicamente necesario para que pueda realizar su función. Al generar el consenso podemos recuperar la mayoría de las mutaciones estabilizantes seleccionadas independientemente. En esta misma línea, la manera lógica de generar una proteína hipertermoestable es resolver el consenso de un grupo de proteínas naturalmente termoestables, ya que las presiones de selección se habría encargado de potenciar esta característica de las proteínas. Sin embargo si observamos el paradigma flexibilidad / actividad, correríamos el riesgo de generar una proteína que no sea activa en absoluto. De hecho, si este andamiaje hiperestable se pensara utilizar en el diseño de una catálisis en particular, no sería útil, ya que la falta de flexibilidad en la proteína impediría cualquier tipo de actividad.

La teoría que explica las bases termodinámicas que sustentan la diferencia en termoestabilidad de una secuencia consenso con respecto a las parentales, fue formulada en 1994 por Boris Steipe y nos remite al ensamble canónico de secuencias. Asumiendo que la distribución de Boltzman aplica en el ensamblaje canónico de secuencias podemos predecir que el residuo que se observa más frecuentemente posee la más alta adecuación de todas las alternativas posibles en todas las posiciones de la proteína. Con nuestro trabajo hemos logrado predecir un gran número de mutaciones estabilizantes logrando un aumento en termoestabilidad de casi 30° C en un barril $(\beta/\alpha)_8$.

Debido a la naturaleza estocástica de las múltiples interacciones entre residuos que ocurre en las proteínas naturales, es sumamente difícil predecir como afectara una sola mutación a la estabilidad global del polipéptido. El cambio de 49 aminoácidos (con respecto a la secuencia parental de *E. coli*) en un polipéptido de 200 residuos implica un cambio del 25% en la secuencia de la proteína. Existen pocos estudios que no impliquen varios ciclos de evolución dirigida, en los que se haya modificado tal cantidad de residuos y se haya logrado obtener una proteína soluble. Se asume que el aumento en estabilidad es independiente y aditivo por cada residuo substituido (distribución de Boltzman) sin embargo, para poder asignar algún valor definitivo a cada mutación individual es necesario caracterizar 49 mutantes sencillas. Este tipo de análisis es tan solo uno de los que se deben realizar en este nuevo andamiaje, ya que es necesario evaluar como se distribuyeron los residuos consenso en la estructura tridimensional, para poder inferir patrones de correlación y de establecimiento de nuevas interacciones electrostáticas que estén mejorando el empacamiento de la enzima.

El homologar la estructura observada de los plegamientos en la naturaleza, con un sistema termodinámico que es descrito a través de sus microestados más comunes nos ha permitido descubrir con bastante certeza los aminoácidos que están determinando la estabilidad de proteínas problema. La obtención de una fitasa con una temperatura media de desnaturalización de entre 15° a 26° C mayor que cualquiera de las parentales por parte del grupo de Lehmann es uno de los antecedentes mas impactantes del éxito de esta estrategia. Hay que resaltar nuestros resultados preliminares muestran un aumento en estabilidad de casi 30 grados, demostrando que el acercamiento consenso es una estrategia de termoestabilización generalizable a varios tipos de plegamientos y en especial al barril TIM.

El haber logrado estabilizar una proteína con un plegamiento $(\beta/\alpha)_8$ es de gran importancia para la ciencia de proteínas, ya que permite pensar en la estabilización de casi cualquier enzima, ampliando las posibilidades de incorporar casi cualquier actividad enzimática en un proceso industrial. Hay que recordar la amplia distribución filogenética y funcional de este plegamiento, por lo que su estabilización es de extrema importancia. Ahora, tenemos una estructura que ha sido seleccionada por la evolución como una de las formas tridimensionales mas exitosas en cuanto a versatilidad de funciones se refiere, y que además se logró termoestabilizar.

El potencial de este plegamiento en la búsqueda de biocatálisis no naturales es enorme, lo demuestran múltiples evidencias: a) la plasticidad en la autonomía del plegamiento que demostró Shukla al cambiar de lugar las unidades (β/α) en el barril y obtener proteínas plegadas, b) la demostración de que las unidades (β/α)₄ son capaces de plegarse y formar proteínas catalíticamente activas hecha por Sterner, (hablándonos de una posible evolución de este plegamiento por duplicación y fusión) y c) nuestro trabajo, que estabilizó de manera rápida y sencilla este exitoso plegamiento.

La generación de barriles TIM estables abre un abanico de posibilidades en el estudio de la relación estructura función en proteínas. Mediante el diseño de proteínas quiméricas fusionadas a andamiajes estables (43) es posible estudiar proteínas que normalmente serían lábiles en su forma natural. También es posible implementar la búsqueda de plegamientos nuevos, al utilizar dominios estables que ya se han seleccionado a lo largo de toda la historia evolutiva, como templados alrededor de los cuales secuencias codificantes naturales adquieran una estructura tridimensional estable.

**ESTA TESIS NO SALE
DE LA BIBLIOTECA**

IX. CONCLUSIONES

Diseñamos una secuencia consenso utilizando un alineamiento de secuencias homólogas de PRAI, una enzima que presenta el plegamiento de barril (β/α)₈. La secuencia del gen sintético fue determinada a partir de la secuencia de aminoácidos respetando el uso de codones del organismo hospedero. Realizamos la síntesis del gen sintético mediante el ensamblaje de oligonucleótidos alternados en reacciones secuenciales de PCR. Evaluamos la capacidad de complementación del nuevo barril TIM consenso, demostrando que es capaz de complementar el crecimiento de una cepa Δ PRAI. Se logró obtener proteína soluble a partir de nuestro sistema de expresión. Purificamos el barril (β/α)₈ utilizando dos métodos cromatográficos distintos, hasta alcanzar la pureza necesaria para realizar experimentos termodinámicos. Realizamos la caracterización biofísica del barril TIM consenso mostrando un aumento en termoestabilidad de casi 30° C en relación con la proteína utilizada como control positivo (PRAI de *E. coli*).

X. PERSPECTIVAS

Es necesario continuar con la caracterización biofísica de la nueva proteína, hasta lograr una estandarización en los experimentos de desnaturalización térmica, para obtener un dato concluyente y poder definir exactamente la ganancia en termoestabilidad. Son necesarios estudios de permeación en gel para estimar el estado de oligomerización de la nueva proteína, así como su posterior purificación a partir de una cepa deletada en el gen de PRAI para evitar cualquier posible contaminación del producto silvestre. Realizar ensayos de desnaturalización con urea permitirá explorar la estabilidad desde otros aspectos distintos a la temperatura. Curvas de dicroísmo circular permitirán evaluar la formación de estructuras secundarias similares a las de la proteína nativa.

Se propone un mapeo detallado de los lugares en donde se definieron residuos consenso para tratar de estimar la formación de nuevas interacciones electrostáticas, que pudieran incrementar la termoestabilidad de la proteína diseñada. Es necesario evaluar comparativamente la naturaleza fisicoquímica de los nuevos residuos, para tratar de establecer si su presencia esta contribuyendo a aumentar el ΔG del plegamiento haciendo, mas difícil la reacción reversible hacia el estado desplegado.

Se realizarán dinámicas moleculares para estimar como están cambiando los patrones de vibración del polipéptido, en relación con la proteína utilizada como control, y confirmar si realmente existió un aumento en la rigidez global del plegamiento. Se realizará una filogenia conteniendo a todas las secuencias parentales que definieron el consenso, para comparar la similitud entre un probable ancestro y la secuencia consenso; esto permitirá realizar reconstrucciones filogenéticas que permitan confirmar si existió un ancestro termófilo en determinados grupos de bacterias.

XI. REFERENCIAS

1. Nei, masatoshi pp 1-7
Molecular evolutionary genetics
Columbia university press 1987. USA
2. Li, Wen-Hsiung Li pp 59 - 63
Fundamentals of Molecular Evolution
Sinauer associates, Inc 1999 USA
3. Yano JK, Poulos TL.
New understandings of thermostable and peizostable enzymes.
Curr Opin Biotechnol. 2003 Aug;14(4):360-5.
4. Bolon DN, Voigt CA, Mayo SL.
De novo design of biocatalysts.
Curr Opin Chem Biol. 2002 Apr;6(2):125-9.
5. Arnold FH, Wintrode PL, Miyazaki K, Gershenson A.
How enzymes adapt: lessons from directed evolution.
Trends Biochem Sci. 2001 Feb;26(2):100-6.
6. Wise EL, Rayment I.
Understanding the importance of protein structure to nature's routes for divergent evolution in TIM barrel enzymes.
Acc Chem Res. 2004 Mar;37(3):149-58
7. Eijsink VG, Bjork A, Gaseidnes S, Sirevag R, Synstad B.,
Rational engineering of enzyme stability.
J Biotechnol. 2004 Sep 30;113(1-3):105-20.
8. Dr. Segovia L.
Comunicación personal
9. Baltzer L. DeGrado F.
Egineering and design. Expanding the protein world
Curr. Op. in Struct. Biol. 14;455-457
10. Baltzer L. DeGrado F.
Elucidation of determinants of protein stability through genome sequence analysis.
Chakravarty S, Varadarajan R.
FEBS Lett. 2000 Mar 17;470(1):65-9.

11. Lehmann M, Pasamontes L, Lassen SF, Wyss M.
The consensus concept for thermostability engineering of proteins.
Biochim Biophys Acta. 2000 Dec 29;1543(2):408-415.
12. Tao H, Cornish VW.
Milestones in directed enzyme evolution.
Curr Opin Chem Biol. 2002 Dec;6(6):858-64.
13. Stemmer WP.
Rapid evolution of a protein in vitro by DNA shuffling.
Nature. 1994 Aug 4;370(6488):389-91.
14. Schmidt-Dannert C, Arnold FH.
Directed evolution of industrial enzymes.
Trends Biotechnol. 1999 Apr;17(4):135-6.
15. Chen R.
Enzyme engineering: rational redesign versus directed evolution.
Trends Biotechnol. 2001 Jan;19(1):13-4.
16. Steipe B.
Consensus-based engineering of protein stability: from intrabodies to thermostable enzymes.
Methods Enzymol. 2004;388:176-86.
17. Adams MW, Kelly RM.
Finding and using hyperthermophilic enzymes.
Trends Biotechnol. 1998 Aug;16(8):329-32.
18. Lehmann M, Kostrewa D, Wyss M, Brugger R, D'Arcy A, Pasamontes L, van Loon AP.
From DNA sequence to improved functionality: using protein sequence comparisons to rapidly design a thermostable consensus phytase.
Protein Eng. 2000 Jan;13(1):49-57.
19. Giver L, Gershenson A, Freskgard PO, Arnold FH.
Directed evolution of a thermostable esterase.
Proc Natl Acad Sci U S A. 1998 Oct 27;95(22):12809-13.
20. Lehmann M, Loch C, Middendorf A, Studer D, Lassen SF, Pasamontes L, van Loon AP, Wyss M.
The consensus concept for thermostability engineering of proteins: further proof of concept.
Protein Eng. 2002 May;15(5):403-11.

21. Perl D, Mueller U, Heinemann U, Schmid FX.
Two exposed amino acid residues confer thermostability on a cold shock protein.
Nat Struct Biol. 2000 May;7(5):380-3.
22. Nikolova PV, Henckel J, Lane DP, Fersht AR
Semirational design of active tumor suppressor p53 DNA binding domain with enhanced stability.
Proc Natl Acad Sci U S A. 1998 Dec 8;95(25):14675-80.
23. Wang Q, Buckle AM, Foster NW, Johnson CM, Fersht AR.
Design of highly stable functional GroEL minichaperones.
Protein Sci. 1999 Oct;8(10):2186-93.
24. Steipe B, Schiller B, Pluckthun A, Steinbacher S.
Sequence statistics reliably predict stabilizing mutations in a protein domain.
J Mol Biol. 1994 Jul 15;240(3):188-92.
25. Hocker B, Jurgens C, Wilmanns M, Sterner R
Stability, catalytic versatility and evolution of the (beta alpha)(8)-barrel fold.
Curr Opin Biotechnol. 2001 Aug;12(4):376-81.
26. Wierenga RK.
The TIM-barrel fold: a versatile framework for efficient enzymes.
FEBS Lett. 2001 Mar 16;492(3):193-8.
27. Nagano N, Orengo CA, Thornton JM.
One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions.
J Mol Biol. 2002 Aug 30;321(5):741-65.
28. Hocker B, Claren J, Sterner R.
Mimicking enzyme evolution by generating new (beta alpha)8-barrels from (betaalpha)4-half-barrels.
Proc Natl Acad Sci U S A. 2004 Nov 23;101(47):16448-53. Epub 2004 Nov 11.
29. Vega MC, Lorentzen E, Linden A, Wilmanns M.
Evolutionary markers in the (beta/alpha)8-barrel fold.
Curr Opin Chem Biol. 2003 Dec;7(6):694-701.
30. Hommel U, Eberhard M, Kirschner K.
Phosphoribosyl anthranilate isomerase catalyzes a reversible amadori reaction.
Biochemistry. 1995 Apr 25;34(16):5429-39.

31. Eberhard M, Tsai-Pflugfelder M, Bolewska K, Hommel U, Kirschner K.
Indoleglycerol phosphate synthase-phosphoribosyl anthranilate isomerase: comparison of the bifunctional enzyme from Escherichia coli with engineered monofunctional domains.
Biochemistry. 1995 Apr 25;34(16):5419-28.
32. Sterner R, Kleemann GR, Szadkowski H, Lustig A, Hennig M, Kirschner K.
Phosphoribosyl anthranilate isomerase from Thermotoga maritima is an extremely stable and active homodimer.
Protein Sci. 1996 Oct;5(10):2000-8.
33. Sharp PM, Li WH.
The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications.
Nucleic Acids Res. 1987 Feb 11;15(3):1281-95.
34. Jansen R, Bussemaker HJ, Gerstein M.
Revisiting the codon adaptation index from a whole-genome perspective: analyzing the relationship between gene expression and codon occurrence in yeast using a variety of models.
Nucleic Acids Res. 2003 Apr 15;31(8):2242-51.
35. Gustafsson C, Govindarajan S, Minshull J.
Codon bias and heterologous protein expression.
Trends Biotechnol. 2004 Jul;22(7):346-53.
36. Voet D.
Biochemistry
Third Ed. Vol 1. John Wiley & Sons Inc. 2004 USA
37. Sambrook J., Fritsch E.F., Maniatis T.
Molecular cloning. A laboratory Manual
Second Ed. Vol 1, 2, 3. Cold Spring Harbor Laboratory Press 1989
38. Vogel H.J., Bonner D.M.
Acetylornithinase of Escherichia coli: partial purification and some properties.
J Biol Chem. 1956 Jan;218(1):97-106.
39. <http://www.changbioscience.com/primo/primoo.html>
40. <http://www.kazusa.or.jp/codon/>

41. Vazquez R. pp 77 - 85
Termodinámica biológica
Primera edición AGT Editor S.A. Mexico 2002
42. Ivens A, Mayans O, Szadkowski H, Wilmanns M, Kirschner K.
Purification, characterization and crystallization of thermostable anthranilate phosphoribosyltransferase from Sulfolobus solfataricus.
Eur J Biochem. 2001 Apr;268(8):2246-52.
43. Protein Eng Des Sel. 2004 Jul;17(7):553-5. Epub 2004 Aug 27.
A thermostable enzyme as an experimental platform to study properties of less stable homologues.
Lill H, Hisabori T, Groth G, Bald D.
44. Kumar S, Nussinov R.
Salt bridge stability in monomeric proteins.
J Mol Biol. 1999 Nov 12;293(5):1241-55.
45. The QIAexpressionist
A handbook for high- level expression and purification of 6xHis-tagged proteins
June 2003 Fifth Edition
46. Dodge C, Schneider R, Sander C.
The HSSP database of protein structure-sequence alignments and family profiles.
Nucleic Acids Res. 1998 Jan 1;26(1):313-5.
47. Shukla A, Guptasarma P.
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)₈-barrel.
Proteins. 2004 May 15;55(3):548-57