UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAESTRIA Y DOCTORADO EN CIENCIAS BIOQUIMICAS

INSTITUTO DE BIOTECNOLOGIA

DEPARTAMENTO DE INGENIERÍA CELULAR Y BIOCATÁLISIS

"Explorando la relación estructura-función de las asas β/α de un barril $(β/α)_8$: implicaciones para el diseño de nuevas funciones enzimáticas"

T E S I S

PARA OBTENER EL GRADO ACADEMICO DE

DOCTOR EN CIENCIAS

PRESENTA:

MC. Adrián Ochoa Leyva

DIRECTOR DE TESIS: DR. XAVIER SOBERON MAINERO

CUERNAVACA, MORELOS, Junio de 2011

El presente trabajo se realizó bajo la dirección del
Dr. Francisco Xavier Soberón Mainero
en el **Departamento de Ingeniería Celular y Biocatálisis** del
Instituto de Biotecnología
de la **Universidad Nacional Autónoma de México**.
Para la realización del mismo se contó con el financiamiento del Consejo Nacional de Ciencia y Tecnología (CONACYT).

Miembros del Jurado:

PRESIDENTE.
Dr. Mario Soberón Chávez

SECRETARIO
Dr. Lorenzo Segovia Forcella

VOCAL
Dra. Rosario Muñoz Clares

SUPLENTE.
Dr. Luis G. Brieba de Castro

SUPLENTE
Dr. Xavier Soberón Mainero

Agradecimientos:

A CONACYT, la UNAM y el IBT.

A mi esposa Liz Carol

A mi hijo Luis Adrian

Al Dr. Xavier Soberón Mainero

Al Dr. Francisco Barona Gómez

A mi comité tutoral:
Dr. Francisco Barona Gómez
Dr. Enrique Rudiño Piñera

Al personal del Laboratorio:
Dr. Joel Osuna
Dra. Gloria Saab
Dr. Humberto Flores
Biol. Filiberto Sanchez
Sr. Javier Dorantes López

A la unidad de síntesis y secuenciación del IBt-UNAM:
Dr. Paul Gaytan
M.C. Jorge A. Yañez
Q.I. Santiago Becerra
M.C. Eugenio Lopez Bustos

A los integrantes del laboratorio Morett-Soberón del IBT

A los integrantes del laboratorio Barona-Montiel del Langebio

A mi familia

Al Langebio

A mis amigos en Cuernavaca

A mis amigos en Irapuato

INDICE

Índice de Figuras

Índice de tablas

Abreviaturas

ADN:	Ácido Desoxirribonucleico
CAT:	Cloranfenicol Acetil Transferasa
CdRP:	1-(o-carboxifenilamino)-l-deoxirribulosa 5-fosfato
PCR:	Reacción en Cadena de la Polimerasa
PDB:	Banco de Datos de Proteína
PRA:	N- (5´- fosforribosil) antranilato
CdRP:	1-[(2-carboxifenilamino)-1-deoxirribulosa 5-fosfato
PRFAR:	N'-[(5'-fosforribulosil)-formimino]-5-aminoimidazol-4-carboxamida-ribonucleótido
ProFAR:	N'-[(5'-fosforribosil)-formimino]-5 aminoimidazol-4-carboxamida-ribonucleótido
TrpA:	Subunidad alfa de Triptofano Sintetasa
TrpF:	N-(5'-fosforribosil)-antranilato Isomerasa
PriA:	Fosforribosil isomerasa A
DTT:	ditiotreitol
PMSF:	fluoruro de fenilmetilsulfonilo
RPM:	Revoluciones por minuto
UFC:	Unidad formadora de colonia
SCLE:	Intercambio sistemático de asas catalíticas, del inglés: Systematic Catalytic Loop Exchange
SFLA:	Adaptabilidad de la relación estructura-función del asa, del inglés: Structure-Function Loop Adaptability
IPTG:	isopropil-β-D-1-tiogalactopiranósido

Resumen

Uno de los principales objetivos del diseño de enzimas es construir proteínas con propiedades mejoradas o nuevas. Esta capacidad de diseñar enzimas con las actividades deseadas tiene una aplicación importante en la industria química, agrícola y farmacéutica. Para alcanzar este objetivo normalmente se emplean dos estrategias de ingeniería de proteínas: el diseño racional y la evolución dirigida. Recientemente, el desarrollo de algoritmos computacionales ha permitido el diseño de enzimas que han cambiado de función, pero cuyas nuevas actividades están órdenes de magnitud por debajo de la eficiencia catalítica de la enzima silvestre. Probablemente es necesario involucrar cambios de secuencia más drásticos que nos permitan explorar más ampliamente el espacio de secuencia de la proteína de interés. Con ello en mente en el presente trabajo planteamos una nueva estrategia de exploración del espacio de secuencia en una enzima $(beta/alpha)_8$ isomerasa. Dicha estrategia involucró el intercambio de asas completas y la introducción de variabilidad de secuencia en los residuos que las unieron con la enzima molde, además del rediseño de los residuos catalíticos. La enzima molde se denomina TrpF, PRA isomerasa o PRAI y está presente en la ruta de biosíntesis del triptófano. En este trabajo mostramos como se pueden adaptar funcionalmente asas de diferentes enzimas en la enzima TrpF y cómo esta adaptación funcional puede ser medida con un ensayo *in vivo*. La inserción de variabilidad en ambos lados de cada asa intercambiada se hizo con el objetivo de imitar el mecanismo natural con el cual las asas de los anticuerpos adquieren variabilidad en sus sitios de unión. También obtuvimos detalles a nivel molecular de algunas variantes mediante análisis de cinética enzimática, dicroísmo circular y filtración en gel. Como último punto, siguiendo esta estrategia de intercambio de asas y de rediseño del sitio activo discutimos la obtención de una nueva actividad enzimática en la enzima TrpF.

1. Introducción

Actualmente uno de los retos mas activos en el campo de la investigación en biocatálisis y biotransformación es el descubrimiento, diseño y optimización de enzimas (*1, 2*). ¿Cómo podemos encontrar o diseñar enzimas que catalizen importantes reacciones químicas, incluso cuando éstas no existan en la naturaleza? ¿Cómo podemos optimizar diferentes enzimas para reconocer sustratos no biológicos y para que funcionen en condiciones de reacción industriales? ¿Cómo ensamblar de manera eficiente múltiples enzimas en nuevas rutas de biosíntesis de varios pasos? La mayoría de las estrategias utilizadas comúnmente para encontrar respuestas a estas preguntas incluyen aproximaciones de bioinformática, evolución dirigida, diseño racional y una combinación de todas ellas (*3, 4*).

Las enzimas que podemos observar hoy en día son el producto de la evolución biológica que ha tomado millones de años y suelen catalizar una determinada reacción con alta especificidad y enantioselectividad. Por ello las enzimas son ampliamente utilizadas en muchos procesos industriales como catalizadores biológicos. Sin embargo, a pesar de la perfecta adaptación a su rol fisiológico, la actividad y la estabilidad a menudo están muy lejos de lo que la química orgánica o los procesos industriales de gran escala necesitan (*5*). Por lo que con el fin de aprovechar plenamente el potencial de las enzimas en los procesos industriales y su aplicación práctica aún quedan muchos retos que mejorar. Las cualidades que normalmente se necesitan mejorar son: la eficiencia, la disponibilidad en altas cantidades, los precios bajos, la poca inhibición por producto y la mejor estabilidad (*3, 4*).

Uno de los principales objetivos del rediseño de enzimas es construir proteínas con propiedades mejoradas o nuevas. Esta capacidad de diseñar enzimas con las actividades deseadas a partir de una enzima silvestre tiene una aplicación práctica importante en la industria química, agrícola y farmaceútica (*1, 2*). Para alcanzar este objetivo normalmente se emplean dos estrategias de ingeniería de proteínas: la primera es el diseño racional y la segunda es la evolución dirigida (*4*). A través del diseño racional podemos modificar específicamente una determinada región o propiedad de una enzima, pero para ello es necesario tener información previa sobre su secuencia, estructura y función. Por otro lado, la evolución dirigida surge como una alternativa al

diseño racional y no requiere de un conocimiento detallado de la estructura ni del mecanismo de reacción de la enzima a modificar. Esta estrategia se enfoca en la generación de variabilidad al azar en el gen de la proteína de interés y la posterior selección de la actividad deseada, imitando el proceso de la evolución natural (*6-8*). Por estas características la evolución dirigida es ampliamente utilizada como primera estrategia para modificar enzimas tanto en laboratorios académicos como industriales. Aunque en la optimización de enzimas los mejores resultados se han obtenido al combinar ambas estrategias, en las que por ejemplo primero se realizan mutaciones específicas por diseño racional y posteriormente la enzima es sometida a ciclos de evolución dirigida o viceversa {Chica, 2005 #26;Chica, 2005 #59}.

Recientemente, el desarrollo de potentes algoritmos computacionales ha permitido la aplicación de enfoques novedosos, diseñando proteínas que han cambiado de función, pero cuyas nuevas actividades están órdenes de magnitud por debajo de la eficiencia catalítica correspondiente a la enzima silvestre (*9, 10*). Probablemente sea necesario involucrar cambios de secuencia más drásticos que nos permitan explorar más ampliamente el espacio de secuencia de la proteína de interés (*4*). Con ello en mente, el presente trabajo plantea una estrategia que involucra el intercambio de asas completas y la introducción de variabilidad de secuencia en los residuos que las unirán con la enzima molde, además del rediseño de los residuos catalíticos de una enzima $(\beta/\alpha)_8$ isomerasa.

El entendimiento de la relación estructura-función de las asas es de gran interés en la ingeniería de proteínas (*11*) debido a que es conocido que las asas juegan un rol crítico en la función enzimática, incluyendo la catálisis (*12*), la especificidad de sustrato (*13*) y las interacciones proteína-proteína (*14*). Nuestro trabajo se enfoca en enzimas que poseen el plegamiento de barril $(\beta/\alpha)_8$. Este plegamiento es uno de los más representados comúnmente en las enzimas conocidas, además posee una gran versatilidad funcional y estructural (*15*). Convenientemente los residuos del sitio activo de las enzimas que poseen este plegamiento se encuentran localizados en la sección C-terminal de las hebras β así como en las asas que unen las hebras β con las hélices α (asas β/α), las cuales se consideran que no influyen en la estabilidad de la enzima (*15, 16*). En este sentido, el objetivo del presente proyecto es tratar de entender cómo se

puede adaptar la relación estructura-función de las asas β/α en el plegamiento de barril $(\beta/\alpha)_8$ y con ello proponer estrategias para diseñar nuevas funciones enzimáticas en estas proteínas.

Las tres enzimas que constituyen nuestro modelo de estudio son: TrpF (PRA isomerasa o PRAI, EC 5.3.1.24) de *Escherichia coli*, HisA (ProFAR isomerasa, EC 5.3.1.16) de *Thermotoga maritima* y PriA (PRA y ProFAR isomerasa, EC 5.3.1.24 y 5.3.1.16) de *Streptomyces coelicolor*. La enzima TrpF está presente en la ruta de biosíntesis del triptófano y cataliza un rearreglo de Amadori del sustrato PRA hacia el producto CdRP (*17*). La enzima HisA se encuentra en la ruta de biosíntesis de la histidina y utiliza un mecanismo de reacción similar al de TrpF para convertir una aminoaldosa en una aminocetosa: del sustrato ProFAR al producto PRFAR (*18*). La enzima PriA es homológa de HisA y fue encontrada formando parte de la biosíntesis de triptófano e histidina en algunas bacterias del género de los actinomicetos (*19*). PriA posee tanto la actividad de PRA como de ProFAR isomerasa y es muy similar estructuralmente a HisA (*20*). Se ha sugerido que cambios conformacionales de las asas β/α 5 y 6 son esenciales para que esta enzima posea ambas actividades catalíticas (*20, 21*). Cambios en la longitud e identidad de secuencia de las asas se han asociado a menudo con la evolución de nuevas actividades enzimáticas en la naturaleza (*22*).

2. Antecedentes

Existen diversos ejemplos donde se muestran cambios de actividad o especificidad con una o pocas mutaciones pero con parámetros catalíticos lejos de parecerse a los de las enzimas silvestres. Teóricamente la evolución de diferentes funciones enzimáticas en superfamilias mecanísticamente diversas se inició por procesos que involucraron una o pocas mutaciones, por lo que los experimentos *in vitro* de este tipo proporcionan información sobre las primeras etapas de la evolución natural divergente (*23*). Sin embargo, muy a menudo los alineamientos de secuencias de proteínas revelan muchas inserciones y deleciones, lo cual resalta su importancia en el diseño natural (*16, 24, 25*). Aunado a ello, las secuencias de enzimas que presentan diferentes funciones dentro de una superfamilia típicamente comparten una identidad de secuencia menor al 40%. Esto puede explicar por qué unas pocas mutaciones puntuales no pueden producir variantes que catalicen la "nueva" reacción con valores altos de k_{cat} o de k_{cat}/k_M, por lo cual es probable que para lograr mejores valores catalíticos sean necesarias estrategias de mutagénesis *in vitro* más agresivas (*4*). Pero además se requiere entender cómo es que cambios o rearreglos de secuencia como inserciones o deleciones podrían adaptarse para mejorar o cambiar funciones enzimáticas (*24*).

En el transcurso del desarrollo de nuestro proyecto se reportó un trabajo de un rediseño agresivo del sitio activo de una glioxilasa II (GlxII) la cual fue convertida a una β-lactamasa (IMP-1) mediante inserciones/deleciones e intercambio no sistemático de secciones de asas (*26*). Inicialmente, GlxII e IMP-1 compartían una baja identidad de secuencia, diferentes sitios de unión de metales y asas de diferentes tamaños y secuencias. La modificación de la enzima GlxII incluyó la eliminación de todo un dominio, el rediseño del sitio de unión a metales y la inserción de variabilidad natural y al azar en las asas de especificidad, así como la aplicación de diferentes estrategias de evolución dirigida. La variante ganadora adquirió la actividad de β-lactamasa y resultó mantener un 59% de identidad de secuencia con su progenitora y un 25% de identidad con la enzima blanco, pero aún con todas las modificaciones hechas los parámetros cinéticos fueron considerablemente mucho más bajos que los de la enzima silvestre IMP-1. Una de las conclusiones importantes de este trabajo es que el rediseño de las

asas es claramente esencial para explorar más ampliamente el espacio de secuencia funcional de una enzima (*25*). Además, dado que la mejor enzima obtenida en este trabajo tiene una actividad enzimática muy baja, también demuestra que aún falta entender como adaptar mejor la relación estructura-función de las asas y con ello poder mejorar la actividad enzimática.

3. Hipótesis:

Las asas β/α de los barriles $(\beta/\alpha)_8$ contienen información funcional (incluida la especificidad de sustrato) que puede ser trasladada y adaptada a la enzima TrpF de *E. coli* mediante novedosas estrategias experimentales, permitiéndonos ampliar el conocimiento para diseñar nuevas funciones enzimáticas.

4. Objetivos:

4.1 Objetivo general:

Entender como adaptar la relación estructura-función de las asas β/α de diferentes barriles $(\beta/\alpha)_8$ en la enzima TrpF de *E. coli* mediante intercambio de asas, introducción de variabilidad en sus sitios de unión, diseño racional del sitio activo y evolución dirigida, y con ello explorar la posibilidad de diseñar la nueva función de ProFAR isomerasa en esta enzima.

4.2 Objetivos específicos:

- Concluir la caracterización estructural del primer método de intercambio de asas β/α en un barril $(\beta/\alpha)_8$ modificado: SCLE, de sus siglas en inglés: Systematic Catalytic Loop Exchange.
- Diseñar una nueva estrategia para intercambiar asas β/α en la enzima TrpF que nos permita introducir variabilidad en ambos sitios de unión de cada asa y analizar la relación estructura-función de cada asa intercambiada en esta enzima.
- Desarrollar una estrategia para medir cuantitativamente la adaptabilidad de la relación estructura-función de las asas β/α intercambiadas en TrpF.
- Rediseñar el sitio activo de TrpF de tal manera que sea más grande y esto nos permita analizar sí sólo con estos cambios es posible obtener la nueva actividad de ProFAR isomerasa.
- Intercambiar las asas β/α importantes para la función de ProFAR isomerasa de PriA en TrpF, así como sus residuos catalíticos y analizar *in vivo* la nueva actividad de ProFAR isomerasa en TrpF.

5. Metodología

Manipulación de ADN

Se utilizaron los procesos estándar para la preparación de plásmidos, reacciones de digestión, reacciones de ligación, transformación y electroforesis. Todas las construcciones se hicieron en el plásmido pDAN5-ped1 (*16*). Las variantes analizadas en fusión con la enzima CAT (Cloranfenicol Acetil Transferasa) se construyeron en el plásmido pDAN5-trpF-loxP-CAT (*16*). Tanto los productos de PCR como los plásmidos empleados en este trabajo se purificaron y analizaron en geles de agarosa al 1 %.

Intercambio de asas en la secuencia del gen de TrpF

Para cada asa se diseñaron dos oligonucleótidos parcialmente complementarios en 12 pb entre sí: uno corresponde a la hebra no codificante y el otro a la hebra codificante. En ambos oligonucleótidos se introdujo un codón NNS el cual reemplaza al respectivo residuo bisagra. Las librerías se construyeron por PCR de manera independiente, utilizando el correspondiente par de oligonucleótidos para cada asa y los dos oligos que flanquean el gen de TrpF. De esta manera, la primera mitad del gen de cada librería fue amplificada utilizando como templado el plásmido pDAN5_trpF y los oligonucleótidos Hind3AOL como 5´-primer (el cual tiene un sitio de restricción *HindIII*) y el oligonucleótido no codificante del asa correspondiente como 3´-primer. La segunda mitad se amplificó utilizando el mismo plásmido como templado y el correspondiente oligonucleótido codificante del asa como 5´-primer y el oligonucleótido NheIAOL (el cual tiene el sitio de restricción *NheI*) como 3´-primer. Los productos amplificados se purificaron de un gel de agarosa al 1%. Finalmente, cada librería se construyó por "overlapping-extension PCR" utilizando como templado el par de productos de PCR anteriormente purificados y los oligonucleótidos HindIIIAOL como 5´-primer y NheIAOL como 3´-primer. Estos productos finales se purificaron de la misma manera que los anteriores y luego se digirieron con las enzimas de restricción *HindIII* y *NheI*. Los productos de la reacción se purificaron y se ligaron en el vector pDAN5 previamente digerido con las mismas enzimas. Para confirmar la correcta inserción de cada asa, la distribución de los residuos en el sitio NNS y la eficiencia de construcción de nuestro

método, se secuenciaron aproximadamente 20 plásmidos de cada librería. Para mayor detalle ver el manuscrito de la sección 2 de resultados (24).

Inserción de mutaciones puntuales

Las mutaciones puntuales se insertaron utilizando la estrategia de megaprimer (46). Para el caso de las librerías con sitios de variabilidad NNS se utilizó la misma estrategia y siempre asegurando tener la librería final al menos 5 veces representada experimentalmente.

Fusiones al reportero de plegamiento CAT

Para fusionar las variantes a CAT se subclonaron del plásmido pDAN5 al plásmido pDAN5-trpF-loxP-CAT utilizando la estrategia previamente descrita (*16*).

Métodos de selección

Para analizar *in vivo* las actividades enzimáticas de PRA isomerasa (TrpF) y de ProFAR isomerasa (HisA) se utilizaron las cepas de selección de *E. coli* JM101Δ*trpF* y HFRG6, respectivamente (*20*). La primera cepa tiene una deleción del gen de TrpF de su genoma y por lo tanto en medio mínimo no complementa la actividad de PRA isomerasa. Por otro lado la cepa HFRG6 tiene una mutación puntual en HisA que la convierte en una enzima no funcional y por lo tanto en medio mínimo no complementa la actividad de ProFAR isomerasa.

Ensayos de complementación

Los ensayos de complementación de las diferentes librerías en ambas cepas de selección se hicieron en placas de medio mínimo M9 sólido. Primeramente se transformó una alícuota de células electrocompetentes con el ADN de cada librería y se incubaron por una hora a 37 ºC en agitación. Posteriormente cada alícuota fue lavada cuatro veces con medio mínimo M9. Para ello se centrifugaron por 2 minutos a 5000 rpm, se desechó el sobrenadante y se adicionó 1 ml de M9. Se repitió este proceso 4 veces. Al final se resuspendió la pastilla celular en 1 ml de M9 y posteriormente se esparció en las placas de M9 sólido. En los casos indicados, el medio se suplementó

con 50 µM de IPTG y 0-1 µg/ml de triptófano o histidina (según corresponda la actividad a seleccionar). Las placas se incubaron a 30 ºC y se aislaron las colonias nuevas que aparecían cada 24 horas.

Expresión de proteínas en la cepa MC1061*ΔtrpF*

Para analizar el nivel de expresión de las diferentes mutantes de residuos puntuales por alanina se utilizó la cepa MC1061*ΔtrpF*. Cada variante se transformó en células calcio competentes y posteriormente se inoculó con una colonia 5 ml de LB/ampicilina, los cuales fueron crecidos por 12 horas a 30 ºC. Estos cultivos se centrifugaron y la pastilla se resuspendió en 500 µl del amortiguador de sonicación y posteriormente se sonicó a 4 ºC. Se separó la fracción soluble de la insoluble mediante centrifugación a 11,000 rpm por 10 minutos y se cargaron 10 µl del extracto soluble en un gel de SDS-poliacrilamida al 12.5%.

Análisis de la eficiencia funcional

La fracción de variantes que retienen la actividad de PRA isomerasa fue calculada con base al número de variantes que complementaron la auxotrofía del gen de TrpF en la cepa de selección JM101*ΔtrpF* en medio mínimo M9. Para ello, cada librería se transformó por triplicado en células electrocompetentes y posteriormente se lavaron con medio M9 (ver sección de ensayos de complementación). Del cultivo lavado se esparcieron diluciones de aproximadamente 1000-1500 células viables en dos diferente medios de cultivo suplementados con ampicilina (200 µg/ml): LB agar y M9 agar. Las placas se incubaron a 30 ºC por 144 horas y cada 24 horas se contaron las unidades formadoras de colonias (UFCs). La fracción de variantes funcionales se calculó como el cociente del número de UFCs encontradas bajo la presión de selección de la actividad de PRA isomerasa (medio M9) entre el número de UFCs crecidas sin esta presión de selección (medio LB). Este valor fue el primer componente del valor de SFLA (structure-function loop adaptability). Para cada librería se seleccionaron aproximadamente 20 colonias de las placas de M9 y se secuenciaron los respectivos plásmidos.

Análisis de la eficiencia de plegamiento

La fracción de variantes con capacidad de plegarse se estimó fusionando cada librería al gen del reportero de plegamiento de cloranfenicol acetil transferasa (CAT), utilizando el método descrito previamente (*16*). Todas estas librerías, así como los controles positivos y negativos (*16*), se transformaron por triplicado en células electrocompetentes de la cepa MC1061*ΔthiE*. Posteriormente estas células se lavaron con medio M9 (ver sección de ensayos de complementación) y del cultivo lavado se esparcieron diluciones de aproximadamente 1000-1500 células viables en dos diferente medios de cultivo suplementados con ampicilina (200 µg/ml): LB agar con y sin cloranfenicol (20 µg/ml). Las placas se incubaron a 30 ºC por 18 horas y la eficiencia de plegamiento se calculó como el cociente del número de UFCs crecidas en las placas de selección de plegamiento (LB ampicilina/cloranfenicol) entre el número de UFCs crecidas en las placas sin esta presión de selección (LB ampicilina). Este valor fue el segundo componente del valor de SFLA (structure-function loop adaptability).

Purificación de proteína y ensayo de actividad enzimática

Cada pastilla del litro de cultivo de células se resuspendió en 25 ml de amortiguador de fosfato de potasio 10 mM, pH 7.6, 50 mM NaCl, 5% (v/m) de glicerol, 0.1 mM DTT, 0.1 mM PMSF y 2.5 mg de lisozima. Se sonicó a 4 ºC (20 segundos, 6 veces con intervalos de 30 segundos cada vez y un pulso de 50%). Posteriormente se centrifugó a 11000 rpm por 20 minutos a 4 ºC para separar la fracción soluble de la insoluble. La fracción soluble de cada extracto celular se filtró con una membrana de 0.44 µm de diámetro y se vertió a una columna de sefarosa-níquel (HisTrap FF crude 5 ml, GE Healthcare), la cual había sido previamente equilibrada con el amortiguador de fosfato de potasio 50 mM pH 7.6 y 300 mM de NaCl. La proteína etiquetada con el "His-tag" se eluyó aplicando un gradiente linear de 1 mM a 300 mM de imidazol en 10 volúmenes de columna. Las fracciones con la proteína pura se mezclaron y se concentraron hasta obtener un volumen final de 3 ml utilizando el sistema Amicon Ultra-15 (MILLIPORE) a 4 ºC. Posteriormente se cargó esta muestra en una columna de filtración en gel Superdex 200 (GE Healthcare Biosciences), la cual previamente se equilibró con el amortiguador HEPES 50 mM, pH 7.6 y 100 mM de NaCl.

Posteriormente, las fracciones con la proteína pura se mezclaron y se concentraron hasta obtener un volumen final de 3 ml utilizando el sistema Amicon Ultra-15 (MILLIPORE) a 4 ºC. Este volumen final fue dializado 3 veces a 4 ºC contra 1 lt de amortiguador desgasificado de fosfato de potasio 10 mM, pH 7.6, 1 mM EDTA y 1 mM 2-Mercaptoetanol. La concentración de proteína final se calculó usando el método de Bradford (*28*) con el ensayo "Bio-Rad Protein Assay" (BIORAD). La actividad enzimática de PRA isomerasa se determinó usando el protocolo previamente reportado (*29*) y sólo con algunas modificaciones menores (*21*). Cada medida de la actividad enzimática representa el promedio de por lo menos 3 experimentos independientes, utilizando enzima fresca purificada cada vez.

Cinéticas de crecimiento en medio mínimo M9

Para estos ensayos, cada variante se inoculó por cuadruplicado en placas de 96 pozos con medio LB suplementado con ampicilina (200 µg/ml) y posteriormente estas placas se incubaron por 12 horas a 30 ºC. También se preparó una placa de 96 pozos con medio mínimo M9 (48 pozos para la cepa JM101*ΔtrpF* y los otros 48 pozos para la cepa HFRG6). Esta placa se suplementó con la correspondiente cantidad del respectivo aminoácido para cada variante. Cada pozo de medio mínimo se inoculó con 2 µl de la respectiva variante de la placa de 96 pozos de LB ampicilina. Las placas se incubaron a 30 ºC en agitación y se midió la absorbancia ($OD_{600\ nm}$) del cultivo celular cada 4 horas durante 5 días. Con los datos obtenidos de este experimento se graficó la cinética de crecimiento de cada variante.

Construcción de los modelos tridimensionales

Cada modelo se hizo utilizando como molde la estructura de TrpF de *E. coli* (PDB: 1pii). Sobre esta estructura se insertaron las mutaciones puntuales utilizando el programa SWISS-PDB Viewer. Posteriormente se corrió una minimización energética y la estructura resultante fue minimizada pero ahora en presencia del análogo del producto de la reacción de PRA isomerasa (rCDRP) con el servidor del RosettaDesign. Se seleccionó la estructura mejor calificada y como último paso se hizo un análisis de modos vibracionales de baja frecuencia utilizando el servidor de ElNemo. Se seleccionó

el modelo con la mejor calificación y fue minimizado con el programa UCSF Chimera en presencia del análogo del producto de la reacción de PRA isomerasa (rCDRP). Los modelos resultantes son los analizados en el Capítulo 4 del presente trabajo.

6. Resultados y discusión

En el Capítulo 1 se analiza la capacidad de la estructura de la enzima TrpF-LoxP para soportar el intercambio sistemático de diferentes asas β/α, proyecto que fue iniciado en la Tesis de Maestría (*30*) y el cual fue concluido en el inicio del doctorado (*16*). En el Capítulo 2 se presenta el análisis de la relación estructura-función de diferentes asas β/α intercambiadas en el asa 6 de TrpF (*24*). Para este segundo capítulo se desarrollaron nuevas estrategias experimentales tanto de intercambio de asas como de diseño de las conexiones con la enzima molde (bisagras).

En el Capítulo 3 se discute la utilización de dos estrategias para explorar la nueva actividad de ProFAR isomerasa en la enzima TrpF. La primera es mediante el rediseño del sitio activo y el diseño de bisagras y la segunda es mediante el intercambio de asas y el diseño de bisagras.

Capítulo 1. **Efecto en la estructura de TrpF-LoxP del intercambio sistemático de diferentes asas β/α**

En este capítulo se hizo la caracterización estructural de las variantes de la enzima TrpF-LoxP, las cuales tienen modificadas diferentes asas β/α como consecuencia del intercambio sistemático de sus asas catalíticas (SCLE del inglés: Systematic Catalytic Loop Exchange) (*30*). De esta manera se analizó la tolerancia estructural del intercambio sistemático de tres asas β/α de la enzima TrpF-LoxP. El gen de esta enzima tiene insertada la secuencia de un sitio de reconocimiento *loxP* en la posición correspondiente al asa α/β 4, el cual codifica para una inserción de 20 aminoácidos en la enzima TrpF y por ello es un barril $(\beta/\alpha)_8$ modificado no existente en la naturaleza (*31*).

Las asas 2, 4 y 6 de la enzima TrpF-LoxP se reemplazaron por otras asas β/α de diferentes tamaños y secuencias y el porcentaje de plegamiento de cada librería se analizó mediante su fusión con el gen reportero de plegamiento de la enzima cloranfenicol acetil transferasa (CAT). De esta manera si nuestra proteína fusionada es soluble, CAT también es soluble y por lo tanto provee de resistencia contra cloranfenicol a la célula que contiene esta fusión (*30*). La estrategia del SCLE incluyó la introducción de variabilidad a saturación mediante la introducción del codon NNS en las dos últimas posiciones de las hebras β que preceden a cada asa intercambiada y que apuntan hacia el interior del barril $(\beta/\alpha)_8$. Estos sitios se diseñaron con el objetivo de permitir un mejor ajuste estructural de las nuevas asas así como para introducir variabilidad en las áreas importantes para la catálisis. Además, el diseño del SCLE permite el intercambio sistemático de asas tanto a la misma posición estructural de la cual provienen como a posiciones diferentes (*30*).

El sistema de selección de plegamiento *in vivo* se validó con diferentes construcciones como controles, las cuales fueron probadas en diferentes cepas de *E. coli*. Este análisis mostró que la única cepa que no reporta falsos positivos es la MC1061Δ*thiE* y que además la concentración óptima de cloranfenicol para discernir entre proteínas plegadas y desplegadas es de 20 μg/ml, lo cual permite un nivel de selección de plegamiento y a la par una alta viabilidad celular. Por lo tanto, todas

nuestras librerías fusionadas a CAT se evaluaron utilizando estas condiciones de presión de selección de plegamiento. Los resultados mostraron que entre el 30% y 90% de las variantes generadas en las diferentes librerías están plegadas. Para confirmar con mayor precisión la eficiencia de este método en la selección de variantes plegadas se hizo un análisis de tipo "Western blot" del extracto celular soluble de 72 variantes seleccionadas por crecer en presencia de cloranfenicol. Los resultados mostraron que todas las fusiones son solubles, lo cual confirma la eficiencia de nuestro método para seleccionar proteínas plegadas (*30*).

Los espectros obtenidos por dicroísmo circular en el UV lejano, así como el cambio en los espectros de la emisión de fluorescencia de algunas variantes sin la fusión con CAT fueron muy similares con los de la enzima TrpF-LoxP, demostrando que las interacciones globales de la estructura se mantienen aún después de intercambiarle asas de diferentes posiciones y de quitarle la fusión con CAT (*16*). Además, su alto nivel de expresión en la fracción celular soluble sugiere que las variantes están correctamente plegadas y que los pequeños cambios observados en los espectros de CD podrían reflejar sólo algunos cambios estructurales a nivel local (*16*). Los resultados en extenso de este capítulo se encuentran en el manuscrito de la página siguiente (*16*).

Con base en los resultados obtenidos al aplicar el método del SCLE al estudio de la estructura de la enzima TrpF-LoxP, concluimos dos aspectos importantes para la continuación del proyecto de doctorado: 1) es posible intercambiar sistemáticamente las asas β/α 2, 4 y 6 de TrpF-LoxP sin afectar su estabilidad, a pesar de que esta enzima es un barril $(\beta/\alpha)_8$ modificado, y 2) las dos posiciones de las hebras β que se mutaron a saturación son importantes para seleccionar proteínas plegadas pero sin función. Basados en estas conclusiones, en el Capítulo 2 del presente trabajo decidimos analizar la relación estructura-función de las asas β/α utilizando como modelo a la enzima original de TrpF. Además diseñamos una nueva estrategia de introducción de variabilidad en los residuos bisagra, los cuales son ahora los dos residuos finales de cada asa intercambiada.

doi:10.1016/j.jmb.2009.02.022 J. Mol. Biol. (2009) 387, 949–964

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Protein Design through Systematic Catalytic Loop Exchange in the $(\beta/\alpha)_8$ Fold

Adrián Ochoa-Leyva, Xavier Soberón*, Filiberto Sánchez, Martha Argüello, Gabriela Montero-Morán and Gloria Saab-Rincón*

Departamento de Ingeniería Celular y Biocatálisis, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Apartado Postal 510-3, Cuernavaca, Morelos 62271, México

Received 18 October 2008; received in revised form 2 February 2009; accepted 10 February 2009
Available online 20 February 2009

Protein engineering by directed evolution has proven effective in achieving various functional modifications, but the well-established protocols for the introduction of variability, typically limited to random point mutations, seriously restrict the scope of the approach. In an attempt to overcome this limitation, we sought to explore variant libraries with richer diversity at regions recognized as functionally important through an exchange of natural components, thus combining design with combinatorial diversity. With this approach, we expected to maintain interactions important for protein stability while directing the introduction of variability to areas important for catalysis.

Our strategy consisted in loop exchange over a $(\beta/\alpha)_8$ fold. Phosphoribosylanthranilate isomerase was chosen as scaffold, and we investigated its tolerance to loop exchange by fusing variant libraries to the chloramphenicol acetyl transferase coding gene as an *in vivo* folding reporter. We replaced loops 2, 4, and 6 of phosphoribosylanthranilate isomerase with loops of varied types and sizes from enzymes sharing the same fold.

To allow for a better structural fit, saturation mutagenesis was adopted at two amino acid positions preceding the exchanged loop. Our results showed that 30% to 90% of the generated mutants in the different libraries were folded. Some variants were selected for further characterization after removal of chloramphenicol acetyl transferase gene, and their stability was studied by circular dichroism and fluorescence spectroscopy. The sequences of 545 clones show that the introduction of variability at "hinges" connecting the loops with the scaffold exhibited a noticeable effect on the appearance of folded proteins. Also, we observed that each position accepted foreign loops of different sizes and sequences.

We believe our work provides the basis of a general method of exchanging variably sized loops within the $(\beta/\alpha)_8$ fold, affording a novel starting point for the screening of novel activities as well as modest diversions from an original activity.

Edited by C. R. Matthews

Keywords: loop grafting; PRAI; directed evolution; CAT; folding reporters

**Corresponding authors.* E-mail addresses: soberon@ibt.unam.mx; gsaab@ibt.unam.mx.

Present address: G. Montero-Morán, Department of Nutritional Sciences, Rutgers, The State University of New Jersey, New Brunswick, NJ 08901, USA.

Abbreviations used: PRAI, phosphoribosylanthranilate isomerase; ePRAI, monofunctional version of E. coli PRAI; ADA, adenosine deaminase; FBPA, fructose-bisphosphate aldolase; PBGS, porphobilinogen synthase; DHDPS, dihydrodipicolinate synthase; αTS, tryptophan synthase α chain; TPS, thiamine phosphate synthase; Ure, urease; MR, mandelate racemase; dsDNA, double-stranded DNA; CAT, chloramphenicol acetyl transferase; Cm, chloramphenicol; WT, wild type; EDTA, ethylenediaminetetraacetic acid; BME, β-mercaptoethanol.

Introduction

It is recognized that the evolution of proteins has proceeded through processes where gene duplication and point mutations have played a major role.[1] This is especially true for the relatively recent evolutionary times, but this simple process is incapable of explaining the first appearance of primordial functional proteins and many of the more drastic modifications apparent in extant protein families and superfamilies. These additional modifications suggest processes involving modular exchange between segments of unrelated proteins.[2–4]

Understanding these natural processes is a very valuable asset to the engineering of proteins, in particular when directed evolution techniques are employed.[5] For the study of the most fundamental questions, it is advantageous to rely on suitable protein architectures, several of which have become model systems due to their versatility and/or potentially wider relevance. Proteins such as those possessing the immunoglobulin fold[6] or the $(\beta/\alpha)_8$ fold[7] can be viewed as "scaffolds" for engineering due to their apparently modular architecture and evolutionarily variable loops. Combining protein design and directed evolution with scaffold-based protein libraries provides an excellent route to engineering new protein functions. The utility of such libraries depends on how tolerant the scaffolds are to randomization, because the selected variants must remain folded and soluble. The $(\beta/\alpha)_8$ or TIM barrel scaffold should be an ideal starting point for engineering novel enzymatic activities by rational design and directed evolution.

TIM barrel proteins are the most common fold among protein catalysts, constituting approximately 10% of all known enzyme structures. They are catalytically versatile: five of the six enzyme classes contain members with this fold.[3] The barrel structure is composed of eight parallel β-strands in the interior of the protein, surrounded by eight α-helices. The pseudo symmetry of the fold is simultaneously an attractive feature for engineering and a suggestion of the modular construction and evolution of this protein architecture.[8,9] The residues at the active sites of all known TIM barrel enzymes are located on the catalytic face of the barrel, which is composed of the C-terminal end of the β-strands and residues coming from the β/α loops (the loops that link β-strands with α-helices). In contrast, the remainder of the fold, including the opposite face of the barrel, is important for conformational stability.[10] It has long been hypothesized that varying the residues of the active site, including the loops, might alter enzymatic function without affecting the stability of the fold.[11–14] Indeed, it has been shown that engineering of protein loops can afford functional changes, notably substrate specificity. This has been demonstrated for proteins with TIM barrel[15] and other folds.[16–18] Furthermore, recent work demonstrated that more profound changes in the enzymatic function are possible through loop swapping as a component of the directed evolution strategy.[19–21]

On the other hand, the design of new enzymes by generating variability through point mutation of an existing gene has shown modest progress.[22] As stated above, for the natural evolution of proteins, this is likely because more drastic changes are needed. Following a related idea, it has been suggested that new enzymes can be generated by recombining different segments of modern proteins, and the concept has been shown to be successful.[4,23]

Here we describe Systematic Catalytic Loop Exchange (SCLE), a new strategy for exchanging loops on a TIM barrel fold. As a scaffold, we chose a variant of phosphoribosylanthranilate isomerase (PRAI-LoxP) from *Escherichia coli*, a monomeric enzyme where we have previously done significant work.[24,25] This gene has an insertion coding for a cre-lox recognition site in the loop linking α-helix 4 and β-strand 5, as previously described.[24] The purpose of using this gene is to carry out *in vivo* recombination of generated variants in future experiments. This gene will be referred to as *trpF*-loxP and the respective protein PRAI-LoxP throughout the rest of the article. We set out to recombine loops compatible with this scaffold that have already been explored and optimized (albeit in different contexts) by natural evolution, hoping that the resulting libraries preserve a high enough percentage of folded proteins to be screened for function. In order to design specific modification sites, we investigated the tolerance of PRAI-LoxP to loop exchange, as estimated by stable folding *in vivo*. We decided to initially generate variability toward 3 of the 8 (β/α) loops, where there is higher participation of their residues in substrate binding and catalysis.[3] In this communication, we describe the exchange of 14 different (β/α) loops within the PRAI-LoxP structure. The exchanged loops belong to eight different proteins sharing the same fold but have diverse functions. Functional diversity will be explored in the future after *in vivo* recombination of the different loop libraries combined with random mutagenesis and additional designed mutations.

Results

Design of the loop exchange in PRAI

As an ultimate goal, the present work aims at contributing to the generation of *de novo* protein activities. The scaffold with the cre-lox insertion is derived from a modified monofunctional version of *E. coli* PRAI (WT ePRAI), in which part of the gene coding for the bifunctional IGPS-PRAI protein has been excised to express the PRAI gene separately.[26] The three-dimensional structure of the *E. coli* enzyme [Protein Data Bank (PDB) entry 1PII] was used to identify the sites for loop exchange. Our exchange strategy focused specifically on loops 2, 4, and 6, where residues identified as binding and catalytic sites are more abundant in a set of TIM barrel proteins analyzed.[3] The loop-donor proteins were selected under the following criteria: they should

Table 1. Characteristics of the exchanged loops in the PRAI-LoxP scaffold

Enzyme (EC number)	Position of the exchanged loop[a]	Sequence of the loop	Lengths
MR (EC 5.1.2.2)	β/α loop 2	GYPAL	5
FBPA (EC 4.1.2.13)	β/α loop 2	SNGGASFIAGKGVKSDVPQ	19
Ure (EC 3.5.1.5)	β/α loop 2	GGTGPAAGTHATTCTPG	17
PRAI WT (5.3.1.24)	β/α loop 2	VATSPRCVN[b]	9
DHDPS (EC 4.2.1.52)	β/α loop 4	PYYNRPS	7
TPS (EC 2.5.1.3)	β/α loop 4	LGQEDLH	7
MR (EC 5.1.2.2)	β/α loop 4	EPTLEHD	7
FBPA (EC 4.1.2.13)	β/α loop 4	DLSEES	6
αTS (EC 4.2.1.20)	β/α loop 4	DVPVQQS	7
Ure (EC 3.5.1.5)	β/α loop 4	EDWGAT	6
ADA (EC 3.5.4.4)	β/α loop 4	GDELGFPGSLF	11
PBGS (EC 4.2.1.24)	β/α loop 4	AAMDG	5
PRAI WT (5.3.1.24)	β/α loop 4	GNEE[b]	4
DHDPS (EC 4.2.1.52)	β/α loop 6	TGNL	4
PBGS (EC 4.2.1.24)	β/α loop 6	PAGAY	5
αTS (EC 4.2.1.20)	β/α loop 6	SRAGVTGAENRAALP	15
PRAI WT (5.3.1.24)	β/α loop 6	NGQGGSGQRFD[b]	11

[a] Structural position in which the WT loop was replaced for the new loop in the PRAI-LoxP scaffold.
[b] Sequence corresponding to the WT loops from PRAI-LoxP.

be TIM barrels, they should comprise diverse functionalities and substrate kinds, and they should preferably have a monomeric state; if this last condition was not met, their oligomerization interface should not involve the catalytic site (i.e., the top of the barrel). Based on these criteria, the selected loop donors were the proteins adenosine deaminase (ADA; PDB code 2ada), fructose-bisphosphate aldolase (FBPA; PDB code 1dos), porphobilinogen synthase (PBGS; PDB code 1l6s), dihydrodipicolinate synthase (DHDPS; PDB code 1dhp), tryptophan synthase α chain (αTS; PDB code 1xc4), thiamine phosphate synthase (TPS; PDB code 1xi3), urease (Ure), and mandelate racemase (MR) from *E. coli*. The structures from Ure and MR were modeled from the homologous proteins from *Klebsiella aerogenes* (PDB code 2kau) and *Pseudomonas putida* (PDB code 1mdr), respectively. The amino acid sequences of loops 2, 4, and 6 of these proteins are shown in Table 1. The loops were inserted at the zone delimited by the last position of the previous β-strand and the first residue of the next α-helix, as shown in Fig. 1a. For loop 2, the exchanged fragment is located between residues Val31 and Asn39; for loop 4, between residues Gly82 and Glu85; and finally, for loop 6, between Asn127 and Arg135 (numbering according to gene reported by Kirschner *et al.*[26]). In order to have a general strategy to exchange loops occupying topological positions different from their original site in the donor protein, we designed "general" double-

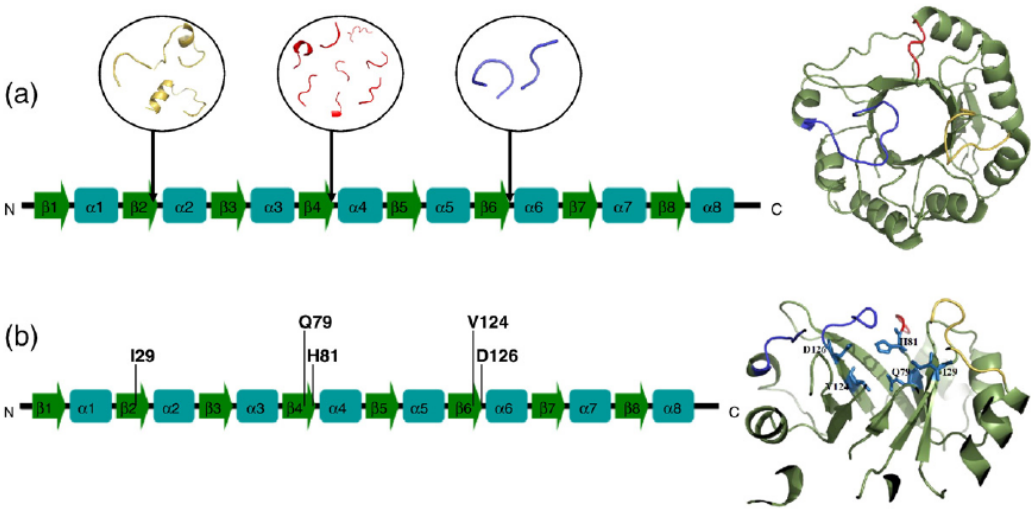

Fig. 1. Secondary-structure elements and tertiary structure (PDB code 1pii) of WT PRAI (ribbon diagram showing a view from the top of the central β-barrel). (a) The loops 2, 4, and 6 are shown in yellow, red, and blue, respectively. The circular insets show the loops to exchange in the PRAI-LoxP scaffold at the three structural positions (β/α 2, β/α 4, and β/α 6). (b) The mutated positions are indicated and a side-view slice of the TIM barrel structure of the PRAI scaffold is shown on the right.

stranded DNA (dsDNA) connectors that would direct the insertion to any of the particular locations explored in PRAI-LoxP (i.e., loop 2, loop 4, or loop 6). There were two sets of dsDNA connectors, the NH2 connector that coded for β2, β4, and β6 (on the side 5′ from the loop) and the CO connector that coded for α2, α4, and α6 (on the side 3′ from the loop). These connectors had different single base overhangs at both ends to allow directional ligation of the incoming dsDNA coding for the loop with the resulting replacement of the original loop present in PRAI, as shown in Fig. 2a. A schematic representation of the strategy used to exchange the loops in the gene of PRAI-LoxP is shown in Fig. 2a and b.

A change of function would likely require, in addition to the loop transplant, the substitution of the last residues in the β-strands, as these are also typically part of the binding and catalytic sites.[27] Therefore, another component in the design of the NH2 connectors was the introduction of variability in the final positions of the β-strand preceding the loop to be exchanged and pointing inward into the barrel (where catalysis takes place), as shown in Fig. 1b. In the case of the NH2 connector for loop 2, the last residue in the β-strand points toward the external shell; therefore, variability was introduced in the previous position, Ile29. In the cases of loops 4 and 6, the residues preceding the loops, His81 and Asp126, respectively, point toward the interior of the barrel; therefore, we decided to introduce variability in these positions as well as in the previous residues also pointing to the interior of the barrel, that is,

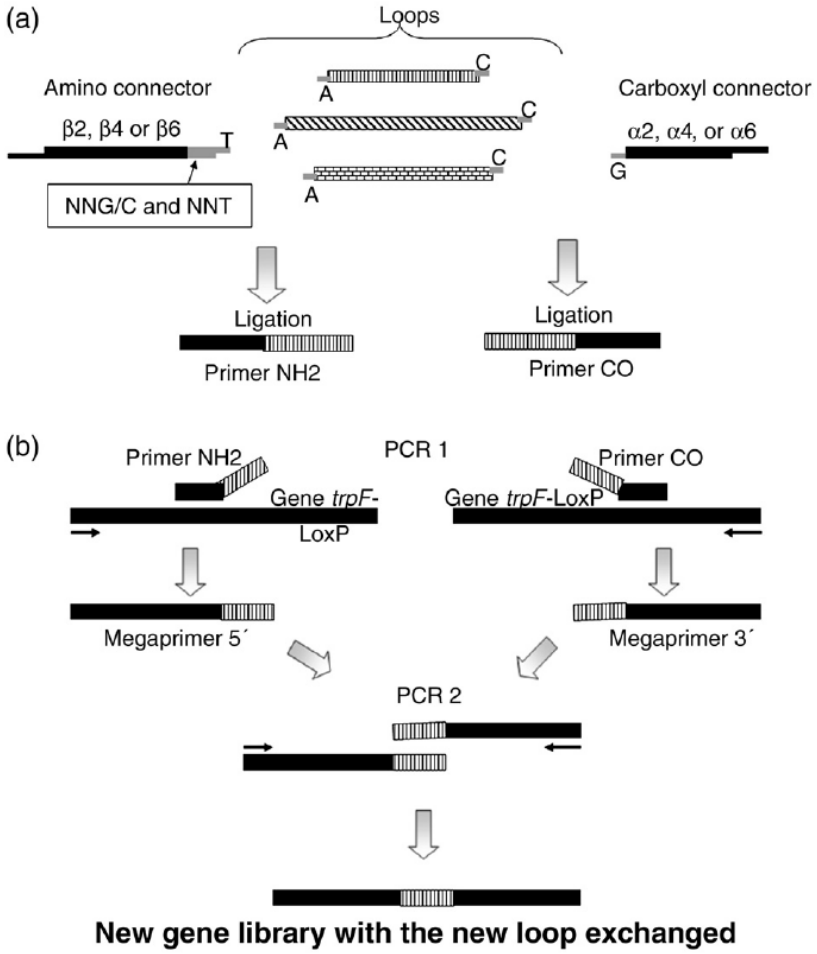

Fig. 2. Strategy for systematic loop exchange in the trpF-loxP gene. (a) dsDNA coding for each loop is ligated with the amino and the carboxyl connector independently, yielding two primers. The amino connectors (that prime to either β2, β4, or β6 coding sequences), the carboxyl connectors (prime to either α2, α4, or α6 coding sequences) and the "loops" (dsDNA coding for the diverse loops) all had different single base overhang at both ends to allow directional ligation. The mutated sites shown in the box are encoded in the oligonucleotides that constitute the amino connectors. (b) The resulting primers from the ligation described in (a) are used in combination with the corresponding end primers (arrowed lines) to amplify both the 5′ and the 3′part of the gene, respectively, by PCR using the *trpF*-loxP gene as template. The only common sequence between the two obtained megaprimers is the coding sequence for the exchanged loop; these are used in a second-round PCR to obtain the final gene with exchanged loop sequence by PCR primer extension.

Gln79 and Val124, respectively. The degenerate codon NNG/C was used when no other restrictions were present, allowing exploration of all 20 amino acids; however, the requirement of fixing the last base to T in the last codon of the NH2 connector restricts the variability of this position, preventing the exploration of residues Met, Gln, Lys, Glu, and Trp. Similarly, the CO connectors had a G base overhanging in the noncoding strand, designed to anneal with the overhanging C in the coding strand of the guest loop. This made unavoidable the introduction of a size conservative mutation from Gln to His in one of the exchanged loops, specifically loop 2 from FBPA from *E. coli*.

Folding selection method

The selection of folded proteins through the use of folding reporter genes has been previously documented.[28,29] Briefly, in the system we chose, the gene or family of genes to be screened for folding is fused to a reporter gene, in this case the gene that codes for chloramphenicol acetyl transferase (CAT) and provides resistance to the antibiotic chloramphenicol (Cm). A protein capable of folding will allow the correct folding of CAT and, therefore, the resultant clones will be resistant to Cm, while no growth will be observed for those clones bearing a folding-deficient fusion gene product.

To validate the selection method, we made a construction of the wild-type (WT) PRAI gene fused to CAT gene as a positive control and, as negative controls, two proteins proven to have poor folding: a circular permutation of PRAI previously described[30] and an engineered antibody provided by Dr. J. Osuna (unpublished result). Additionally, a construction of WT PRAI with two stop codons between PRAI and CAT was also engineered. The resultant plasmids were transformed in four *E. coli* strains, XL1-Blue, JM101Δ*trpF*, JMB9, and MC1061Δ*thiE* and plated on Luria broth (LB) medium plates containing ampicillin (200 μg/ml) and Cm at concentrations ranging from 0 to 100 μg/ml. We found that JM101, XL1-Blue, and JMB9 strains show lower resistance to Cm. Additionally, as also noticed by others,[31] the JM101 and XL1-Blue gave many false-positive results. In contrast, selection with the strain MC1061Δ*thiE* gave consistent results. None of the negative controls, including the fusion of WT PRAI with two stop codons, showed growth at Cm concentrations above 15 μg/ml (Table 2). The positive control, on the other hand, showed a higher proportion of colonies under selective conditions. In this case, we observed that Cm concentration as low as 20 μg/ml is enough to discriminate unfolded variants while maintaining high viability of cells bearing viable genes.

Assessment of folding in selected clones

To further confirm the accuracy of positive clone scorings, Western blot analysis of the cell extract of 8 clones selected in Cm plates from each of nine libraries was carried out. As observed in Fig. 3, of 72 clones selected, all but one expressed variable but clearly detectable amounts of fusion protein in the soluble fraction. This demonstrates that the method is reliable for the identification of true positive folded proteins.

Ratio of folded proteins in the generated libraries

Using the conditions described in the folding selection method section, we evaluated the fraction of folded proteins in each loop-exchanged library. Figure 4 shows the percentage of fusion proteins conferring Cm resistance in libraries generated on loop 2, loop 4, and loop 6, respectively. We then took the survival ratio of a population of cells transformed with the PRAI-LoxP-CAT fusion as our higher limit and normalized the survival rate observed for the different libraries with this value. The percentage of folded proteins for the three different libraries generated in loop 2 had a very similar value, around 30–40%, and in fact they were the lowest values found in the whole set of constructions. Exchanges at loop 4, where more libraries were generated, afforded a higher viability, from 40% to approximately 70% of folded peptides, depending on the incoming guest loop. Finally, in the case of loop 6, for which only three libraries were constructed, we observed the highest viability, ranging from 50% to 90%.

In order to further investigate the properties of these clones, 10 colonies were isolated from the plate

Table 2. Resistance toward Cm conferred by different CAT fusions when expressed in MC1061Δ*thiE*

Variant	Description[a]	Concentration of Cm (μg/ml)[b]				
		0	5	10	15	20
pDan5	No insert	+++	—	—	—	—
Vh-CAT	Antibody fused to CAT	+++	+++	+	—	—
Per3-β4/α3-CAT	Permutation of PRAI fused to CAT	+++	+++	+	—	—
PRAI-Stop-CAT	PRAI WT with the insertion of two stop codons between their and CAT fusion.	+++	—	—	—	—
PRAI-CAT	PRAI WT fused to CAT	+++	+++	+++	+++	+++

Resistance was estimated by the growth rate.

[a] Description of the genes from the different variants.

[b] Cells were grown in LB medium plates supplemented with ampicillin and with different concentrations of Cm. The number of colonies appearing in the presence of Cm was counted after 18 h and compared with that without Cm. The percentages of colonies that survived to different concentrations of Cm relative to the control without it are as follows: +++, 100%; ++, 60%; +, ≤10%.

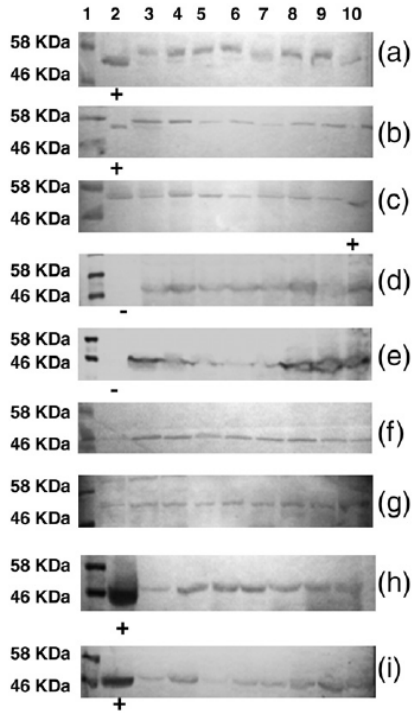

Fig. 3. Western blot analysis from soluble extract with an antibody rose against CAT protein. The expected molecular mass of the proteins fused to CAT is approximately 50 kDa. Lane 1 shows a molecular mass marker, lane 2 or lane 10 shows either a positive or negative control indicated by a + or − symbol under the lane, according to the case. Lanes 2–10 show the soluble fraction of eight independent colonies chosen from LB/ampicillin/Cm plates from (a) loop 2 FBPA, (b) loop 6 αTS, (c) loop 4 Ure, (d) loop 4 MR, (e) loop 6 DHDPS, (f) loop 2 MR, (g) loop 4 αTS, (h) loop 4 DHDPS, and (i) loop 4 FBPA libraries.

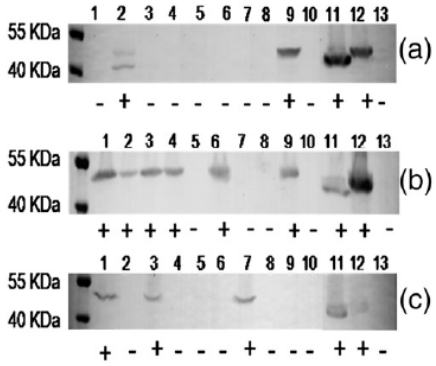

Fig. 5. Western blot analysis from soluble extract from colonies isolated without selective pressure using an antibody rose against CAT protein. The expected molecular mass of the proteins fused to CAT is approximately 50 kDa. 1–10, Western blots of the soluble fractions of 10 clones randomly chosen from the LB/ampicillin plates from each loop library. 11, WT PRAI; 12, clone selected in presence of Cm (20 μg/ml); 13, plasmid pDAN5. The isolated plasmid DNAs from analyzed clones were retransformed in the strain MC1061 Δ*thiE* and plated on LB supplemented with ampicillin and Cm. Those that conferred Cm resistance are shown as (+) and the ones that did not are shown as (−). (a) Loop 2 library from MR. (b) Loop 4 library from Ure. (c) Loop 6 library from PBGS.

without selection of representative libraries of each loop comprising the broad range of survival rates observed, that is, loop 2 from MR, loop 4 from Ure, and loop 6 from PBGS. Figure 5a–c shows the Western blot of the cell extract soluble fraction of the selected clones with an antibody rose against CAT. It is important to note that the results of percentage of clones showing the presence of fusion protein product in the gel summarized in Table 3 are in good agreement with those found by Cm resistance *in vivo*. Plasmid DNA was retransformed in MC1061 Δ*thiE E. coli* competent cells. There is an invariable correspondence between growth in Cm and the presence of fusion protein in the Western experiment, indicating that Cm resistance is an adequate criterion to evaluate the ratio of folded protein variants (Fig. 5a–c). All these plasmids were sequenced and, as expected, none of the sequences found in clones reported as nonfolded by Western blot analysis are present in the clones reported as folded by the genetic selection method.

Sequence analysis

Sequence analysis of a few clones (about 20 from each selective and nonselective condition) was carried out from each library to verify if the insertions were introduced as planned, as well as to find out if

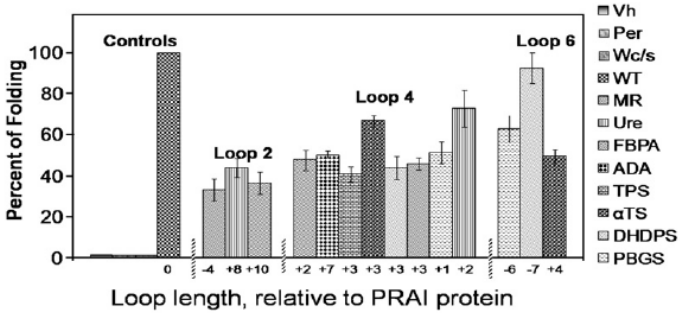

Fig. 4. Percentage of folded proteins fused to CAT was calculated as described in Materials and Methods. Values and error bars represent the average and standard deviation for each library analyzed. Controls are designated as follows: Vh, antibody; Per, circular permutation of PRAI; Wc/s, PRAI-LoxP with stop codons before the fusion to CAT; WT, PRAI-LoxP. The differences in length of loops of the libraries, relative to the PRAI WT protein, are shown in the *x*-axis.

Table 3. Percentage of folding for loop libraries by Western blot analysis

Library[a]	Percentage of mutants positive to fusion to CAT[b]
Loop 2 of MR	28
Loop 4 of Ure	60
Loop 6 of PBGS	43

[a] Name of the library analyzed by Western blot.
[b] The percentage of folding was calculated by dividing the number of clones that present fusion with CAT by the total of clones analyzed. This value was normalized by the sequence found as WT or with an aberrant construction in the 10 clones.

there were amino acid preferences at the end of the β-strand preceding the inserted loops in the folded proteins. From the sequence analysis of 545 clones selected and nonselected (354,250 pb), we conclude that between 70% and 100% of the generated mutants were well constructed, demonstrating the correct loop replacement and the presence of variability in mutagenized positions in the different libraries.

Sequence analysis of clones under selective pressure

There is a true biased representation of some amino acids as a result of design. After correction by normalizing the observed frequencies of each amino acid in selected *versus* nonselected clones for each library, we still observe overrepresentation of some amino acids in selected clones. However, the size of the sample is very small to draw conclusions.

Libraries of loop 2

Some amino acids at the variable positions had either a significantly higher or lower frequency than expected from chance. Some loops show a strong bias for a specific amino acid, as is the case for the insertion of loop 2 from Ure, which shows a positive selection for both serine and alanine at position 29, as well as a negative selection for threonine. Loop 2 from FBPA, on the other hand, showed a strong positive selection for threonine and an equally strong negative selection for leucine at that position. Finally, the exchange of loop 2 with the library representing the loop 2 from MR had its highest preference for arginine and a negative selection against tryptophan.

Libraries of loop 4

The same analysis was performed for the two randomized residues upon replacement of loop 4 in PRAI-LoxP. The insertion of loop 4 from TPS shows a strong preference for arginine at position 79, while position 81 shows negative selection against proline. The replacements of the rest of loops 4 do not display significant preference for any particular residue at position 79, with the exception of loops from MR and DHDPS, which show negative selection against isoleucine and glycine, respectively. Position 81 on the other hand, shows a small but significant preference for serine and threonine when the loop from PBGS is introduced and for serine when the guest loop is from MR. Interestingly, serine is counterselected when the inserted loops 4 derive from the αTS and from DHDPS. Negative selection was also observed against glycine when loop 4 was replaced by that from ADA. These results indicate that some residues preceding the exchanged sequence are more compatible with the new loop than with others. Perhaps more important was the observation of some correlated preferences. With the size of the sample analyzed, the probability of finding the same pair of residues twice is quite small. In five of the eight libraries with loop 4 exchanged, clones with repeated pairs of residues were found among the approximately 20 selected clones that were sequenced. For instance, the pairs S–T, D–T, and N–P at positions 79–81 were found twice when loop 4 from Ure replaced the PRAI loop. In the library with replacement of loop 4 from PBGS, the combination H–V was selected twice, while in that from ADA, the pair S–R was found. The pair R–V was selected in the library from the DHDPS loop and the loop from MR showed two repeated pairs, T–V and P–N. These combinations were not present in the sample of randomly chosen clones, so the possibility of overrepresentation of these amino acids in the library is unlikely. When loop 4 of αTS, of TPS, and of FBPA are exchanged, there is no significant preference for specific combinations at positions 79 and 81.

Libraries of loop 6

In the three libraries constructed, there were some significant selections at position 124. In the case of the PBGS loop, leucine was selected, while lysine was counterselected; in the case of the guest loop from DHDPS, there was a small but significant preference for cysteine, while for the guest loop from the α subunit of αTS, there was a preference for threonine. In both DHDPS and αTS guest loops, there was a negative selection for serine at position 124. On the other hand, no preference was evident for any residue at position 126, but counterselection was observed for serine in the case of the DHDPS guest loop and for threonine in the case of the αTS guest loop. Just as observed in the libraries from loop 4, there were combinations at the two positions that were selected twice. The combination T–C in both PBGS and DHDPS libraries and the combination T–R in the αTS library were observed twice in folded clones.

Structural CD and fluorescence spectroscopic analysis of some selected clones

Considering the drastic changes introduced in the PRAI-LoxP sequence, the relatively high percentages of folded proteins that were obtained according to the Cm resistance test are somewhat surprising. To find out if some selected mutants were folded and stable in the absence of CAT, we further investigated

their structural properties. Figure 6 shows the far UV–CD spectra of the purified proteins overlapped with the spectrum of the PRAI-LoxP (the parental gene product used for the constructions).[24] Some subtle changes in the spectra could be observed, suggesting some structural differences. The secondary-structure predictions based on the CD spectra were carried out. Albeit these results have to be taken with reservation, it is interesting to note that loop 4 from the αTS showed a calculated content of secondary-structure elements closest to that of PRAI-LoxP protein (14% α-helix, 38% β-strand), while variants with loop 2 and loop 6 exchanged show an increase in the α-helical content (28%). Loop 2 from FBPA forms a helical structure in the context of its parental protein. The increase in signals at 208 and 222 nm suggests that this loop may also be helical in the context of the PRAI-LoxP scaffold. The increase in helical structure upon substitution of loop 6 by the corresponding loop from PBGS is harder to explain, since in this case it is a small loop of only five residues compared to the 11 residues in loop 6 from PRAI-LoxP. However, the shortening of this loop may have contributed to the propagation or better formation of α-helix 6, which shows only one turn in the crystal structure of the WT protein[32] and/or in the configuration of the adjacent α-helix 5, which is not a well-structured helix in the native protein. Nonetheless, it is important to highlight that all the spectra are characteristic of folded proteins. Figure 7 shows the fluorescence emission spectra of the selected variants under native and denaturing conditions. The blue shift observed under native conditions for all the proteins is characteristic of folded proteins, and their red shift in urea indicates the exposure of the fluorophores to the aqueous environment upon loss of tertiary structure. The fluorescence spectrum of WT PRAI was recorded for comparison (Fig. 7a). The center of mass of the spectra under denaturing conditions is the same for all the proteins (around 366), while under native conditions all the PRAI-LoxP mutants show a blue shift higher than that of the WT ePRAI protein

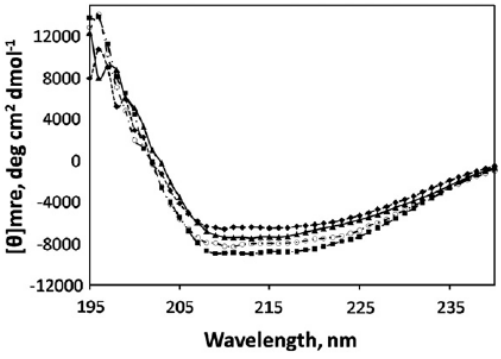

Fig. 6. Far-UV CD spectra of representative constructs engineered in this study. Triangles, PRAI-LoxP; open circles, Mut_L2_FBPA; diamonds, Mut_L4_αTS; squares, Mut_L6_PBGS.

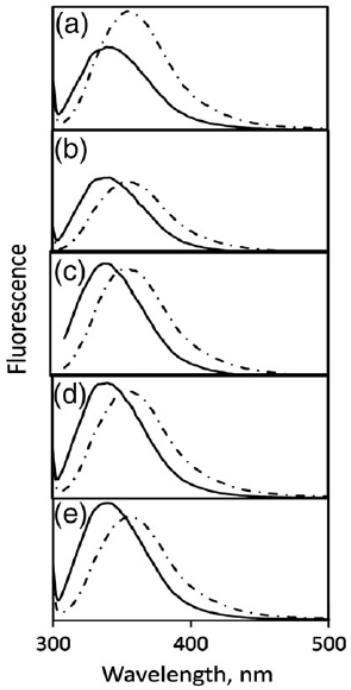

Fig. 7. Fluorescence emission spectra of representative constructs engineered in this study under native conditions (continuous line) and at 9 M urea (dashed line). (a) WT ePRAI, (b) PRAI-LoxP, (c) Mutant_L2_FBPA, (d) Mutant_L4_αTS, and (e) Mutant_L6_PBGS.

(around 349 nm *versus* 352 nm for the WT enzyme). Interestingly, the WT PRAI spectrum shows an increment of fluorescence upon unfolding, while the rest of the variants show a slight decrease in fluorescence. This is indicative of less fluorescence quenching among the fluorophores in the folded state of the variants, but the same behavior was observed for the parental PRAI-LoxP protein, indicating that this difference is related to the introduction of the lox-P loop at the bottom of the barrel and not to the exchanged loops at the catalytic face.

Figure 8 shows the loss of secondary structure of the variants in 8 M urea. It is noteworthy that there is some residual secondary structure in the proteins after incubation for 2 h at denaturing conditions. This is in agreement with the high stability of the secondary structure observed for the proteins subjected to thermal unfolding (data not shown). This behavior was shared by all the variants, including the parental PRAI-LoxP protein. Therefore, we attributed it to the presence of the insertion coded by the cre-lox recognition sequence.

Oligomeric state of the mutants

The oligomeric state of the mutants was investigated by size-exclusion chromatography using three different initial protein concentrations ranging

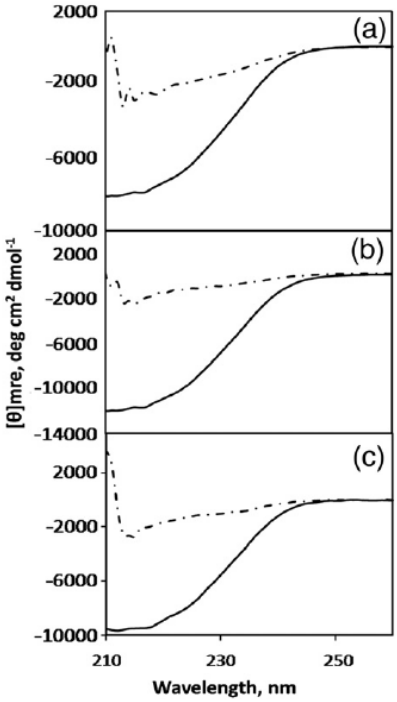

Fig. 8. Far-UV CD spectra of representative constructs engineered in this study under native conditions (continuous line) and at 9 M urea (dashed line). (a) Mutant_L2_FBPA, (b) Mutant_L4_ αTS, and (c) Mutant L6 PBGS.

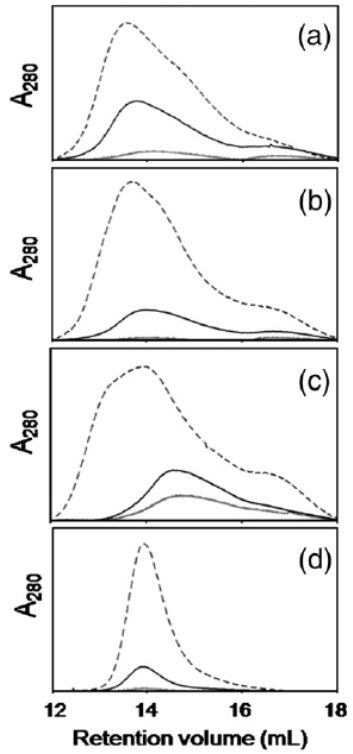

Fig. 9. Analytical gel-filtration chromatograms of PRAI-LoxP and its mutants using a Superose HR12 column. The initial protein concentrations were 2 μM (thin lines), 20 μM (thick lines), and 100 μM (broken lines). (a) PRAI-LoxP, (b) Mutant_L2_FBPA, (c) Mutant_L4_αTS, and (d) Mutant_L6_PBGS.

from 2 to 100 μM. Table 4 shows the apparent molecular masses of the variants determined by their elution profile. As shown in Fig. 9, all the proteins, including the PRAI-LoxP parental protein, are present as a mixture of monomer–dimer as judged by their elution profile. This behavior has also been observed for the WT ePRAI.[33] The variant with loop 6 from PBGS shows dimer as the dominant species (Fig. 9d). In the other cases, the dimer seems to be in a fast equilibrium with higher oligomers when the protein concentration increases to 100 μM, as suggested by the increase in apparent molecular mass and the deformation of the peaks, especially for the variant with loop 4 from αTS (Fig. 9c).

Discussion

In this study, the PRAI-LoxP protein was chosen as a scaffold and its tolerance to loop exchange was investigated by fusing the libraries to the CAT coding gene as an *in vivo* folding reporter. When such fusion proteins are soluble, they provide

Table 4. Apparent molecular masses and deduced association states

Protein[a]	Calculated molecular mass (kDa)[b]	Apparent molecular mass (kDa)[c]			Deduced association state
		2.0 μM	20 μM	100 μM	
PRAI-LoxP	24.1	22.9+43.3	22.9+47.6	49.8	Monomer−dimer
Mut_L2_FBPA	25.0	23.8+46.5	23.9+46.7	24.3+49.7	Monomer−dimer
Mut_L4_αTS	24.3	40	40	23.9+46.4	Dimer−monomer
Mut_L6_PBGS	23.7	46.2	46.6	46.4	Dimer

[a] The proteins were dissolved in 20 mM potassium phosphate buffer (pH 7.6) and 300 mM KCl. Each protein carries a His-tag at its N-terminus and the protein concentrations were 2, 20, and 100 μM.
[b] The calculated molecular mass was theoretically calculated from the sequences of the respective proteins.
[c] Apparent molecular masses were calculated using the protein elution volumes and a calibration curve generated using elution volumes for proteins of known molecular mass.

resistance to Cm to their host cells. Our design allows the exchange of the loops in the same position or in other topological positions from the original protein in PRAI. As an ultimate goal, the present work aims at contributing to the generation of *de novo* protein activities by applying selection regimes to the folded proteins from the loop libraries. This approach allows the exploration of a more divergent sequence space and is not limited to point mutations. Also, it enables the introduction of variability to areas important for catalysis while maintaining a higher probability for a folded protein.

The behavior of the controls for folding selection allowed us to evaluate different *E. coli* strains for the selection of true positive clones and the identification of the proper Cm concentration to discriminate the unfolded variants while maintaining the viability of cell growth. As mentioned, JM101 and XL1-Blue strains gave many false-positive results.[31] On the other hand, the MC1061Δ*thiE* was the only strain tested with zero false positives. The MC1061 strain is a Str^r mutant that has a point mutation (K43R) within the *rpsL* gene (encoding ribosomal protein S12).[34] Mutation K43R has been suggested to stabilize the *ram* state by increasing the tRNA affinity in the A site of the ribosome.[35–37] The influence of translational processes on protein folding has been amply debated, and a recent study suggests its influence on the relative abundance of some proteins, including CAT.[35,38] These peculiarities may be related to the different expression behavior of these proteins in the strains we have tested. We propose that the change in the stability of the ribosome is probably the main difference between the MC1061 and the other strains that permit the proper expression of our fused proteins, but more experiments are needed to corroborate this hypothesis.

An important finding of this study is that each position accepted foreign loops of different sizes and sequences (most likely due to the allowance of hinge variability). Our data suggest that the sequence of the loop inserted has much less impact on structural adaptations than does the site of the insertion (Fig. 3). This has also been suggested for ubiquitin, where the structural effects of a loop insertion depend primarily on the site of the insertion and much less on the sequence of the insert.[39] A larger data set will be needed to draw more drastic conclusions regarding any significant tolerance difference between other alternative loop positions in the barrel scaffold. The relatively high tolerance to the diversity of sequences tested and the lengths of the inserted loops is in agreement with the general notion that the active-site end of the TIM barrel is inherently tolerant to variability.[10] The high proportion of chimeric proteins (30–90%) that remained foldable and relatively stable after the introduction of foreign loops also indicates the feasibility of our approach to afford useful libraries for subsequent functional screening.

It is clear from the data that the introduction of variability at the hinges connecting the loops with the scaffold has a noticeable effect for obtaining folded proteins. The distribution of amino acids found in randomized positions without selective pressure of the analyzed samples shows only the normal bias inherent to the combinations NNG/C and NNT in the oligonucleotides used for mutagenesis. Thus, the amino acid preferences observed with the introduction of each particular loop are indicative of the importance of the position adjacent to the loop for conformational fitness. The constraints imposed on these positions are likely to derive both from the set of interactions with the guest loop and from those corresponding to tightly packed environments within the core of the protein, as suggested by the lower viability observed in libraries from loop 2, where only one of the barrel-core positions was randomized. It is interesting to note that the amino acids with higher propensity at the hinge positions are dependent on the guest loop and always different from the original residue at the host protein. However, these mutated positions may have been selected for function rather than stability in the host protein because it is well documented that amino acids constituting the catalytic site may be in unstable conformations to favor catalysis.[40] A sequence alignment of all nonredundant PRAI proteins shows that the residues mutated are well conserved (I29 by 74%, Q79 by 96%, H81 by 91%, V124 by 80%, and D126 by 98%). Furthermore, computational analysis of the protein stability with ΔΔG Rosetta and Rosetta Design algorithms[41] shows that these mutated residues are destabilizing in the PRAI structure. These observations suggest a role that is more functional than structural for these sites. Since our selection method is based on stability, the appearance of different amino acids is to be expected. The small size of the sample and the different patterns found with various loops precludes a systematic analysis of the sequence preferences, but the fact that many sequences did not produce folded proteins highlights the importance of introducing variability in the libraries. It will be interesting to compare the preferences observed for folding with those observed when searching for activity.

Overall, the far UV–CD and fluorescence emission spectra of the purified chimeras are remarkably similar to the parental protein spectra, demonstrating that the global interactions that maintain the structure are preserved in the chimera proteins tested (Figs. 6 and 7). The rather small differences observed may reflect some local structural changes. Far UV–CD spectra are more sensitive to the formation of secondary-structure elements; however, the purification of fair amounts of the different proteins from the soluble cell extract fraction and the blue shift in the fluorescence spectra suggest that the proteins are properly folded and not molten globule intermediates.[42,43] Interestingly, two of the tested constructs seem to have a higher α-helical content. The guest loop of one of them, loop 2 from 1,6-biphosphate aldolase (Mut_L2_FBPA), forms indeed an α-helix in its original protein. This suggests that this loop probably retains its secondary structure in the PRAI-LoxP scaffold—a likely fact, since interactions that maintain α-helical structure depend on

local contacts.[44] The other, loop 6 from PBGS is a shorter loop than the host loop (Mut_L6-PBGS). Loop 6 in PRAI forms part of a mobile lid that opens and closes the active site of the enzyme. PRAI from *E. coli* has a canonical TIM barrel structure, although helix 6 is of only one-turn length and the segment corresponding to helix 5 is not recognized as such in the crystal structure. These two helices are in the vicinity of the exchanged loop 6. It is possible that a change in this loop promotes the formation or stabilization of α-helical structure. Also, it is important to note that the percentages of secondary structure predicted between PRAI-LoxP and Mut_L4 αTS are the same. The original loop 4 from αTS is only three amino acids longer in length than the same loop from PRAI and the tertiary structures of both are very similar. This probably explains the minimal structural changes in this mutant with respect to those in the other mutants and PRAI-LoxP. The selection of a robust protein as starting point as well as the limitation of mutagenesis to areas relevant to function but not so committed to structural integrity is likely to be important for finding a high proportion of folded chimeras from which to search for activity. The proposed strategy allows exploration of distant mutants not reachable by any of the conventional directed evolution strategies, traversing through extensive valley regions to new peaks in the folding-sequence landscape. Thus, search of novel function is restricted to screening the small population of folded proteins instead of a vast amount of nonviable proteins.

Several observations on natural protein evolution were inspiring for our work. First, it is clear that different members of enzyme superfamilies, which have common ancestors but sometimes rather different catalytic activities,[45] incorporate indels or insertions in the loops besides point mutations. Second, the mechanism of antibody generation, which can be viewed as a natural "directed evolution" machinery, relies on a few scaffolds with variability concentrated at the loops.[46,47] Furthermore, in the antibody system, the loop repertoire is itself limited, suggesting the efficiency of employing modular exchange as opposed to randomization.[48] Third, in enzymes with the TIM barrel fold, the active site is always composed of loops and residues in the same face of the protein architecture.

It is tempting then to speculate that even modest diversion from an original activity and specificity would benefit from the use of loops as modular components of variability. Our work suggests the possibility of "structural evolvability" of the TIM barrel fold by loop grafting. A combinatorial approach would lead to the evolution of new proteins with the same scaffold hopefully generating novel binding and catalytic activities. Nature has probably used this mechanism to evolve different functions in the same fold, as hypothesized in the transition from microbial phosphotriesterase-like lactonase into modern bacterial phosphotriesterase, an enzyme that degrades a synthetic insecticide introduced in the 20th century, by insertions in loop 7.[12] Certainly, insertions and deletions into loops, or loop grafting, are believed to be a primary mechanism for creating enzyme diversity.[19,20] Here we demonstrate that in a protein with the TIM barrel architecture, a general method for exchanging variable-sized loops is possible and efficient enough to provide a reasonable starting point for the selection of new enzymatic activities. The mutability of these loops in isolation offers a broad scope for further engineering of multiple loops simultaneously.

Materials and Methods

Construction of the loop insertions in the PRAI gene scaffold

We designed two oligonucleotides, one that corresponds to coding DNA strand and the other to noncoding DNA strand, for assembling the loops and connectors. The coding oligonucleotide of the different loops had a hanging cytosine at the 3′ end, and the noncoding oligonucleotide includes a hanging adenine at the 3′end. On the other hand, the coding oligonucleotide of the NH2 connectors includes a dangling thymine at the 3′ end and the noncoding oligonucleotide for the CO connector had a dangling guanine at the 3′ end (Fig. 2a). Each coding oligonucleotide was hybridized with its respective noncoding oligonucleotide to allow the production of 20 dsDNA—14 for loops and 6 for connectors.

In the NH2 connector of loop 2, an NNS codon was introduced in the coding position for Ile29 to generate variability. In the case of loops 4 and 6, two NNS and NNT codons were introduced to generate variability at positions Gln79, His81 (loop 4), Val124, and Asp126 (loop 6), respectively. The dsDNA NH2 and CO connectors were independently ligated with dsDNA of each loop to produce two primers. These primers are named NH2 and CO, respectively. The 14 different libraries of PRAI-LoxP were constructed in three steps (Fig. 2b). First, the amino halves of the *trpF*-loxP genes were constructed by PCR using a trpF-loxP containing pDAN5 vector (pDAN5B1A4LinkerB5A8PRAI)[24] as template and the oligonucleotide 5′-ATGACAGTCCGAAGCTTCAGGAGGGGTGTTGATG-3′ with a HindIII restriction site (underlined) as 5′-primer and the NH2 megaprimers as 3′-primer. Second, the carboxyl halves were constructed by PCR using the same vector as template and the CO megaprimer as 5′-primer and the oligonucleotide 5′-ATTGGTTTGCCGCTAGCT-CATTAATATGCGCG-3′ with NheI restriction site (underlined) as 3′-primer. Third, each library was amplified by overlapping extension PCR, using the two PCR products previously obtained and the oligonucleotides 5′-ATGACAGTCCGAAGCTTCAGGAGGGGTGTTGATG-3′ with a HindIII restriction site (underlined) as 5′-primer and 5′-ATTGGTTTGCCGCTAGCTCATTAATATGCGCG-3′ with an NheI restriction site (underlined) as 3′-primer. The amplified loop libraries were cloned into pDAN5 vector (pDAN5B1A4LinkerB5A8PRAI) using HindIII and NheI restriction sites, yielding 14 different libraries of plasmids with the new loop exchanged in the corresponding position in the *trpF*-loxP gene.

The plasmids obtained were used to transform *E. coli* XL1-BlueMrf′ cells (Stratagene) by electroporation. Transformed cells were plated on LB medium containing ampicillin to select for plasmid uptake. The individual loop libraries contained from 2.4×10^4 to 4.5×10^5 different variants as estimated from the numbers of grown colonies.

Grown colonies were picked from the plates and amplified in liquid medium to purify plasmid DNA. Plasmid DNA of 283 clones (approximately 20 clones from each loop library) grown under such nonselective conditions were sequenced.

Construction of vector for fusion to CAT gene

The CAT fusion vector pDAN5-*trpF*-loxP-CAT was derived from pDAN5B1A4LinkerB5A8PRAI. pDAN5-*trpF*-LoxP-CAT was constructed in three steps. First a BamHI site was deleted in the original vector pDAN5-*trpF*-loxP by using PCR of whole plasmid (MEGAWHOP)[49] with the complementary oligonucleotides 5′-TGGCTTTAATGAGGGTCCATTCGTTTGTGA-3′ and 5′-TCACAAACGAATGGACCCTCATTAAAGGCA-3′ as overlapping primers, yielding a vector with the deletion of BamHI restriction site by the change in one nucleotide (underlined). This vector was named pDAN5-*trpF*2. Second, the *trpF*-CAT fusion was amplified by PCR using a *trpF*-CAT containing pT4 vector (pT4-*trpF*-CAT) kindly provided by Dr. Lorenzo Segovia (unpublished results) as template and the oligonucleotide 5′-CAGTCGAAGCTTCAGGAGGGGTGTTGATGCTGGAGAAT-AAGGTAT-3′ with HindIII site (underlined) as 5′-primer, and 5′-TGCCGCTAGCTTAGTGGTGGTGGTGGTGATCTTTA-3′ with NheI site (underlined) as 3′-primer. The amplified fragment was cloned into pDAN5-*trpF*2 using HindIII and NheI restriction sites, introducing a new BamHI site between *trpF* and CAT genes and yielding pDAN5-*trpF*-CAT vector. Sequences inserted into the multiple cloning site of pDAN5-*trpF*-CAT between the restriction sites HindIII and BamHI led to the expression of proteins fused to CAT at their C-terminus.

Construction of positive and negative controls for folding selection

The positive control to select folded proteins was the WT PRAI protein fused to CAT (pDAN5-*trpF*-CAT vector). Two proteins proven to have poor folding, a circular permutation of PRAI previously described[30] and an engineered antibody (kindly provided by Dr. J. Osuna, unpublished result) fused to CAT, were used as negative controls. Additionally, a construction of WT PRAI with two stop codons between PRAI and CAT fusion was used as negative control.

The circular permutation constructed for this study was the β4/α3 variant previously reported.[30] For its construction, new amino and carboxyl termini were created between α-helix 3 and β-strand 4 using the oligonucleotide 5′-AGTCCGAAGCTTCAGGAGGGGTGTTGATGCGCGATCTGGCAGTCCAACTGCATGG-3′ with HindIII site (underlined) as 5′-primer and the oligonucleotide 5′-CCCGCCAGAGGGATCGTAATATGCGCGCAGCGTC-3′as 3′-primer to generate the new amino termini at β-strand 4. The oligonucleotide 5′-TACGATCCCTCTGGCGGG-CTCGAGAATAAAGTATGT-3′ as 5′-primer and the oligonucleotide 5′-TGCC-GGATCCGTGATCCGATAACACCTTAGCTTTG-3′ with BamH1 site (underlined) as 3′-primer were used to amplify the new C-terminal segment. Both amplifications used a WT PRAI containing vector (pUC18-*trpF*-WT) as template. The two amplified fragments were used to amplify the permuted gene by overlapping extension PCR using oligonucleotide 5′-AGTCCGAAGCTTCAGGAGGGGTGTTGATGCGCGATCTGGC-AGTCCAACTGCATGG-3′ with HindIII site (underlined) as 5′-primer and the oligonucleotide 5′-TGCCGGATCCGTGATCCGATAACACCTTAGCTTTG-3′ with BamH1 site (underlined) as 3′-primer. The amplified fragment was cloned into pDAN5-*trpF*-CAT using HindIII and BamHI restriction sites, yielding the variant Per3-β4/α3-CAT in fusion with CAT.

The fusion of the antibody gene to CAT gene was assembled by PCR using the previously described antibody (Vh) containing pT4 vector as template and the oligonucleotides 5′-AGTCCGAAGCTTCAGGAGGGGTGTTGATGGACGTCCAGCTCCAGCAGTCTGGA-3′ with HindIII site (underlined) as 5′-primer and the oligonucleotide 5′-TGCC-GGATCCACTAGTGACAGTGACCAGAGTACCTTG-3′ with BamH1 site (underlined) as 3′-primer. The amplified fragment was cloned into pDAN5-*trpF*-CAT using HindIII and BamHI restriction sites, yielding the Vh-CAT variant.

The construction of WT PRAI with two stop codons between PRAI and CAT fusion was assembled by PCR using a WT PRAI containing vector (pUC18-*trpF*-WT) as template and the oligonucleotide 5′-ATGACAGTCCGAAGCTTCAGGAGGGGTGT-TGATG-3′ with HindIII site (underlined) as 5′-primer and the oligonucleotide 5′-TGCCGGATCC*TCATTA*ATATGCGCGCAGCGTC-3′with BamH1 site (underlined) and with the two stop codons (italic) as 3′-primer. The amplified fragment was cloned into pDAN5-*trpF*-CAT using HindIII and BamHI restriction sites, yielding the PRAI variant with two stop codons before the fusion to CAT gene.

Chloramphenicol resistance analysis

E. coli MC1061Δ*thiE* cells were transformed with the different variants. Dilutions of the transformed cells were spread on LB medium plates containing Cm (0–100 μg/ml) and ampicillin (200 μg/ml) and incubated for 18 h at 30 °C and the number of colonies under each condition was quantified. The vector pDAN5-*trpF*-CAT, containing a fusion of the WT *trpF* gene fused to CAT gene was used as a control to normalize the survival rate under each condition. A concentration of Cm that discriminated between the negative controls and the WT PRAI, with the highest survival rate for the last one, was picked as the selection condition. In this case, we observe that a Cm concentration as low as 20 μg/ml was adequate to discriminate unfolded variants while maintaining high viability of cells bearing viable genes.

Assessment of expression of soluble fusion protein in selected clones

The *E. coli* MC1061Δ*thiE* strain was transformed by electroporation with each of the constructed loop libraries fused to CAT. Following shaking of the transformants for 1 h at 37 °C, the cells were streaked on LB agar plates containing Cm (20 μg/ml) and ampicillin (200 μg/ml). After incubation at 30 °C for 18 h, eight colonies from each library were selected and precultured in TB liquid medium supplemented with ampicillin (200 μg/ml, for maintenance of plasmid) and incubated for 12 h at 37 °C in the shaker. These precultures (200 μl) were used to inoculate 5 ml of fresh TB liquid medium supplemented with ampicillin (200 μg/ml) and incubated for 18 h at 18 °C. Cells were harvested by centrifugation (Eppendorf/5804-R5, FA-45-30-11 rotor, 5 min, 4000 rpm at 4 °C). The cells were resuspended in 200 μl of 10 mM potassium phosphate buffer at pH 7.6 supplemented with protease inhibitor cocktail [Complete, Free EDTA (ethylenediaminetetraacetic acid), Roche]. The cells were lysed by sonication

(Branson Sonifier 450; 10 s six times at 30-s intervals, 50% pulse, 0 °C) and centrifuged again (Eppendorf/5804-R5, FA-45-30-11 rotor, 20 min, 11,000 rpm at 4 °C) to separate the soluble and insoluble fractions of the cell extract. After separation of 10 μL of soluble extract on a 13% SDS-PAGE and transfer to nitrocellulose membranes (Amersham Pharmacia Bioscience) for 90 min at 80 mA in a semidry transfer unit (Hoefer SemiPhor-Amersham Pharmacia Biotech), Western blotting was performed according to standard protocols,[50] using affinity-purified anti-CAT-digoxigenin (Roche) at a 1:3000 dilution as a primary antibody. The second antibody used was the anti-digoxigenin-AP (Roche) at a 1:5000 dilution and visualized with BCIP/NBT alkaline phosphatase substrate solution (Sigma) for the detection by colorimetric reaction.

Determination of the percentage of folding of the loop libraries

Each loop library fused to CAT gene (pDAN5-Loop Library-CAT), as well as the controls pDAN5-*trpF*-CAT containing the WT PRAI, pDAN5-*trpF*C/S-CAT containing the WT PRAI with two stop codons between *trpF* and CAT gene fusion, pDAN5-Vh-CAT containing the antibody fused to CAT, and pDAN5-per3-CAT including the circular permutation of PRAI, were independently transformed in *E. coli* MC1061Δ*thiE* electrocompetent cells in triplicate. Dilutions of transformed cells were spread in two separate LB medium agar plates, one supplemented with Cm (20 μg/ml) and ampicillin (200 μg/ml), and the other only with ampicillin (200 μg/ml). These plates were incubated for 18 h at 30 °C and the numbers of colonies grown between these two conditions for each plasmid were compared. The folding percentage of the libraries was calculated as the ratio of number of colonies grown under the selective pressure [Cm (20 μg/ml) and ampicillin (200 μg/ml)] to the number of colonies grown without this selective pressure [only ampicillin (200 μg/ml)]. This value was corrected by the number of WT contaminant colonies found under selective pressure for each loop library.

In vivo selection for soluble variants

The *E. coli* MC1061Δ*thiE* strain was transformed by electroporation with the corresponding loop library fused to CAT. Following shaking of the transformants for 1 h at 37 °C, the cells were streaked onto LB agar plates containing Cm (20 μg/ml) and ampicillin (200 μg/ml) and incubated at 30 °C for 18 h. Grown colonies were selected from the plates and grown in liquid LB medium to purify the plasmid DNA. A total of 262 clones grown under selective pressure for folding were sequenced (approximately 18 clones from each loop library).

Statistical analysis of sequences

The sequences found in the mutagenized positions under selective pressure (ampicillin and Cm) were compared with the sequences found without selective pressure (ampicillin). The amino acids observed in these positions were converted to frequencies and the difference between these indicates the discrepancy of occurrence for each residue between these two conditions. The average and the standard deviation of the frequencies were used to determine with a 95% confidence which amino acids were negatively or positively selected in proteins folded for each loop library.

Western blot analysis

The *E. coli* strain MC1061Δ*thiE* was transformed by electroporation with three loop libraries and pDAN5-*trpF*-CAT containing the WT *trpF*. The libraries used for this analysis were loop2 from MR, loop 4 from Ure, and loop 6 from PBGS. After 1-h incubation at 37 °C under shaking, the cells were streaked onto LB plates containing ampicillin (200 μg/ml) and incubated at 37 °C for 12 h. Ten grown colonies were scratched from the plates and resuspended in 5 ml of TB liquid medium supplemented with ampicillin (200 μg/ml, for maintenance of plasmid) and incubated for 12 h at 37 °C in the shaker. Precultures (200 μl) were used to inoculate 5 ml of fresh TB liquid medium supplemented with ampicillin (200 μg/ml) and incubated for 18 h at 18 °C. Cells were harvested and treated as described above. Total protein concentration was measured by Bradford reagent (Bio-Rad). After separation of 11 μg of soluble extract on a 13% SDS-PAGE, the protein was transferred to nitrocellulose membranes (Amersham Pharmacia Bioscience) and analyzed as previously described.

Cloning of selected variants into pET28b(+)

In order to produce the selected variants in the absence of CAT as well as to introduce a C-terminal His6 tag, the *trpF*-loxP and the genes from four true positive colonies grown in the presence of Cm (20 μg/ml) were subcloned from pDAN5-*trpF*-CAT into pET28b(+) vector. To this end, the PRAI-LoxP and the PRAI variant genes (Mut_L4_Ure, Mut_L4_αTS, Mut_L6_PBGS, and Mut_L2_FBPA) were amplified by PCR using the oligonucleotides 5′-GCCA-TACCATGGGGGAGAATAAGGTATG-TGGC-3′ with a NcoI site (underlined) as 5′-primer and 5′-GTCCGAAA-GCTTTCATT-AGTGGTGGTGGTGGTGGTGGGATCCA-TATGCGCGCA-3′with a HindIII site (underlined) as 3′-primer. The amplified product was ligated with pET28b (+), yielding the PRAI-LoxP and the PRAI variants cloned in the pET28b(+) vector. All constructs were sequenced entirely to exclude inadvertent mutations.

Analysis of protein solubility

The mutants Mut_L2_FBPA, Mut_L4_ αTS, Mut_L6_PBGS, and PRAI-LoxP were produced in 5-ml cultures of *E. coli* Rosetta2 cells (Novagen) as described above. Each culture was centrifuged (Eppendorf/5804-R5, FA-45-30-11 rotor, 5 min, 4000 rpm at 4 °C). The cells were resuspended in 0.4 ml of 10 mM potassium phosphate, 0.5 mM EDTA (pH 7.6), 50 mM NaCl, 5% glycerol, 0.1 mM DTT, 0.1 mM PMSF, and 2.5 mg lysozyme and lysed by sonication (Branson Sonifier 450; 10 s six times at 30-s intervals, 50% pulse, 0 °C). The soluble and insoluble fractions of the cell extract were separated. The soluble fractions were applied to SDS-PAGE [13% (w/v) acrylamide].

Expression of PRAI-LoxP variants and protein purification

The expression of the PRAI-LoxP variants was performed in *E. coli* Rosetta2 cells (Novagen) transformed with various pET28b(+) plasmids containing the encoding sequences. To this end, 1 l of LB medium supplemented with kanamycin [50 μg/ml, for maintenance of pET28b(+)] and Cm (25 μg/ml, for maintenance of pRARE) was inoculated with a preculture and incubated at 37 °C. After

an OD_{600} of 0.7 was reached, expression was induced by addition of 1 mM IPTG, and growth was continued for another 15 h at 20 °C. The cells were harvested by centrifugation (Eppendorf/5804-R5, F34-6-38 rotor, 5 min, 4000 rpm at 4 °C) and resuspended in 25 ml of 10 mM potassium phosphate buffer at pH 7.6, 0.5 mM, 50 mM NaCl, 5% glycerol, 0.1 mM DTT, 0.1 mM PMSF, and 2.5 mg of lysozyme, lysed by sonication (Branson Sonifier 450; 20 s six times at 30-s intervals, 50% pulse, 0 °C), and centrifuged again (Eppendorf/5804-R5, F34-6-38 rotor, 20 min, 13000 rpm at 4 °C) to separate the soluble from the insoluble fraction of the cell extract.

The variants Mut_L2_FBPA (mutant with loop 2 from the FBPA library), Mut_L4_αTS (mutant with loop 4 from the αTS library), Mut_L6_PBGS (mutant with loop 4 from the PBGS library), and PRAI-LoxP were purified from the soluble cell fraction. To this end, the extract was loaded onto a nickel Sepharose column (HisTrap FF crude 5 ml, GE Healthcare) previously equilibrated with 50 mM potassium phosphate and 300 mM NaCl buffer at pH 7.6. The column was equilibrated with the same buffer, and the bound His6-tagged protein was eluted by applying a linear gradient from 1 to 500 mM imidazole. Fractions with pure protein were pooled and these were concentrated in the Amicon Ultra-15 system (Millipore) until a final volume of 300 μl was reached. This volume was loaded onto a gel-permeation column (Sephacryl S200, GE Healthcare) that had been previously equilibrated with 50 mM potassium phosphate, 1 mM EDTA, and 0.4 mM DTT buffer at pH 7.6. The different proteins were eluted with the same buffer, and the fractions with pure protein were pooled.

Analytical methods

The different proteins were concentrated using 10-kDa cutoff Amicon® membranes and dialyzed against 4× 1-l degassed 10 mM sodium phosphate, 1 mM EDTA, and 1 mM 2-mercaptoethenol, pH 7.6 buffer. Protein concentration was estimated using an extinction coefficient of 22,900 M^{-1} cm^{-1} at 280 nm. Protein samples were prepared in the same buffer and in 9 M urea, 10 mM sodium phosphate, 1 mM EDTA, and 1 mM 2-mercaptoethanol, pH 7.6 buffer, with a final protein concentration of 16.5 μM for CD studies and 2.0 μM for fluorescence studies.

CD spectra were recorded with a JASCO model J-715 spectropolarimeter equipped with a Peltier temperature control supplied by Jasco. Spectra were collected from 260 to 190 nm. Buffer conditions were 10 mM potassium phosphate, 1 mM EDTA, and 1 mM β-mercaptoethanol (BME) at pH 7.6 and 25.0 °C. Eight replicate spectra were collected from each sample to improve signal-to-noise ratio. The final protein concentration was 16.5 μM, and spectra were collected in a 0.01-cm path-length cell. The secondary-structure prediction was performed using the CDSSTR algorithm, which requires data from 190 to 240 nm.[51–55] Loss of secondary structure was measured in samples of the different proteins in 8.0 M urea incubated at 25 °C for at least 2 h before the spectrum was recorded.

Fluorescence emission spectra were recorded on an LS50B spectrofluorimeter (Perkin Elmer, Norwalk, CT) equipped with a thermostated cell compartment. Spectra were collected from 300 to 540 nm using a 295-nm excitation wavelength at 25 °C (excitation and emission slit widths were 4.0 nm). Protein samples were prepared at 50 μg/ml concentration either in 10 mM phosphate, 1 mM EDTA, and 1 mM BME buffer at pH 7.6 or in 9.0 M urea in the same buffer. Spectra were recorded after 2 h incubation at 25 °C.

FPL size-exclusion chromatography data analysis

The purified enzymes were analyzed by size-exclusion chromatography to determine their nature of oligomerization in an Äkta FPLC system equipped with a UV detector and a size-exclusion Superose HR12 column from Amersham Biosciences (Uppsala, Sweden). The samples were eluted with a 20 mM phosphate, 300 mM KCl, and 1 mM BME buffer at pH 7.6 at a flow rate of 0.5 ml/min. The column was calibrated with lysozyme, lectin, and bovine serum albumin. The proteins were injected at initial concentrations of 2, 20, and 100 μM.

Acknowledgements

This work was supported by grant IN220802 from PAPIIT and CONACyT 49590-Q to G.S.R. We thank Dr. Helena Wright and Dr. Vilmos Fülop for experimental help in the overexpression and purification of the PRAI-LoxP and PRAI variants. We thank Azucena Carrillo Hernández, Yara Sánchez Corrales, and Paulina Tapia Quiroz for their technical assistance in the construction of some of the libraries.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2009.02.022

References

1. Grauer, D. a. & Li, W. -H. (2000). *Fundamentals of Molecular Evolution*. Sinauer Associates, Inc., Sunderland, MA.
2. Graziano, J. J., Liu, W., Perera, R., Geierstanger, B. H., Lesley, S. A. & Schultz, P. G. (2008). Selecting folded proteins from a library of secondary structural elements. *J. Am. Chem. Soc.* **130**, 176–185.
3. Nagano, N., Orengo, C. A. & Thornton, J. M. (2002). One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J. Mol. Biol.* **321**, 741–765.
4. Voigt, C. A., Martinez, C., Wang, Z. G., Mayo, S. L. & Arnold, F. H. (2002). Protein building blocks preserved by recombination. *Nat. Struct. Biol.* **9**, 553–558.
5. Shao, Z. & Arnold, F. H. (1996). Engineering new functions and altering existing functions. *Curr. Opin. Struct. Biol.* **6**, 513–518.
6. Nilsson, B. (1995). Antibody engineering. *Curr. Opin. Struct. Biol.* **5**, 450–456.
7. Wierenga, R. K. (2001). The TIM-barrel fold: a versatile framework for efficient enzymes. *FEBS Lett.* **492**, 193–198.
8. Hocker, B., Beismann-Driemeyer, S., Hettwer, S., Lustig, A. & Sterner, R. (2001). Dissection of a (betaalpha)8-barrel enzyme into two folded halves. *Nat. Struct. Biol.* **8**, 32–36.
9. Lang, D., Thoma, R., Henn-Sax, M., Sterner, R. & Wilmanns, M. (2000). Structural evidence for

evolution of the beta/alpha barrel scaffold by gene duplication and fusion. *Science*, **289**, 1546–1550.
10. Urfer, R. & Kirschner, K. (1992). The importance of surface loops for stabilizing an eightfold beta alpha barrel protein. *Protein Sci.* **1**, 31–45.
11. Sterner, R. & Hocker, B. (2005). Catalytic versatility, stability, and evolution of the (betaalpha)8-barrel enzyme fold. *Chem. Rev.* **105**, 4038–4055.
12. Afriat, L., Roodveldt, C., Manco, G. & Tawfik, D. S. (2006). The latent promiscuity of newly identified microbial lactonases is linked to a recently diverged phosphotriesterase. *Biochemistry*, **45**, 13677–13686.
13. Wright, H., Noda-Garcia, L., Ochoa-Leyva, A., Hodgson, D. A., Fulop, V. & Barona-Gomez, F. (2008). The structure/function relationship of a dual-substrate (betaalpha)8-isomerase. *Biochem. Biophys. Res. Commun.* **365**, 16–21.
14. Patrick, W. M. & Blackburn, J. M. (2005). In vitro selection and characterization of a stable subdomain of phosphoribosylanthranilate isomerase. *FEBS J.* **272**, 3684–3697.
15. Cheon, Y. H., Park, H. S., Kim, J. H., Kim, Y. & Kim, H. S. (2004). Manipulation of the active site loops of D-hydantoinase, a (beta/alpha)8-barrel protein, for modulation of the substrate specificity. *Biochemistry*, **43**, 7413–7420.
16. Forrer, P., Stumpp, M. T., Binz, H. K. & Pluckthun, A. (2003). A novel strategy to design binding molecules harnessing the modular nature of repeat proteins. *FEBS Lett.* **539**, 2–6.
17. Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P. & Pluckthun, A. (2003). Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Biol.* **332**, 489–503.
18. DeLano, W. L., Ultsch, M. H., de Vos, A. M. & Wells, J. A. (2000). Convergent solutions to binding at a protein–protein interface. *Science*, **287**, 1279–1283.
19. Park, H. S., Nam, S. H., Lee, J. K., Yoon, C. N., Mannervik, B., Benkovic, S. J. & Kim, H. S. (2006). Design and evolution of new catalytic activity with an existing protein scaffold. *Science*, **311**, 535–538.
20. Tawfik, D. S. (2006). Biochemistry. Loop grafting and the origins of enzyme species. *Science*, **311**, 475–476.
21. Li, S., Li, B., Fei, Y., Jiang, D., Sheng, Y., Sun, Y. & Zhang, J. (2007). Exon grafting yields a "two active-site" lysozyme. *Biochem. Biophys. Res. Commun.* **358**, 997–1001.
22. Barona-Gómez, F., Ochoa-Leyva, A. & Soberón, X. (2008). Advances and perspectives in protein engineering: From natural history to directed evolution of enzymes. In *Advances in Protein Physical Chemistry* (García-Hernández, E. & Fernández-Velasco, D. A., eds), pp. 407–438, Transworld Research Network, Kerala, India.
23. Tsuji, T., Onimaru, M. & Yanagawa, H. (2006). Towards the creation of novel proteins by block shuffling. *Comb. Chem. High Throughput Screen.* **9**, 259–269.
24. Saab-Rincon, G., Mancera, E., Montero-Moran, G., Sanchez, F. & Soberon, X. (2005). Generation of variability by in vivo recombination of halves of a (beta/alpha)8 barrel protein. *Biomol. Eng.* **22**, 113–120.
25. Soberon, X., Fuentes-Gallego, P. & Saab-Rincon, G. (2004). In vivo fragment complementation of a (beta/alpha)(8) barrel protein: generation of variability by recombination. *FEBS Lett.* **560**, 167–172.
26. Eberhard, M., Tsai-Pflugfelder, M., Bolewska, K., Hommel, U. & Kirschner, K. (1995). Indoleglycerol phosphate synthase-phosphoribosyl anthranilate isomerase: comparison of the bifunctional enzyme from *Escherichia coli* with engineered monofunctional domains. *Biochemistry*, **34**, 5419–5428.
27. Farber, G. K. & Petsko, G. A. (1990). The evolution of alpha/beta barrel enzymes. *Trends Biochem. Sci.* **15**, 228–234.
28. Maxwell, K. L., Mittermaier, A. K., Forman-Kay, J. D. & Davidson, A. R. (1999). A simple in vivo assay for increased protein solubility. *Protein Sci.* **8**, 1908–1911.
29. Sieber, V., Martinez, C. A. & Arnold, F. H. (2001). Libraries of hybrid proteins from distantly related sequences. *Nat. Biotechnol.* **19**, 456–460.
30. Akanuma, S. & Yamagishi, A. (2005). Identification and characterization of key substructures involved in the early folding events of a (beta/alpha)8-barrel protein as studied by experimental and computational methods. *J. Mol. Biol.* **353**, 1161–1170.
31. Seitz, T., Bocola, M., Claren, J. & Sterner, R. (2007). Stabilisation of a (betaalpha)8-barrel protein designed from identical half barrels. *J. Mol. Biol.* **372**, 114–129.
32. Priestle, J. P., Grutter, M. G., White, J. L., Vincent, M. G., Kania, M., Wilson, E. *et al.* (1987). Three-dimensional structure of the bifunctional enzyme *N*-(5′-phosphoribosyl)anthranilate isomerase-indole-3-glycerol-phosphate synthase from *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **84**, 5690–5694.
33. Akanuma, S. & Yamagishi, A. (2008). Experimental evidence for the existence of a stable half-barrel subdomain in the (beta/alpha)8-barrel fold. *J. Mol. Biol.* **382**, 458–466.
34. Casadaban, M. J. & Cohen, S. N. (1980). Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *J. Mol. Biol.* **138**, 179–207.
35. Chumpolkulwong, N., Hori-Takemoto, C., Hosaka, T., Inaoka, T., Kigawa, T., Shirouzu, M. *et al.* (2004). Effects of *Escherichia coli* ribosomal protein S12 mutations on cell-free protein synthesis. *Eur. J. Biochem.* **271**, 1127–1134.
36. Hosaka, T., Tamehiro, N., Chumpolkulwong, N., Hori-Takemoto, C., Shirouzu, M., Yokoyama, S. & Ochi, K. (2004). The novel mutation K87E in ribosomal protein S12 enhances protein synthesis activity during the late growth phase in *Escherichia coli*. *Mol. Genet. Genomics*, **271**, 317–324.
37. Carter, A. P., Clemons, W. M., Brodersen, D. E., Morgan-Warren, R. J., Wimberly, B. T. & Ramakrishnan, V. (2000). Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature*, **407**, 340–348.
38. Horjales, S., Cota, G., Senorale-Pose, M., Rovira, C., Roman, E., Artagaveytia, N. *et al.* (2007). Translational machinery and protein folding: evidence of conformational variants of the estrogen receptor alpha. *Arch. Biochem. Biophys.* **467**, 139–143.
39. Ferraro, D. M., Hope, E. K. & Robertson, A. D. (2005). Site-specific reflex response of ubiquitin to loop insertions. *J. Mol. Biol.* **352**, 575–584.
40. Shoichet, B. K., Baase, W. A., Kuroki, R. & Matthews, B. W. (1995). A relationship between protein stability and protein function. *Proc. Natl Acad. Sci. USA*, **92**, 452–456.
41. Cheng, G., Qian, B., Samudrala, R. & Baker, D. (2005). Improvement in protein functional site prediction by distinguishing structural and functional constraints on protein family evolution using computational design. *Nucleic Acids Res.* **33**, 5861–5867.
42. Kuwajima, K. (1989). The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins*, **6**, 87–103.

43. Baum, J., Dobson, C. M., Evans, P. A. & Hanley, C. (1989). Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig alpha-lactalbumin. *Biochemistry*, **28**, 7–13.
44. Creighton, T. E. (1992). *Proteins: Structures and Molecular Properties*, 2nd edit. W. H. Freeman Company, New York.
45. Babbitt, P. C. & Gerlt, J. A. (1997). Understanding enzyme superfamilies. Chemistry as the fundamental determinant in the evolution of new catalytic activities. *J. Biol. Chem.* **272**, 30591–30594.
46. Tomlinson, I. M., Walter, G., Marks, J. D., Llewelyn, M. B. & Winter, G. (1992). The repertoire of human germline VH sequences reveals about fifty groups of VH segments with different hypervariable loops. *J. Mol. Biol.* **227**, 776–798.
47. Chothia, C., Lesk, A. M., Gherardi, E., Tomlinson, I. M., Walter, G., Marks, J. D. *et al.* (1992). Structural repertoire of the human VH segments. *J. Mol. Biol.* **227**, 799–817.
48. Almagro, J. C., Quintero-Hernandez, V., Ortiz-Leon, M., Velandia, A., Smith, S. L. & Becerril, B. (2006). Design and validation of a synthetic VH repertoire with tailored diversity for protein recognition. *J. Mol. Recognit.* **19**, 413–422.
49. Miyazaki, K. (2003). Creating random mutagenesis libraries by megaprimer PCR of whole plasmid (MEGAWHOP). *Methods Mol. Biol.* **231**, 23–28.
50. Sambrook, J. F. E. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edit. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
51. Whitmore, L. & Wallace, B. A. (2008). Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers*, **89**, 392–400.
52. Whitmore, L. & Wallace, B. A. (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **32**, W668–W673.
53. Sreerama, N. & Woody, R. W. (2000). Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* **287**, 252–260.
54. Compton, L. A. & Johnson, W. C., Jr (1986). Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal. Biochem.* **155**, 155–167.
55. Manavalan, P. & Johnson, W. C., Jr (1987). Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Anal. Biochem.* **167**, 76–85.

Capítulo 2. **Explorando la adaptabilidad de la relación estructura-función de las asas β/α de TrpF**

En esta parte del proyecto analizamos el rol funcional del intercambio de asas β/α en la actividad catalítica de la enzima TrpF de *E. coli*. Dado que la anterior estrategia fue para intercambiar asas de manera aleatoria e insertando variabilidad en los dos últimos residuos de la hebra beta de la estructura de TrpF que anteceden a cada asa intercambiada, decidimos desarrollar una nueva estrategia de intercambio de asas. Esta nueva estrategia incluye la introducción de variabilidad de secuencia (NNS) en los dos últimos residuos de cada asa intercambiada, los cuáles son los residuos que la unirán con el molde de TrpF. Además, las asas intercambiadas se seleccionaron para analizar aspectos específicos de cada una de ellas al ser transplantada al molde de TrpF. Esta estrategia se diseñó imitando al sistema de evolución natural de los anticuerpos, en el cual se introduce variabilidad de secuencia en ambos lados del asa de especificidad H3. Además el asa H3 puede ser codificada por diferentes genes lo cual aporta mayor diversidad y diversifica la unión a diferentes antígenos (*32-34*).

La introducción de variabilidad en los residuos bisagra que conectan las nuevas asas con el resto de la proteína puede permitirles el muestreo de conformaciones que de otra manera no serían posibles (*35*), lo que incluye la exploración de posibles orientaciones óptimas para la catálisis. Con esta nueva estrategia experimental se reemplazó el asa β/α 6 de TrpF. Dicha asa se encuentra localizada justo arriba del residuo catalítico Asp 126 y está formando una tapa flexible sobre el sitio activo, la cual es crítica para la unión del sustrato y para la actividad de PRA isomerasa de TrpF. Las asas utilizadas como reemplazo se seleccionaron por tener relaciones funcionales, estructurales y evolutivas con el asa 6 de TrpF. De esta manera se seleccionó el asa 6 de TrpA, el asa 6 de PriA y el asa 6 de HisA. Adicionalmente, se seleccionaron el asa 6 de MetR sólo por mantener una posición estructural equivalente, mientras que las asas 1 de PriA y 1 de HisA por no tener ningún tipo de relación. Asimismo, para explorar el rol intrínseco del asa 6 original, diseñamos una asa utilizando el programa RosettaDesign (*36*). Con esta asa diseñada se pretende preservar la estructura del asa original pero con diferente secuencia.

La capacidad funcional de cada una de las librerías construidas se evaluó mediante la complementación de la auxotrofía del gen de TrpF en la cepa JM101Δ*trp.* Posteriormente, las librerías se fusionaron al reportero de plegamiento de CAT para evaluar su capacidad de plegamiento usando la metodología de resistencia a cloranfenicol previamente descrita (*16*). Para estimar la capacidad de adaptación (adaptabilidad) en la estructura de TrpF de la relación estructura-función de cada asa intercambiada, desarrollamos un valor denominado SFLA (SFLA: Structure-Function Loop Adaptability). Este valor se utilizó para medir la capacidad de las variantes plegadas de acomodar productivamente el sustrato dentro del nuevo sitio activo. El valor final refleja la adaptabilidad de la relación estructura-función inherente de cada asa para mantener la actividad de PRA isomerasa en el molde de TrpF. Como valor de referencia probamos el efecto de introducir variabilidad sólo en los residuos bisagras Asn127 y Trp138, pero manteniendo el asa original de TrpF.

Para la librería del asa silvestre con variabilidad en las bisagras se encontró que casi todas las variantes plegadas mantienen la actividad de PRA isomerasa, lo cual resulta en un valor de SFLA de 0.84 e indica la alta adaptabilidad de la relación estructura-función del asa original para acomodar de una manera productiva el sustrato de PRA en el sitio activo. Por otro lado, el asa con mejor adaptabilidad estructural y funcional para mantener la actividad de PRA isomerasa en el molde TrpF fue el asa 6 de TrpA con un valor de SFLA de 0.70. En primera instancia este resultado es interesante, ya que esta enzima carece de la actividad de PRA isomerasa, pero concuerda con la sugerencia previa de que esta asa se encuentra cumpliendo un rol funcional equivalente en TrpF y TrpA, funcionando como una tapa que cierra el sitio activo de manera similar (*37*). Además, se ha sugerido que estas dos enzimas comparten el mismo origen evolutivo (*37*). Estas relaciones pueden explicar la buena adaptabilidad del asa de TrpA en TrpF. Por otro lado, a pesar de la alta similitud tanto de estructura como de secuencia entre las asas 6 de PriA y 6 de HisA, se encontró que el asa 6 de PriA muestra una mejor adaptabilidad (SFLA: 0.39) que el asa 6 de HisA (SFLA: 0.28). Este resultado puede ser explicado debido ha que el asa 6 de PriA tiene una participación importante en la especificidad de sustrato en la actividad de PRA isomerasa (*21*) mientras que HisA no posee esta actividad (*38*). Interesantemente,

después de haber hecho mutaciones que indujeron un cambio conformacional del asa 6 de HisA, se obtuvo la actividad de PRA isomerasa con valores catalíticos cercanos a los de la enzima original de TrpF (*39*). Estas observaciones refuerzan el papel de la funcionalidad de esta asa para la actividad de PRA isomerasa. Contrario a ello, el valor de 0.13 de SFLA obtenido para el asa diseñada por el programa RosettaDesign confirma la pérdida del componente funcional de esta nueva asa, a pesar de mantener el 40% de la identidad de secuencia con el asa original de TrpF. De la misma manera se obtuvieron valores bajos de SFLA para el asa 6 de MetR, la cual solo tiene relación con el asa 6 de TrpF por mantener una posición estructural equivalente, y para las asas 1 de PriA y 1 de HisA, las cuales carecen de relaciones funcionales y estructurales con el asa 6 de TrpF.

De manera similar al asa H3 de los anticuerpos, las asas β/α de los barriles $(\beta/\alpha)_8$ varían en longitud, secuencia y estructura. Aunado a ello, se ha sugerido que la modularidad de las asas β/α es una manera utilizada convenientemente por la naturaleza en la evolución de la función de estas enzimas (*40, 41*). Aunque no sabemos si existe o existió un mecanismo sistemático similar al de los anticuerpos en la historia natural de los barriles $(\beta/\alpha)_8$, es tentador especular que los eventos resultado del intercambio de asas han sido capturados por la presión de selección, como se ha sugerido mediante un análisis filogenético en otra familia de enzimas (*42*). En el caso de TrpF es evidente que el asa 6 cumple los requisitos para su evolucionabilidad funcional, sugiriendo que la modularidad de los barriles $(\beta/\alpha)_8$ puede ser favorecida por la neutralidad (a través de un andamiaje estable) y por la plasticidad funcional (a través de sustituciones de asas β/α). De hecho las asas pueden tener propiedades funcionales que son difíciles de interpretar a partir de su secuencia y las cuales podrían ser trasladadas entre andamiajes diferentes usando la estrategia aquí descrita. Esto está implícito en los altos valores de SFLA del asa 6 de TrpA, la cual se había postulado que cumplía un rol similar en TrpA y TrpF; y del asa 6 de PriA la cual esta implicada en la especificidad y actividad de PRA isomerasa de PriA. En apoyo a esta hipótesis, recientemente se ha observado cómo diferentes secuencias de asas pueden reflejar la selección natural de la dinámica molecular de las mismas, la cual puede ser correlacionado con la función (*43*).

Los detalles en extenso correspondientes a este capítulo se presentan en el artículo de la siguiente página. Es importante señalar tres conclusiones de esta sección las cuales fueron fundamentales para el diseño de las siguientes secciones de nuestro trabajo: 1) es posible intercambiar sistemáticamente el asa β/α 6 de TrpF sin perder totalmente la actividad de PRA isomerasa de esta enzima; 2) se puede medir la adaptabilidad de la estructura-función de las asas reemplazadas mediante el SFLA; y 3) es posible trasladar el componente funcional implícito en las asas, siempre y cuando se diseñen sitios de variabilidad en las bisagras, los cuales son los puntos que la unirán con el nuevo andamiaje.

doi:10.1016/j.jmb.2011.05.027 J. Mol. Biol. (2011) xx, xxx–xxx

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Exploring the Structure–Function Loop Adaptability of a $(\beta/\alpha)_8$-Barrel Enzyme through Loop Swapping and Hinge Variability

Adrián Ochoa-Leyva[1], Francisco Barona-Gómez[2], Gloria Saab-Rincón[1], Karina Verdel-Aranda[2], Filiberto Sánchez[1] and Xavier Soberón[1,3]*

[1]*Departamento de Ingeniería Celular y Biocatálisis, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Avenida Universidad 2001, Cuernavaca, C.P. 62210, México*
[2]*Evolution of Metabolic Diversity Laboratory, Laboratorio Nacional de Genómica para la Biodiversidad (Langebio), CINVESTAV-IPN, Km 9.6 Libramiento Norte, Carretera Irapuato-León, Irapuato, C.P. 36822, México*
[3]*Instituto Nacional de Medicina Genómica (INMEGEN), Secretaría de Salud, México City, Distrito Federal, C.P. 01900, Mexico*

Received 13 December 2010; received in revised form 31 March 2011; accepted 18 May 2011

Edited by C. R. Mathews

Keywords:
protein evolution;
protein engineering;
structure–function relationship;
loop swapping;
enzyme design

Evolution of proteins involves sequence changes that are frequently localized at loop regions, revealing their important role in natural evolution. However, the development of strategies to understand and imitate such events constitutes a challenge to design novel enzymes in the laboratory. In this study, we show how to adapt loop swapping as semiautonomous units of functional groups in an enzyme with the $(\beta/\alpha)_8$-barrel and how this functional adaptation can be measured *in vivo*. To mimic the natural mechanism providing loop variability in antibodies, we developed an overlap PCR strategy. This includes introduction of sequence diversity at two hinge residues, which connect the new loops with the rest of the protein scaffold, and we demonstrate that this is necessary for a successful exploration of functional sequence space. This design allowed us to explore the sequence requirements to functional adaptation of each loop replacement that may not be sampled otherwise. Libraries generated following this strategy were evaluated in terms of their folding competence and their functional proficiency, an observation that was formalized as a Structure–Function Loop Adaptability value. Molecular details about the function and structure of some variants were obtained by enzyme kinetics and circular dichroism. This strategy yields functional variants that retain the original activity at higher frequencies, suggesting a new strategy for protein engineering that incorporates a more divergent sequence exploration beyond that limited to point mutations. We discuss how this approach may provide insights into the mechanism of enzyme evolution and function.

Introduction

Molecular evolution of proteins involves mutation and recombination events through sequence space to create new functional properties.[1] During the study of these evolutionary processes in the laboratory, point mutations have played an important role, but these are only able to search local regions of sequence space. For

*Corresponding author. E-mail address: soberon@ibt.unam.mx.

Abbreviations used: SFLA, Structure–Function Loop Adaptability; PriA, phosphoribosyl isomerase A; CFU, colony-forming units; PTE, phosphotriesterase; PRA, *N*-(5′-phosphoribosyl)anthranilate.

0022-2836/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.

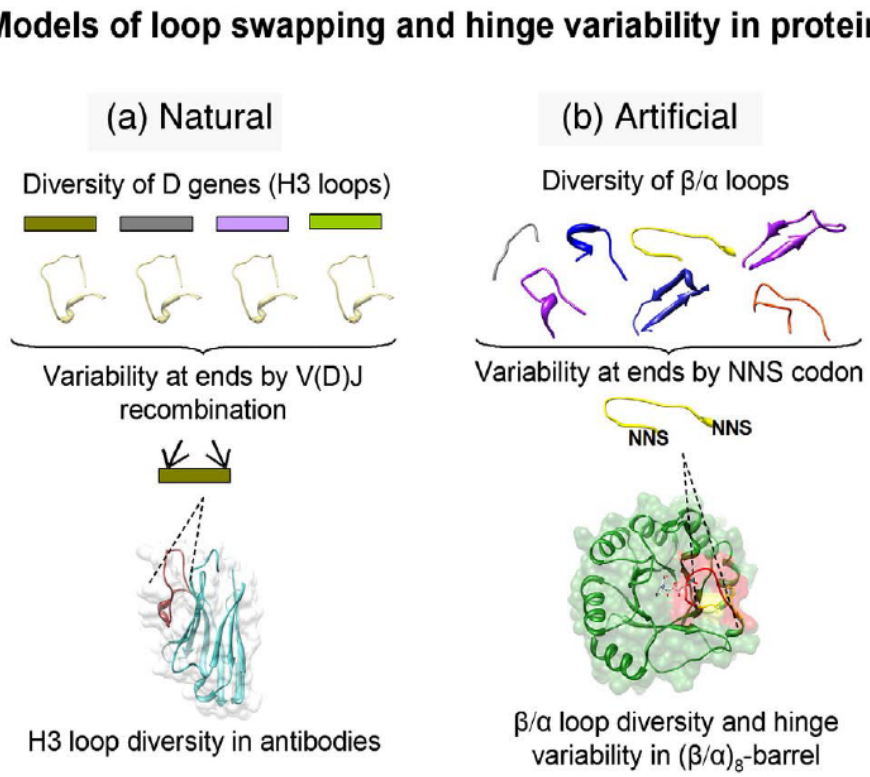

Fig. 1. Models of loop diversity and adaptability in proteins. (a) Natural loop diversity and adaptability in antibodies. The most important loop in antibodies (H3 loop) is encoded by different D genes. Additional variability is introduced by the V(D)J recombination event, generating sequence diversity at both ends of the selected D gene.[5] The diversity of the H3 loops and the introduction of variability in both junctions result in a large number of potential binding sites subsequently selected against the antigens. The three-dimensional (3D) structure of the single domain antibody cAb-Lys2 [Protein Data Bank (PDB) ID: 1rjc][5] is illustrated at the bottom. The H3 loop is shown in brown. (b) Artificial loop diversity and adaptability in $(\beta/\alpha)_8$-barrels proposed in this article. Different β/α loops have embedded functional information representing a diversity source that may represent functional protein modules that can be adapted to a different scaffold through the introduction of variability at both ends. The 3D structure of ecTrpF (PDB ID: 1pii) forming a complex with the product analogue reduced 1′-(2′-carboxyphenylamino)-1′-deoxiribulose 5′-phosphate (rCdRP) is illustrated at the bottom. β/α loop 6 is shown in red. Variability was introduced using the degenerate codon NNS at both ends of the loops (hinges). The 3D structures were rendered using the Chimera package.[8]

exploration of more divergent variability, sequence changes such as insertions, deletions or module exchanges should be taken into account.[2] In natural sequences, these changes are frequently localized at nonstructured specific regions, i.e., loops, revealing their role in protein evolution and in the diversification of numerous enzyme families and superfamilies.[3] Usually, the sequence identity between loops from unrelated enzymes is low or does not exist. Otherwise, there is a high sequence conservation among loops from homologous enzymes, revealing their structure–function relationships.[4] Incorporating such types of sequence changes would be a very valuable asset to the engineering of proteins, but methodologies to successfully mimic such events in the laboratory remain to be developed.

Antibodies are a versatile example of natural loop engineering to design different binding specificities. Recently, it was pointed out that the most important loop (H3 loop) within the antigen binding site is mainly encoded by a particular D gene, inserted between the V and J genes.[5] Picking this gene out of a small pool provides most of the diversity. Additional variability is introduced by the imprecision in V(D)J recombination event, generating sequence diversity at both ends of the selected D gene and, therefore, on the structure of the H3 loop.[5–7] This introduction of variability in both junctions of the H3 loop, in addition with the combination of different D genes, results in a large number of potential binding sites subsequently selected against the antigens. This example provides information about the natural functional adaptation of this loop in antibodies (Fig. 1).

Establishing the structure–function relationship of loops, which in some cases lack electron density in

crystallographic structural models, is a fundamental question in protein science. Loops are known to play critical roles of enzyme function, including catalysis,[9] substrate specificity[10] and protein–protein interactions.[11] Thus, the development of novel methods for engineering of loops would promote the construction of enzymes with novel substrate specificities or functions.[12] In order to face this challenge, our group has recently developed a novel systematic methodology, which was employed to investigate the structural evolvability of a modified $(\beta/\alpha)_8$-barrel enzyme by different loop exchanges.[13] We have originally focused on this protein fold since the $(\beta/\alpha)_8$-barrel is one of the most common folds among protein catalysts and has great functional and structural versatility.[14] Conveniently, the residues of their active sites are located at the C-terminal ends of β-strands as well as in loops that link β-strands with α-helices (i.e., β/α loops), which are considered not to contribute to protein stability.[13,14]

The present work focuses on the effectiveness that loop swapping and hinge variability may have in exploring the functional sequence space of a $(\beta/\alpha)_8$-barrel enzyme, allowing a more divergent sequence exploration,[2] which thus far has largely been limited to point mutations. Additionally, to mimic the natural system of antibodies that produces diversity on junctions of their H3 loop[5] and on the basis of the assumption that junctions might require different conformations to accommodate each loop replacement, we randomized the residues that connect the new loops with the rest of the protein scaffold (Fig. 1). To this end, the β/α loop 6 of monofunctional *N*-(5′-phosphoribosyl)anthranilate (PRA) isomerase from *Escherichia coli* (ecTrpF) was selected.[15] This enzyme catalyzes the Amadori rearrangement of PRA to 1′-(2′-carboxyphenylamino)-1′-deoxiribulose 5′-phosphate (CdRP),[16] which is the third step in the synthesis of tryptophan from chorismic acid. The loop 6 is located just above the catalytic residue Asp126, and it forms a long and flexible lid over the active-site pocket, which is critical for substrate binding and PRA isomerase activity.[17]

Loop replacements were performed with an overlap PCR strategy (Fig. 3a) that includes the introduction of sequence variability at the amino and the carboxy ends, i.e., the hinge residues of each loop replacement, mimicking the natural system of antibodies to generate diversity at both ends of the H3 loop.[5] The final libraries were evaluated in terms of their folding competence and their functional proficiency, an observation that was formalized as a Structure–Function Loop Adaptability (SFLA) value. Details about the function and stability of some variants were obtained by enzyme kinetic studies of PRA isomerase activity, circular dichroism (CD) and size-exclusion chromatography. The observation that varying the hinge residues of loops allows preservation of function on a large number of variants suggests a whole new strategy for protein engineering, incorporating a more divergent sequence exploration, beyond that limited to point mutations.

Results

Loop replacement and hinge variability design

It has been suggested that β/α loop 6, which connects the β-strand 6 to helix 6, fulfills an equivalent functional role, as a lid that closes down upon substrate binding in three different $(\beta/\alpha)_8$-barrel enzymes: TrpF (PRA isomerase, EC 5.3.1.24), TrpC (indole-3-glycerol-phosphate synthase, EC 4.1.1.48) and TrpA (α-subunit of tryptophan synthase, EC 4.2.1.20).[18] These enzymes catalyze consecutive reactions in tryptophan biosynthesis, which share conserved elements of substrate specificity, including the phosphate binding site, supporting the idea that they share a common evolutionary origin.[18] In histidine biosynthesis, the enzyme HisA (ProFAR isomerase, EC 5.3.1.16) uses a reaction mechanism similar to that of TrpF to convert an identical aminoaldose moiety into the corresponding aminoketose.[19] Additionally, PRA isomerase activity can be established in HisA,[20] and after several rounds of mutagenesis, it was observed that conformational changes in its β/α loop 6 are important to increase this activity.[21] Moreover, an enzyme homologue of HisA, dubbed phosphoribosyl isomerase A (PriA), which takes part in both histidine and tryptophan biosynthesis, therefore harboring both ProFAR and PRA isomerase activities, was discovered.[22] PriA is structurally similar to HisA, and conformational changes in β/α loops 5 and 6 have been suggested as important for its PRA isomerase activity.[23,24]

The loops to be used as replacements were selected with the aim of covering a range of functional, structural and evolutionary relationships with loop 6 of ecTrpF (Loop 6 Wt) (Table 1 and Fig. 2). Loop 6 of tryptophan synthase α chain from *E. coli* (Loop 6 ecTrpA), loop 6 of PriA from *Streptomyces coelicolor* (Loop 6 scPriA) and loop 6 of ProFAR isomerase from *Thermotoga maritima* (Loop 6 tmHisA) have all three types of relationships, as was described above. Loop 6 of methyltetrahydrofolate–corrinoid-dependent methyltransferase from *Moorella thermoacetica* (Loop 6 mtMetR) was chosen only for its structural equivalent position. Additionally, loops 1 of scPriA (Loop 1 scPriA) and tmHisA (Loop 1 tmHisA), which do not have functional, structural or evolutionary relationships to Loop 6 Wt, were selected. Moreover, to explore the intrinsic functional role of the original loop 6, we designed a new loop, predicted to be compatible with the same backbone structure but with a different sequence, using the

Table 1. Sequence and structural characteristics of loop replacements

Loop replacement	Sequence[a]	Length	Structural similarity (RMSD)[b]	Sequence identity[c] (%)	Type of relationships with Loop 6 Wt
Loop 6 Wt	**N**GQGGSGQRFD**W**	12	0 Å on 12 residues	100	—
Loop 6 scPriA	**D**IAKDGTLQGP**N**	12	2.7 Å on 8 residues	17	Functional, structural and evolutionary
Loop 6 tmHisA	**E**KDGTLQEHD**F**	11	2.1 Å on 9 residues	27	Functional, structural and evolutionary
Loop 6 ecTrpA	**S**RAGVTGAENRAALP**L**	16	2.3 Å on 12 residues	25	Functional, structural and evolutionary
Loop 6 mtMetR	**P**LILPANVA**Q**	10	2.8 Å on 7 residues	0	Structural equivalent position
Loop 6 Rosetta	**N**GNEGDGVEHD**W**	12	0.2 Å on 12 residues	40	Theoretically calculated with the same backbone structure
Loop 1 scPriA	**D**GQAVRLVHGESGTETSYG**S**	20	—[d]	—[d]	None
Loop 1 tmHisA	**R**GKVARMIKGRKENTIFYEK**D**	21	—[d]	—[d]	None

[a] Sequence of each loop used for the replacement. The hinge residues mutated to saturation are shown in bold and underlined.
[b] Minimal RMSD calculated by a structural superimposition of the α-carbons of loop 6 of ecTrpF with each loop replacement.
[c] Calculated on the basis of the structural superimposition described in footnote b.
[d] Fewer than three residues aligned; loop 6 of ecTrpF does not match with loop replacement.

Rosetta Design web server.[25] The resulting lowest free-energy sequence (Loop 6 Rosetta) differs by 60% from the original loop 6.

The analysis of protein crystal structures has shown that loop movements can take two main forms: (i) hinge motions that are not constrained by tertiary packing interactions, which are characterized by the localization of motion to a few main-chain torsion angle changes, and (ii) shear motions that are manifested by small side-chain torsion angle changes along an entire interface and closely packed segments of polypeptides.[26] Asn127 and Trp138 are the end residues of the β/α loop 6 of ecTrpF (Fig. 2), which, in comparison with the end residues identified after a superimposition of liganded and unliganded TrpF structures from *T. maritima*, comprise two well-defined protein hinges (data not shown). In addition, on the basis of previous studies on loop 6 of triosephosphate isomerase,[27] an experimental model for the $(\beta/\alpha)_8$-barrels that has

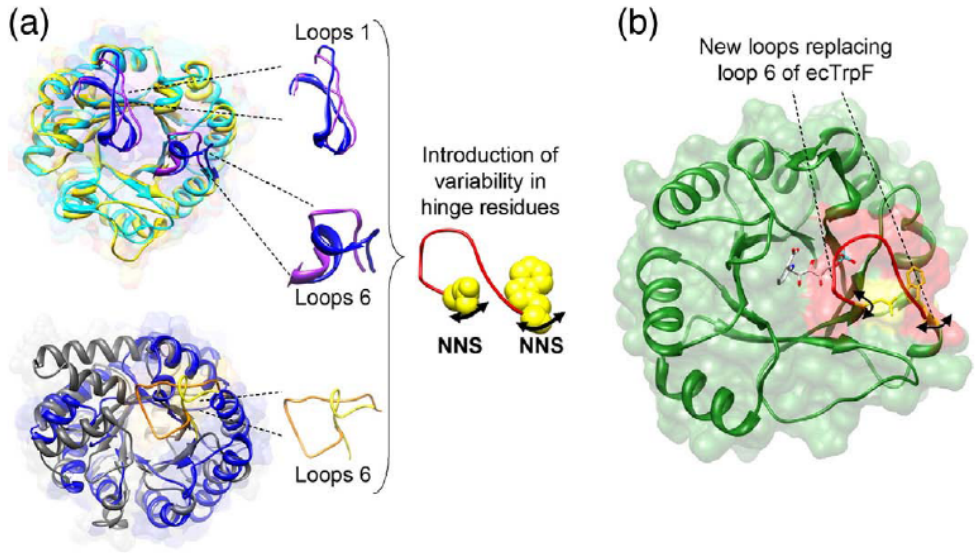

Fig. 2. Loop replacement design. (a) β/α loops from positions 1 and 6 of the scPriA (yellow), tmHisA (cyan), ecTrpA (gray) and mtMetR (blue) enzymes are illustrated. Zoom representation of selected loops is shown; the colors are a follows: blue (tmHisA), purple (scPriA), orange (ecTrpA) and yellow (mtMetR). Sequence variability was introduced using the degenerate codon NNS at both end hinge residues of each loop replacement. (b) The selected loops were used to replace the β/α loop 6 (shown in red) of ecTrpF, which is anchored by hinge residues Asn127 and Trp138. 3D structure of ecTrpF (PDB ID: 1pii) forming a complex with the product analogue rCdRP is illustrated. Docking of rCdRP was inferred from a TrpF structure in complex with rCdRP (PDB ID: 1lbm) and superimposed with the ecTrpF structure. The 3D structures were rendered using the Chimera package from UCSF.[8]

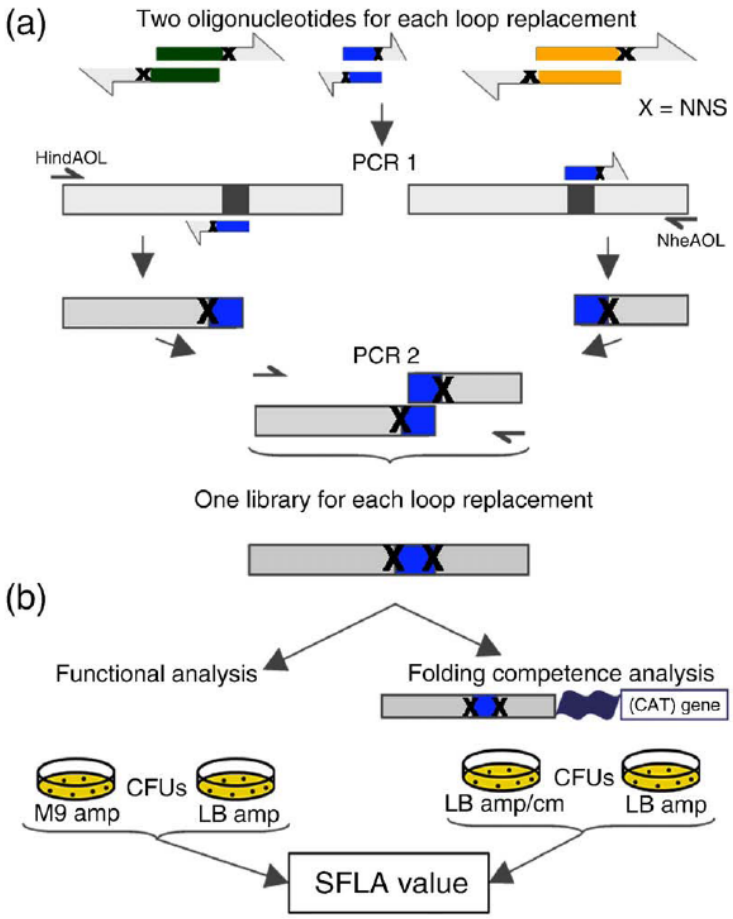

Fig. 3. Library construction and assessment of the SFLA value. (a) One pair of oligonucleotides of different sizes (from 36 bp to 69 bp), which are partially complementary (12 bp), was designed for each loop replacement. The first halves of the *trpF* gene were constructed using the oligonucleotide HindAOL as 5′-primer and the corresponding noncoding loop oligonucleotide as 3′-primer (left, PCR 1). The carboxy halves were constructed using the corresponding coding loop oligonucleotide as 5′-primer and the oligonucleotide NheAOL as 3′-primer (right, PCR 1). Libraries were amplified by overlapping-extension PCR (PCR 2) using the two PCR products from the previous reactions as templates and the oligonucleotides HindAOL and NheAOL. (b) The fraction of functional and folded variants was calculated as described in Materials and Methods. The results of the functional proficiency analysis were the first component of SFLA value, and the results of the folding competence analysis were the second component.

evolutionary relationships with TrpF,[18] these residues would be expected to participate in the movement of the original loop 6. Therefore, each loop replacement could require different hinge residues, such that conformational movements in the new protein scaffold could be coordinated to maintain catalysis. To this end, loop replacements were performed with an overlap PCR strategy that introduces sequence variability at both end residues, allowing the exploration of all 20 amino acids (Fig. 3a).

From the sequence analyses of all libraries, it was concluded that 97% of the generated variants were correctly constructed. The sequence of the hinge residues reveals that there is a true biased representation of some amino acids, which is in agreement with the theoretical distribution expected by an NNS codon (Supplementary Fig. S1).

Scoring the folding competence and functional proficiency

To estimate the folding competence, we fused the libraries to a chloramphenicol acetyl transferase gene using the system that we have previously described as a reliable folding competence reporter.[13] Briefly, if the fusion is stably expressed and soluble in bacteria, they confer resistance to chloramphenicol, thus providing a simple test for folding competence. This folding competence

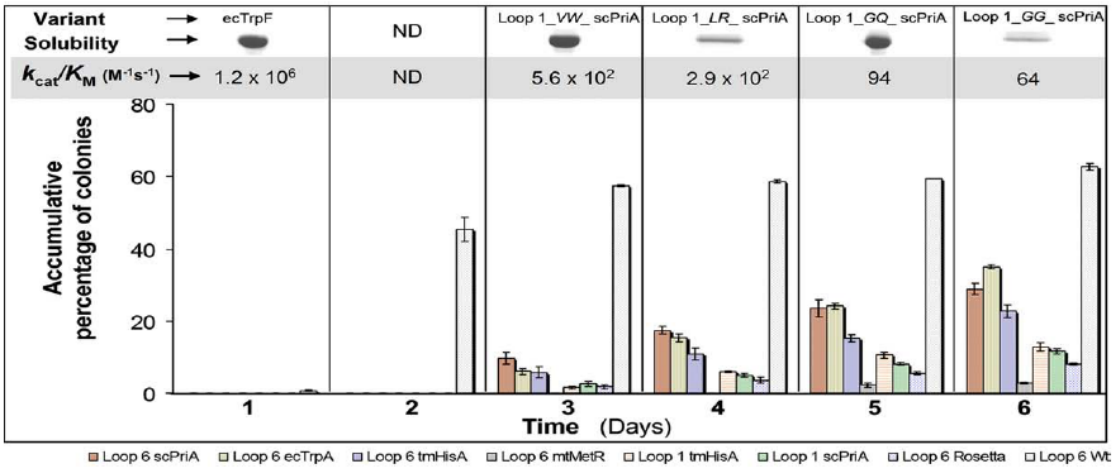

Fig. 4. Kinetics and sensitivity of the *in vivo* functional assay. The percentage values and error bars represent the accumulative average and standard deviation, respectively, of the number of colonies that appeared throughout the functional assay until 6 days. (Top) the catalytic efficiency values (k_{cat}/K_m) of selected variants from the library Loop 1 scPriA, which were isolated at different time points, are shown. The soluble protein level after overexpression and nickel-affinity purification is also shown. ND, values not determined since only library Loop 6 Wt rendered functional variants at this point in time.

selection is independent of the amount of soluble protein.[13] The fraction of folded variants of each library was estimated by comparing the number of colony-forming units (CFU) growing on plates with and without selection pressure for folding competence (Fig. 3b).

The functional proficiency of each library was scored by the ability of their variants to enable growth of the *E. coli* tryptophan auxotroph FBG-Wf (JM101 *trpF*^{4})[23] on minimal media lacking tryptophan. Hence, the fraction of functional variants was estimated by comparing the number of CFU growing on plates of minimal and rich media (Fig. 3b). The number of CFU that appeared on the selection plates was counted every 24 h for 6 days (Fig. 4). The ecTrpF enzyme complements the *trpF* deficiency overnight. Extraction and sequencing of plasmids that complemented at different time points throughout this period of evaluation, just before false positives started to appear, were used to define the functional range of variants in the library. The function conferred by these constricts was confirmed after transformation *de novo* of fresh *E. coli* FBG-Wf cells and growth on minimal media (data not shown). Interestingly, the number of CFU that appeared after every 24 h varies among the different libraries studied (Fig. 4). The lack of stochastic variation in cell growth suggests that gene amplification and point mutation of chromosomal genes[28,29] were not contributing to the observed phenotypes.

Enzyme kinetic analysis qualitatively confirmed the *in vivo* complementing experiments. Variants selected after 3–4 days of incubation in minimal media have the best k_{cat}/K_m values, while the variants selected after 5–6 days have the poorest k_{cat}/K_m values of all variants analyzed (Fig. 4). Furthermore, for at least one library, a correlation between the observed phenotypes and the decreasing levels of catalytic efficiency (k_{cat}/K_m) was observed (Fig. 4 and Table 2). Long complementing times show that our threshold for selection of true functional variants with PRA isomerase activity is at least higher than $k_{cat}/K_m = 64\ M^{-1}\ s^{-1}$. Furthermore, although the expression or solubility levels may be influencing the functional proficiency reported by our *in vivo* assay, the overexpression analysis of selected variants does not show a correlation between the solubility level and the complementing time (Fig. 4 and Supplementary Fig. 2a). Indeed, the expression of several randomly selected variants scored as not active by our functional proficiency assay but soluble by our folding competence assay suggests that differential solubility levels are not related with the functional proficiency of the variants (Supplementary Fig. 2b). We previously observed in the *E. coli* FBG-Wf strain that loss of PRA isomerase activity of different PriA enzyme variants is unrelated to insolubility problems.[23]

Significance of the SFLA value

On the basis of theoretical and statistical analyses supporting that structural changes in loops are strongly coupled to the evolutionary distance of their sequences and with their functional dependence,[4] we conducted experiments that provide data to estimate an SFLA value (Fig. 3b). This value was scored as the ratio between the fractions of functional variants to the total fractions of folded variants and was used as a metric reporting on the

Table 2. *In vitro* and *in vivo* functional analysis of selected variants

Variant[a]	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (M^{-1} s^{-1})	Complementing time (days)[b]
ecTrpF	2028	28.9	1.2×10^6	0.5
Loop 6_*QY*_ecTrpA	0.96	32.4	4.9×10^2	3
Loop 6_*GW*_scPriA	0.78	60.5	2.1×10^2	3
Loop 6_*RL*_tmHisA	1.14	82.4	2.3×10^2	3
Loop 1_*VW*_scPriA	0.84	24.8	5.6×10^2	3
Loop 1_*LR*_scPriA	1.10	63.8	2.9×10^2	4
Loop 1_*GQ*_scPriA	0.96	170.1	94	5
Loop 1_*GG*_scPriA	0.96	250.9	64	6
Loop 6_*EN*_scPriA	ND	ND	ND	~

The overall standard errors of enzyme kinetic parameters are less than 20%.

Plasmidic DNA was extracted from selected variants and used to retransform the *E. coli* strain MC1061*thiE*$^{\Delta}$ and plated on LB supplemented with ampicillin and chloramphenicol. Resistance to this latter antibiotic (Cm^R) is shown in the last column.

~, after 1 week of incubation, this variant does not complement.

[a] The N- and C-terminal hinge residues are shown in italic type.

[b] Time elapsed before the appearance of visible colonies (1 mm) on minimal medium plates. Growth was recorded every 24 h.

number of functional enzyme variants of each library. Hence, the final value determines the fraction of random mutations at the hinge residues for each loop replacement that are capable of folding and retaining sufficient PRA isomerase activity to support tryptophan synthesis and, therefore, growth on minimal media. As a reference, we probed the effect of only introducing variability at the hinge residues Asn127 and Trp138 while maintaining the wild-type loop 6 (Loop 6 Wt).

We found that the fraction of folded and functional variants depends on the loop used as replacement. The values obtained from different libraries vary between 49% and 75% for folding competence and from as little as 3% to as much as 60% for functional proficiency (Fig. 5a). The differences between the higher numbers of folded proteins compared to functional proteins imply the non-retention of PRA isomerase activity in several folded variants. The lack of activity was not a consequence of a decrease in expression levels (Supplementary Fig. 2b) and the subsequent enzyme kinetic analysis, for at least one variant showed that it is catalytically dead (Table 2). This suggests that the SFLA value reflects on the ability of loop replacement with hinge variability to retain the PRA isomerase activity on the ecTrpF scaffold.

The values obtained for Loop 6 Wt show that most of the folded variants retain PRA isomerase activity (70% folded *versus* 60% functional), resulting in an SFLA value of 0.84 (Fig. 5b). The library with the second highest value of functional variants and an SFLA value of 0.70 was Loop 6 ecTrpA, which came as a surprise since this enzyme lacks PRA isomerase activity (Fig. 5a and b). As expected, despite the high structural and sequence similarity between Loop 6 scPriA and Loop 6 tmHisA, the former, which has been related to its PRA isomerase activity,[23,24] showed a larger fraction of functional variants on the ecTrpF scaffold (Fig. 5a). Indeed, the SFLA values (0.39 and 0.28, respectively) of their cognate libraries reflected higher structure–function loop adaptability for Loop 6 scPriA than for Loop 6 tmHisA (Fig. 5b).

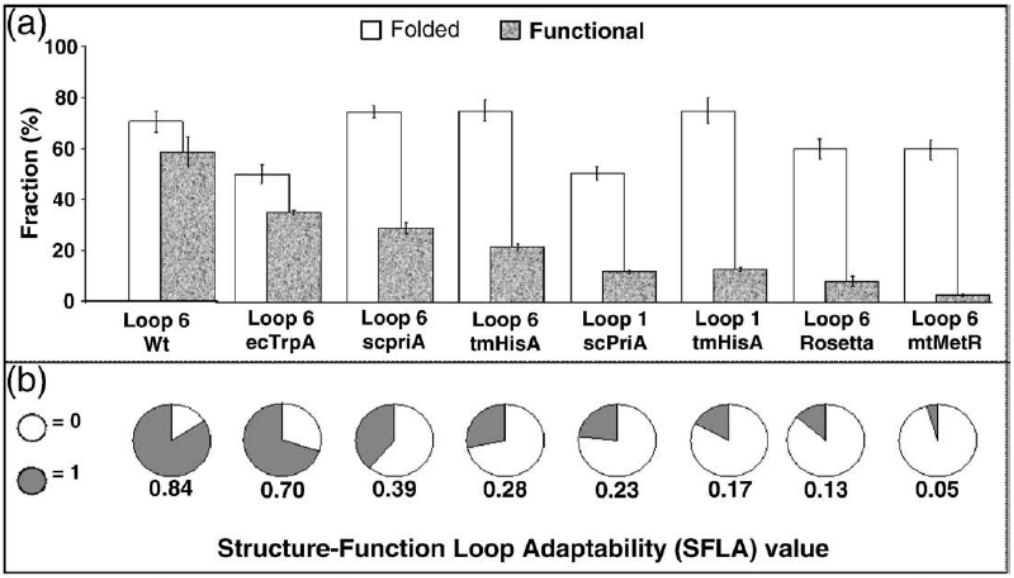

Fig. 5. Analysis of functional proficiency and folding competence and SFLA values of the libraries. (a) Fraction of folded and functional variants, which are the two components of SFLA value. The fraction values and error bars represent the average and standard deviation, respectively. (b) SFLA value (see Materials and Methods).

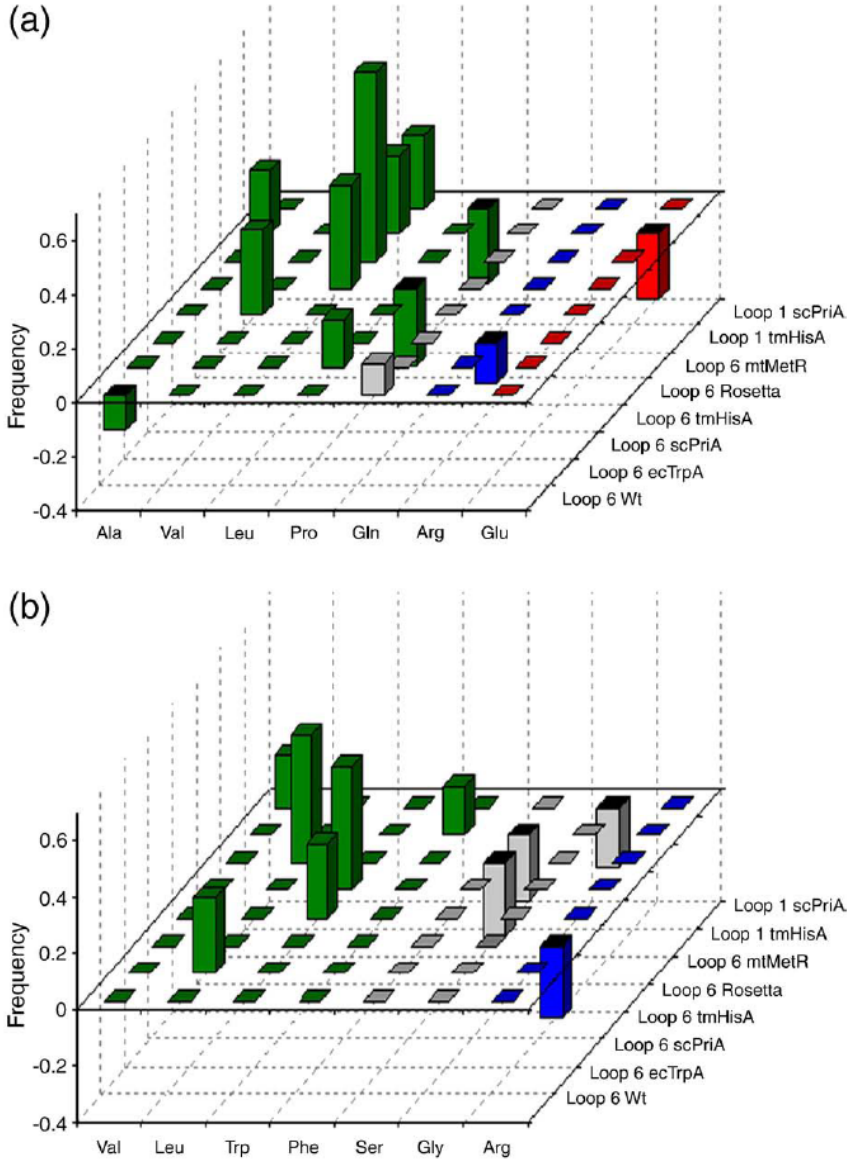

Fig. 6. Sequence analysis of the selected residues in the two hinges.. Sequence analysis of the selected residues in the two hinges. The positive or negative overrepresentation of the normalized frequency for amino acids found on the selected *versus* the nonselected conditions is shown in the amino (a) and carboxy (b) hinges. Amino acids are colored according to their physicochemical properties as follows: green, nonpolar and hydrophobic; gray, polar and uncharged; blue, polar basic; and red, polar acidic. The figure shows only the amino acids with a frequency either higher or lower (2 SD) than expected by the NNS design.

The functional relevance of the results described above is emphasized when it is noted that, despite having a high fraction of folded variants, libraries Loop 6 mtMetR, Loop 6 Rosetta, Loop 1 scPriA and Loop 1 tmHisA have a lower fraction of functional variants, with consequent lower SFLA values of 0.05, 0.13, 0.23 and 0.17, respectively (Fig. 5b). Loop 6 Rosetta, which shares 40% sequence identity with Loop 6 Wt, was compatible with the target protein backbone in folding competence terms, as shown by 60% folded variants; however, only 8% of the variants were functional, yielding a poor SFLA value of 0.13 (Fig. 5a and b).

Hinge variability for each loop replacement

The different SFLA values imply that certain hinge residues that maintain PRA isomerase activity may be

selected. To investigate if there is a pattern of functionally compatible residues in our assay, we performed a sequence analysis for each library complementing the *trpF* deficiency in minimal media. There is a true biased representation of some amino acids without selective pressure as a result of the variability introduced by the NNS codon (Supplementary Fig. S1). Therefore, the frequencies observed under selective pressure *versus* nonselective pressure were normalized per amino acid and for each library. Despite the small number of the samples sequenced (about 20 per library and for each condition), we found a statistically significant overrepresentation, indicating that some hinge residues are more compatible than others with the new loop and/or vice versa (Fig. 6). In the ecTrpF structure, the original hinge residues Asn127 and Trp138 anchor β/α loop 6, and the observed recurrence of residue selection at these positions suggests that the reorganization of each loop replacement with hinge residues is necessary. This analysis also showed a general preference for hydrophobic residues at both hinges, independently of the loop replacement, suggesting the importance of maintaining hydrophobic packing in these regions. Additionally, as suggested by the number of residues that was positively or negatively selected under functional constraints, we found that there is a more restricted acceptance of variability at the amino hinge with respect to the carboxy hinge (Fig. 6).

Enzyme kinetic and structural analyses of a set of variants

To obtain a broader understanding of the functional and structural effects on the ecTrpF scaffold caused by (i) different loop replacements and (ii) different residues at hinge positions maintaining the same loop replacement, we performed steady-state enzyme kinetics of PRA activity (Table 2) and structural studies by CD in a set of variants (Fig. 7). One of the fastest-growing variants complementing the tryptophan auxotrophy, which formed colonies after 3 days on minimal media, was selected for different libraries (Table 2). Also, four variants from library Loop 1 scPriA, which covered the window of our functional assay (3–6 days, Fig. 4), were selected. This selection was complemented by a variant from Loop 6 scPriA (Table 2) that was not able to complement the tryptophan auxotrophy but was shown to be soluble (Supplementary Fig. 2b).

The enzyme kinetic analysis revealed a minimum decrease of 4 orders of magnitude in k_{cat}/K_m with respect to the wild-type ecTrpF upon loop replacement, largely due to a reduction on the k_{cat} (Table 2). The effect of different hinge residues maintaining loop 1 from scPriA was also explored, and the enzyme kinetic analysis of the variants with glycine at both hinges (loop1_GG_scPriA) and with only one glycine (loop1_GQ_scPriA) suggests that binding to the transition state was mainly affected (Table 2). These results also show that there is a relationship between the N-terminal and C-terminal hinges, illustrating their important role in the loop structure–function adaptability.

Since the catalytic activity of variant Loop 6_*EN*_scPriA could not be detected *in vitro*, which is consistent with the lack of functional activity *in vivo* (Table 2), it is safe to state that our functional proficiency assay is able to select for soluble and nonfunctional variants independently of their

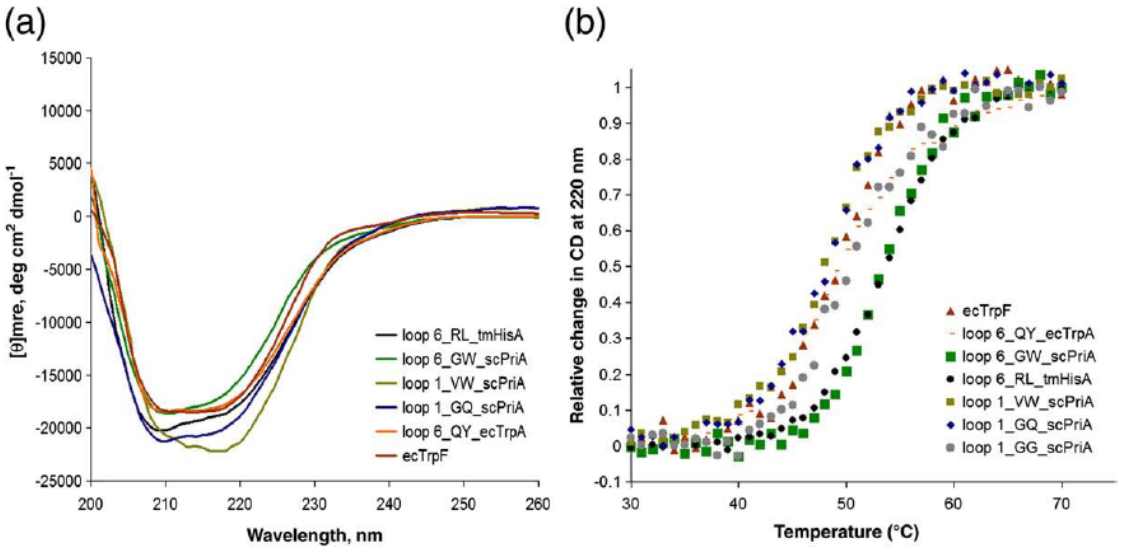

Fig. 7. Structural analysis of selected variants. (a) Far-UV CD spectra of ecTrpF and variants with different loop replacements and with different residues at their hinge positions and (b) thermal unfolding curves.

solubility levels (Supplementary Fig. 2b). These results suggest that variants with a negative functional phenotype could be explained by adoption of incorrect conformations of the catalytic active site. Indeed, K_m values of variants complementing the tryptophan auxotrophy after 5 or 6 days tended to increase, which contrasted with the observation that k_{cat} parameters throughout all the variants analyzed were similar (Table 2).

The far-UV CD spectra analyses suggest only slight structural changes at the secondary structure level as a consequence of loop replacements or modification of hinge residues maintaining the same loop (Fig. 7a and Supplementary Table S1). These slight differences in the CD spectra and in the estimated secondary structure content among the variants analyzed could reflect changes between interactions of secondary structure elements surrounding the original β/α loop 6.[13] The thermal denaturation process (Fig. 7b) showed that all variants form stable structures, with apparent thermal melting temperatures ($T_{m,app}$) that are very similar to those of ecTrpF, around 49 °C, except for the variants with loop 6 from scPriA (Loop 6_GW_scPriA) and loop 6 from tmHisA (Loop 6_RL_tmHisA) that were more thermostable by 4 °C (Supplementary Table S2). This may be related to the fact that both variants exist predominantly as dimers (Supplementary Fig. 3 and Supplementary Table 2), which is in agreement with the previous observations in TrpF from *T. maritima*, where the dimerization state increases the intrinsic thermal stability of the protein.[30]

Discussion

In a manner reminiscent to that of the H3 loop of the antibodies, the β/α loops of $(\beta/\alpha)_8$-barrels vary immensely in length, sequence and conformation (Fig. 1), and it has been suggested that their modularity is a convenient device in the evolution of $(\beta/\alpha)_8$-barrels, as semiautonomous units of critical functional groups.[31,32] Although we are not aware of systematic evidence for a similar mechanism operating in the natural history of $(\beta/\alpha)_8$-barrels, it is tempting to speculate that events resulting in loop swapping have been captured by selective pressure, as previously suggested after phylogenetic analysis.[4] In the case of ecTrpF, it is apparent that β/α loop 6 satisfies the requirements for its functional evolvability, suggesting that, in $(\beta/\alpha)_8$-barrels, the separation of their tightly packed scaffolds (stability face) from their floppy active sites (catalytic face)[33] could be simultaneously promoting structural plasticity[13] as well as functional versatility via active-site loop swapping. Indeed, loops may include structural and dynamic information that is not readily interpreted from their sequences, and these could be translated among different scaffolds. This is implied by the high SFLA values of Loop 6 ecTrpA, which has been postulated to fulfill a function similar to that of Loop 6 ecTrpF,[18] and Loop 6 scPriA, which is an enzyme with PRA isomerase activity.[23] In support of this hypothesis, it was recently observed that different loop sequences may reflect the natural selection of not only chemical properties but also dynamic modes that augment substrate specificity.[34] Our results show that the sequence of any given loop replacement is more important in obtaining functional variants and with less influence in obtaining folded variants (Fig. 5a). Thus, a relationship between loop sequences and function was more prominently found than with proper folding (Fig. 5a).

It is clear that the introduction of variability at hinge residues connecting the loops with the ecTrpF scaffold has a noticeable effect for obtaining variants that retain the PRA isomerase activity. Importantly, an arbitrary pair of connecting residues would have failed to produce a functional protein 40–97% of the time, depending on the particular loop used as replacement. Perhaps not surprisingly, in some cases, for example, the library Loop 6 Wt, the obtained SFLA value is indicative of their high structure–function adaptation (Fig. 5b). Furthermore, the higher number of fastest-growing colonies observed for this library, which only includes variability at hinge residues, supports this notion (Fig. 4). This is in accordance with the fact that loop 6 was highly replaceable by other loops without affecting the folding competence. In contrast, the concomitant loss of functional proficiency of ecTrpF suggests that the high sequence conservation of loop 6 TrpF homologues was evolved mainly under functional constraints. These constraints were probably also related in maintaining an active conformational dynamics of the active site, as was recently observed in another fold.[34]

The SFLA value obtained for Loop 6 Rosetta (0.13), in which the functional constraints were not considered, confirms the lack of functional information embedded on this loop in spite of its 40% sequence identity with Loop 6 Wt (Fig. 5b and Table 1). A comparison of this value with that obtained for Loop 6 Wt (0.84) also highlights the importance of the functional information embedded in the original sequence to retain the PRA isomerase activity of ecTrpF. Low SFLA values were also observed for Loop 6 mtMetR, which only maintains a structurally equivalent position, and Loop 1 scPriA and Loop 1 tmHisA, which lack functional and structural relationships with Loop 6 Wt (Fig. 5b). These results emphasize a recognized challenge in computational enzyme design, recently identified as the main reason for failure to conceive, a priori, an active-site architecture.[35] Within this context, our loop replacement strategy could offer a complementary approach for exploring active-site architectures,

which could then feed back on the enzyme design approaches.

Among heterologous loops, the SFLA value of Loop 6 ecTrpA indicates that it is the most adapted, structurally and functionally, to retain the PRA isomerase activity (Fig. 5b). This result is in agreement with the early suggestion that loop 6 is fulfilling an equivalent function during the catalysis in both TrpF and TrpA, closing down the binding site.[18] Despite the high structural and sequence similarity between Loop 6 scPriA and Loop 6 tmHisA (Fig. 2 and Table 1), we found that Loop 6 scPriA shows a better structure–function adaptability than Loop 6 tmHisA, which could be rationalized assuming that this loop participates in the substrate specificity of the PRA isomerase activity of scPriA,[24] while this activity is absent in tmHisA.[20] Interestingly, in a recent study, wild-type levels of PRA isomerase activity were established in tmHisA after mutations that induce a conformational change in its β/α loop 6.[21]

The selective pressure of our *in vivo* assays (Fig. 3b) reveals a functional and structural association between the hinge residues and each loop replacement (Fig. 6). Consistent with this, the motions of loop 6 in triosephosphate isomerase are controlled by their N- and C-terminal residues, and their mutation increases the appearance of nonfunctional conformations of the original loop, which may not be competent for ligand binding or catalysis[36] or induce nonproductive loop–loop interactions, drastically affecting the kinetic parameters.[37] Therefore, our results highlight the importance of variability at hinge residues to functional adaptation of loops and their probable role on the evolution of the structure–function relationship in $(\beta/\alpha)_8$-barrel enzymes. Interestingly, in another fold, the variability at the end residues of secondary structure elements, including the loops, correlates with their movement across the catalytic cycle and is being subjected to natural selective pressure.[38]

It has previously been shown that only modest levels of PRA isomerase enzymatic activity are sufficient to suppress tryptophan auxotrophy.[20,39] The catalytic efficiency (k_{cat}/K_m) of the wild-type ecTrpF is around 1.2×10^6 M^{-1} s^{-1}, and the catalytic efficiency of our most deficient variant, isolated after 6 days of growth, is in the range of 64 M^{-1} s^{-1}, that is, 5 orders of magnitude less. Assuming that complementing time *in vivo* could be related to the activity *in vitro* (Fig. 4), we decided to establish 6 days as the threshold of a true functional proficiency analysis. This window of observation avoids false positives, and it implies that the selected variants with PRA isomerase activity have a catalytic efficiency better than 64 M^{-1} s^{-1}. In addition, the sensitive assay for PRA isomerase activity is limited to substrate concentrations ≤0.8 mM[40] because of the extremely high fluorescence quantum yields of anthranilate and PRA. Recently, it has been estimated that the minimum PRA isomerase catalytic efficiency that can reliably be identified *in vivo* and confirmed *in vitro* is around 0.3 M^{-1} s^{-1}, suggesting that the lower limit for biologically relevant PRA isomerase activity in *E. coli* could be in the range of 25- to 30-fold less.[39] This figure is 7 orders of magnitude below that found for ecTrpF and 2 orders of magnitude below our most deficient variant, demonstrating that the results reported herein are biologically meaningful.

The *in vitro* analysis of a set of variants contributes to a general understanding of the molecular effects of loop swapping and hinge variability in the enzyme used as scaffold. Despite structural and sequence differences between loop replacements and hinge residues, the decrease in the k_{cat} for all analyzed variants shows a similar trend (Table 2), most likely due to disruption of the array of catalytic residues and/or the phosphate binding sites that are located nearby (Fig. 2). Additionally, we found that the range of K_m values among the analyzed variants depends on the restrictions imposed by different loop replacements and also by different hinge residues. Given the structural position and the high importance of the original loop 6 in the substrate binding and catalysis of ecTrpF, loop replacements presented in this work modified drastically the framework of the active site of ecTrpF, which, consequently, affected their kinetic parameters (Table 2). All these new variants could be selected for altered specificities without necessarily abolishing the original activity of the protein, as it was suggested early for insertions in the β/α loops of TrpF.[41]

Recently, two attempts of loop swapping in the $(\beta/\alpha)_8$-barrel have been reported. The first attempt involved the transplantation of the eight β/α loops from a phosphotriesterase (PTE) by the eight corresponding loops of the Dr0390 protein, a member of the amidohydrolase superfamily. However, the chimeric protein was largely insoluble, and the PTE activity to paraoxon could not be detected on the soluble chimeric protein only obtained after refolding.[42] In the second attempt, two β/α loops were grafted from lactonase from *Rhodococcus erythropolis* and lactonase from *Mycobacterium tuberculosis* to lactonase from *Mycobacterium avium subsp. paratuberculosis* K-10 enzyme, which are PTE-like lactonases. However, loop grafting from AhlA to MCP did not result in a change in the substrate preference, and loop grafting from PPH to MCP had no detectable activity toward the substrate of interest.[43] The authors concluded that the effect of loop grafting is still both unpredictable and uncertain and that there are some rather subtle issues that govern the folding of the chimeric protein that are not well understood.[42,43] It could be that hinge variability to adapt the functional information embedded within the β/α loops of these $(\beta/\alpha)_8$-barrels is necessary, as shown by our results.

The observation that varying the hinges of loops allows preservation of function on a large number of variants suggests a whole new strategy for protein engineering, incorporating a more divergent sequence exploration, beyond that limited to point mutations. The SFLA concept can also be used to study other types of structure–function relationships, including the replacement of other secondary structure elements, such as β-sheet or α-helix, where the introduction of sequence variability at their hinge residues may also prove to be important for obtaining folded and functional proteins.

Materials and Methods

Construction of libraries

All oligonucleotides used in this study are described in Supplementary Table S3. Two oligonucleotides, which are partially complementary (12 bp), were designed for each loop replacement: one that corresponds to the noncoding DNA strand and the other, to the coding DNA strand. In both of them, an NNS codon that replaced the hinge residues was introduced. Libraries were independently constructed by PCR using the corresponding pair of oligonucleotides for each loop replacement, as depicted in Fig. 3a. For each library, the first half of the *trpF* gene was constructed using a *trpF*-containing pDAN5 plasmid as template and the oligonucleotide Hind3AOL as 5′-primer with a HindIII restriction site and the noncoding loop oligonucleotide as 3′-primer (Fig. 3a, PCR 1). The second half was constructed using the same plasmid as template and their corresponding coding loop oligonucleotide as 5′-primer and the oligonucleotide Nhe1AOL with an NheI restriction site as 3′-primer (Fig. 3a, PCR 1). The amplification products were purified from a 1% agarose gel. Libraries were finally constructed by overlapping-extension PCR using the corresponding two PCR products from the previous reactions as templates and the oligonucleotides Hind3AOL as 5′-primer and NheAOL as 3′-primer (Fig. 3a, PCR 2). Amplified products were purified from a 1% agarose gel and then digested with HindIII and NheI, and the purified reaction products were ligated into the pDAN5 plasmid. The resulting libraries contained 10^5 to 10^6 different variants. Approximately 20 plasmids, for each library, were sequenced to confirm loop replacement and to analyze the statistical distribution of the variability introduced at both hinges.

Analysis of functional proficiency

The fraction of variants that retain PRA isomerase activity was selected by complementing the tryptophan auxotroph *E. coli* JM101*ΔtrpF*[23] in M9 minimal media. To this end, each library was used to independently transform electrocompetent cells in triplicate. Dilutions of approximately 1000–1500 viable transformed cells were spread on two different media with ampicillin (200 μg/ml): LB media agar and M9 minimal media agar. Plates were incubated at 30 °C for 6 days, and the CFU were counted every 24 h. The fraction of functional variants was calculated as the ratio of the CFU grown under selective pressure for PRA isomerase activity (M9 media) to the CFU grown without this selective pressure (LB media), and this value was the first component of the SFLA value (Fig. 3b). For each library, approximately 20 colonies were selected from the M9 plates, and their plasmids were sequenced.

Analysis of folding competence

The fraction of variants with folding competence was estimated by fusing all libraries to the chloramphenicol acetyl transferase gene as an *in vivo* folding reporter, using the method previously described.[13] These libraries, as well as the positive and negative controls for selection of folded proteins, were independently used to transform *E. coli* MC1061*ΔthiE* electrocompetent cells in triplicate. Dilutions of approximately 1000–1500 viable transformed cells were spread on two different conditions with ampicillin (200 μg/ml): LB media agar with and without 20 μg/ml of chloramphenicol. Plates were incubated at 30 °C for 18 h. The fraction of variants with folding competence was calculated as the ratio of the number of CFU grown under selective pressure for folding (LB amp plus chloramphenicol) to the number of CFU grown without this selective pressure (LB amp), and this figure was the second component of the SFLA value (Fig. 3b).

Statistical analysis of sequences

The amino acid frequencies observed at mutagenized hinge positions were scored. The plasmid sequences from the colonies grown without selective pressure for PRA isomerase activity were compared with the sequences found under this pressure for each library. The discrepancy in the abundance of each residue between these two conditions was calculated. The average and the standard deviation of the frequencies were used to determine, with 95% confidence, the negative or positive selection for specific amino acids.

Protein production

Genes encoding the selected variants were subcloned from pDAN5 into pET28b(+) plasmid and sequenced. To this end, genes were amplified by PCR using oligonucleotides 5′-GCCATACCATGGGGGAGAA-TAAGGTATGTGGC-3′ with an NcoI site (underlined) as 5′-primer and 5′-GTCCGAAAGCTTTCATTA-GTGGTGGTGGTGGTGGTGGGATCCATATG-CGCGCA-3′ with a HindIII site (underlined) as 3′-primer. The amplified product was ligated into pET28b(+). All constructs were sequenced entirely to exclude inadvertent mutations. Overexpression was performed in *E. coli* Rosetta2 cells (Novagen) after transforming the different pET28b(+) plasmids housing the encoding sequences. For each variant, 1 l of LB media supplemented with kanamycin (50 μg/ml) and chloramphenicol (25 μg/ml) was inoculated with a 5-ml preculture and incubated at 37 °C. After an OD_{600} of 0.7 was reached, expression was induced by adding 0.5 mM IPTG, and growth continued for another 14 h at 20 °C. The cells were harvested by centrifugation, for 5 min, at 4000 rpm and 4 °C.

Protein purification and steady-state enzyme kinetics

Cell pellets were resuspended in 25 ml of 10 mM potassium phosphate buffer at pH 7.6, 50 mM NaCl, 5% glycerol, 0.1 mM DTT, 0.1 mM PMSF and 2.5 mg of lysozyme, lysed by sonication (Branson Sonifier 450; 20 s, six times in 30-s intervals, 50% pulse, 4 °C) and centrifuged again (20 min, 11,000 rpm at 4 °C) to separate the soluble from the insoluble fraction of the cell extract. Variants were purified from the soluble cell fraction by loading the cell extract into a nickel Sepharose column (HisTrap FF crude, 5 ml; GE Healthcare) previously equilibrated with 50 mM potassium phosphate buffer, pH 7.6, and 300 mM NaCl. The bound His6-tagged protein was eluted by applying a linear gradient from 1 mM to 300 mM imidazole. Fractions with pure protein were pooled, and these were concentrated in the Amicon Ultra-15 system (MILLIPORE) to a final volume of 3 ml. Next, the samples were loaded on a Superdex 200 column (GE Healthcare) that was previously equilibrated with 50 mM Hepes buffer (pH 7.6) and 100 mM NaCl. Fractions with pure protein were pooled and concentrated in the Amicon Ultra-15 system to a final volume of 3 ml and dialyzed against 3× 1 l degassed 10 mM sodium phosphate buffer (pH 7.6), 1 mM ethylenediaminetetraacetic acid and 1 mM 2-mercaptoethenol at 4 °C. Protein concentrations were quantified using the Bradford method using the BIO-RAD protein assay. Michaelis–Menten enzyme kinetics of PRA isomerase activity were determined using reported protocols[16] with minor modifications.[24] Each data point represents the average of at least three independent experiments using freshly purified enzyme.

Structural studies by CD

Measurements were carried out using a J-715 spectropolarimeter (JASCO) equipped with a Peltier temperature control supplied by JASCO. Eight replicate spectra were collected on each sample to improve the signal-to-noise ratio. The far-UV CD spectra were collected from 190 to 260 nm at 25 °C. Proteins were measured at 0.3 mg/ml and dissolved in 10 mM potassium phosphate buffer at pH 7.6, 1 mM ethylenediaminetetraacetic acid and 1 mM 2-mercaptoethenol. Eight replicate spectra were collected on each sample to improve the signal-to-noise ratio. The spectra were collected in a 0. 01-cm-path-length cell. The thermal denaturation process was analyzed by measuring the change in ellipticity at 220 nm as a function of temperature. The temperature was increased at a rate of 0.3 °C min^{-1}. Thermal denaturation curves were normalized assuming a linear temperature dependence of the baselines for native and denaturated states. The apparent thermal melting temperature ($T_{m,app}$) was determined by finding a midpoint temperature between the native form (linear interpolation of the native region) and the denatured form (either the lowest point or linear interpolation of the unfolded region) on thermal unfolding curves.

Analytical gel filtration

The intermolecular associations were analyzed by size-exclusion chromatography in an Akta FPLC and a Superdex 200 column (GE Healthcare). Purified protein in an initial volume of 0.15 ml was eluted at a flow rate of 0.4 ml min^{-1} over a Superdex 200 column that was equilibrated with 50 mM Hepes buffer (pH 7.6) and 100 mM NaCl. The apparent molecular masses were determined by comparing the protein elution volumes to a calibration curve, which was obtained using proteins from Gel Filtration Standard of BIO-RAD (151-1901).

Supplementary materials related to this article can be found online at doi:10.1016/j.jmb.2011.05.027

Acknowledgements

We are indebted to Birte Hocker and Frances Arnold for useful discussions and critical reading of the manuscript and to the anonymous reviewers for their constructive comments and very critical suggestions. The authors thank Jorge Yañez, Eugenio López, Santiago Becerra and Paul Gaytán from Unidad de Síntesis y Secuenciación (Instituto de Biotecnología, Universidad Nacional Autónoma de México) for technical assistance on the synthesis of oligonucleotides and sequencing. This work was supported by Mexican Science and Technology Research Council grants 50952-Q and 83039. A.O.-L. thanks doctoral scholarship 201145 from Mexican Science and Technology Research Council.

References

1. Smith, J. M. (1970). Natural selection and the concept of a protein space. *Nature*, **225**, 563–564.
2. Bogarad, L. D. & Deem, M. W. (1999). A hierarchical approach to protein molecular evolution. *Proc. Natl Acad. Sci. USA*, **96**, 2591–2595.
3. Seibert, C. M. & Raushel, F. M. (2005). Structural and catalytic diversity within the amidohydrolase superfamily. *Biochemistry*, **44**, 6383–6391.
4. Panchenko, A. R. & Madej, T. (2005). Structural similarity of loops in protein families: toward the understanding of protein evolution. *BMC Evol. Biol.* **5**, 10.
5. De Genst, E., Silence, K., Ghahroudi, M. A., Decanniere, K., Loris, R., Kinne, J. *et al.* (2005). Strong *in vivo* maturation compensates for structurally restricted H3 loops in antibody repertoires. *J. Biol. Chem.* **280**, 14114–14121.
6. Weill, J. C. & Reynaud, C. A. (1996). Rearrangement/hypermutation/gene conversion: when, where and why? *Immunol. Today*, **17**, 92–97.
7. Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature*, **302**, 575–581.
8. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612.
9. Venkitakrishnan, R. P., Zaborowski, E., McElheny, D., Benkovic, S. J., Dyson, H. J. & Wright, P. E. (2004).

Conformational changes in the active site loops of dihydrofolate reductase during the catalytic cycle. *Biochemistry*, **43**, 16046–16055.
10. Heinis, C., Schmitt, S., Kindermann, M., Godin, G. & Johnsson, K. (2006). Evolving the substrate specificity of O^6-alkylguanine-DNA alkyltransferase through loop insertion for applications in molecular imaging. *ACS Chem. Biol.* **1**, 575–584.
11. Akiva, E., Itzhaki, Z. & Margalit, H. (2008). Built-in loops allow versatility in domain–domain interactions: lessons from self-interacting domains. *Proc. Natl Acad. Sci. USA*, **105**, 13292–13297.
12. Bornscheuer, U. (2009). Meeting report: protein design and evolution for biocatalysis. *Biotechnol. J.* **4**, 443–445.
13. Ochoa-Leyva, A., Soberon, X., Sanchez, F., Arguello, M., Montero-Moran, G. & Saab-Rincon, G. (2009). Protein design through systematic catalytic loop exchange in the (beta/alpha)8 fold. *J. Mol. Biol.* **387**, 949–964.
14. Sterner, R. & Hocker, B. (2005). Catalytic versatility, stability, and evolution of the (betaalpha)8-barrel enzyme fold. *Chem. Rev.* **105**, 4038–4055.
15. Eberhard, M., Tsai-Pflugfelder, M., Bolewska, K., Hommel, U. & Kirschner, K. (1995). Indoleglycerol phosphate synthase-phosphoribosyl anthranilate isomerase: comparison of the bifunctional enzyme from *Escherichia coli* with engineered monofunctional domains. *Biochemistry*, **34**, 5419–5428.
16. Hommel, U., Eberhard, M. & Kirschner, K. (1995). Phosphoribosyl anthranilate isomerase catalyzes a reversible amadori reaction. *Biochemistry*, **34**, 5429–5439.
17. Patrick, W. M. & Blackburn, J. M. (2005). *In vitro* selection and characterization of a stable subdomain of phosphoribosylanthranilate isomerase. *FEBS J.* **272**, 3684–3697.
18. Wilmanns, M., Hyde, C. C., Davies, D. R., Kirschner, K. & Jansonius, J. N. (1991). Structural conservation in parallel beta/alpha-barrel enzymes that catalyze three sequential reactions in the pathway of tryptophan biosynthesis. *Biochemistry*, **30**, 9161–9169.
19. Henn-Sax, M., Thoma, R., Schmidt, S., Hennig, M., Kirschner, K. & Sterner, R. (2002). Two (betaalpha)(8)-barrel enzymes of histidine and tryptophan biosynthesis have similar reaction mechanisms and common strategies for protecting their labile substrates. *Biochemistry*, **41**, 12032–12042.
20. Jurgens, C., Strom, A., Wegener, D., Hettwer, S., Wilmanns, M. & Sterner, R. (2000). Directed evolution of a (beta alpha)8-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc. Natl Acad. Sci. USA*, **97**, 9925–9930.
21. Claren, J., Malisi, C., Hocker, B. & Sterner, R. (2009). Establishing wild-type levels of catalytic activity on natural and artificial (beta alpha)8-barrel protein scaffolds. *Proc. Natl Acad. Sci. USA*, **106**, 3704–3709.
22. Barona-Gomez, F. & Hodgson, D. A. (2003). Occurrence of a putative ancient-like isomerase involved in histidine and tryptophan biosynthesis. *EMBO Rep.* **4**, 296–300.
23. Wright, H., Noda-Garcia, L., Ochoa-Leyva, A., Hodgson, D. A., Fulop, V. & Barona-Gomez, F. (2008). The structure/function relationship of a dual-substrate (betaalpha)8-isomerase. *Biochem. Biophys. Res. Commun.* **365**, 16–21.
24. Noda-Garcia, L., Camacho-Zarco, A. R., Verdel-Aranda, K., Wright, H., Soberon, X., Fulop, V. & Barona-Gomez, F. (2010). Identification and analysis of residues contained on beta→alpha loops of the dual-substrate (beta alpha)8 phosphoribosyl isomerase A specific for its phosphoribosyl anthranilate isomerase activity. *Protein Sci.* **19**, 535–543.
25. Liu, Y. & Kuhlman, B. (2006). RosettaDesign server for protein design. *Nucleic Acids Res.* **34**, W235–W238.
26. Gerstein, M., Lesk, A. M. & Chothia, C. (1994). Structural mechanisms for domain movements in proteins. *Biochemistry*, **33**, 6739–6749.
27. Sun, J. & Sampson, N. S. (1998). Determination of the amino acid requirements for a protein hinge in triosephosphate isomerase. *Protein Sci.* **7**, 1495–1505.
28. Hastings, P. J., Slack, A., Petrosino, J. F. & Rosenberg, S. M. (2004). Adaptive amplification and point mutation are independent mechanisms: evidence for various stress-inducible mutation mechanisms. *PLoS Biol.* **2**, e399.
29. Hendrickson, H., Slechta, E. S., Bergthorsson, U., Andersson, D. I. & Roth, J. R. (2002). Amplification–mutagenesis: evidence that "directed" adaptive mutation and general hypermutability result from growth with a selected gene amplification. *Proc. Natl Acad. Sci. USA*, **99**, 2164–2169.
30. Thoma, R., Hennig, M., Sterner, R. & Kirschner, K. (2000). Structure and function of mutationally generated monomers of dimeric phosphoribosylanthranilate isomerase from *Thermotoga maritima*. *Structure*, **8**, 265–276.
31. Babbitt, P. C. & Gerlt, J. A. (1997). Understanding enzyme superfamilies. Chemistry as the fundamental determinant in the evolution of new catalytic activities. *J. Biol. Chem.* **272**, 30591–30594.
32. Gerlt, J. A. & Babbitt, P. C. (2001). Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu. Rev. Biochem.* **70**, 209–246.
33. Hocker, B., Jurgens, C., Wilmanns, M. & Sterner, R. (2001). Stability, catalytic versatility and evolution of the (beta alpha)(8)-barrel fold. *Curr. Opin. Biotechnol.* **12**, 376–381.
34. Peng, T., Zintsmaster, J. S., Namanja, A. T. & Peng, J. W. (2007). Sequence-specific dynamics modulate recognition specificity in WW domains. *Nat. Struct. Mol. Biol.* **14**, 325–331.
35. Baker, D. (2010). An exciting but challenging road ahead for computational enzyme design. *Protein Sci.* **19**, 1817–1819.
36. Kempf, J. G., Jung, J. Y., Ragain, C., Sampson, N. S. & Loria, J. P. (2007). Dynamic requirements for a functional protein hinge. *J. Mol. Biol.* **368**, 131–149.
37. Wang, Y., Berlow, R. B. & Loria, J. P. (2009). Role of loop–loop interactions in coordinating motions and enzymatic function in triosephosphate isomerase. *Biochemistry*, **48**, 4548–4556.
38. Henzler-Wildman, K. A., Lei, M., Thai, V., Kerns, S. J., Karplus, M. & Kern, D. (2007). A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*, **450**, 913–916.
39. Patrick, W. M. & Matsumura, I. (2008). A study in molecular contingency: glutamine phosphoribosyl-

pyrophosphate amidotransferase is a promiscuous and evolvable phosphoribosylanthranilate isomerase. *J. Mol. Biol.* **377**, 323–336.
40. Leopoldseder, S., Claren, J., Jurgens, C. & Sterner, R. (2004). Interconverting the catalytic activities of (betaalpha)(8)-barrel enzymes from different metabolic pathways: sequence requirements and molecular analysis. *J. Mol. Biol.* **337**, 871–879.
41. Urfer, R. & Kirschner, K. (1992). The importance of surface loops for stabilizing an eightfold beta alpha barrel protein. *Protein Sci.* **1**, 31–45.
42. Xiang, D. F., Kolb, P., Fedorov, A. A., Meier, M. M., Fedorov, L. V., Nguyen, T. T. *et al.* (2009). Functional annotation and three-dimensional structure of Dr0930 from *Deinococcus radiodurans*, a close relative of phosphotriesterase in the amidohydrolase superfamily. *Biochemistry*, **48**, 2237–2247.
43. Chow, J. Y., Wu, L. & Yew, W. S. (2009). Directed evolution of a quorum-quenching lactonase from *Mycobacterium avium* subsp. *paratuberculosis* K-10 in the amidohydrolase superfamily. *Biochemistry*, **48**, 4344–4353.

Capítulo 3. **Rediseño del sitio activo de TrpF basado en el conocimiento de la evolución natural de una familia de barriles $(\beta/\alpha)_8$ isomerasas.**

La reacción de isomerización de PRA catalizada por TrpF requiere de un ácido general para protonar a el oxígeno del anillo de la furanosa y una base general para desprotonar el intermediario de la base de Schiff del cuerpo de la ribosa (Figura 1) (*18*). Dado que los mecanismos de reacción de TrpF y PriA son similares, y que los residuos catalíticos están ubicados en posiciones que podrían ser equivalentes, decidimos construir variantes de TrpF que contuvieran la ubicación estructural e identidad de los residuos catalíticos de PriA, y medir si preservaban la actividad de PRA isomerasa. Además decidimos investigar si estas modificaciones pueden hacer que TrpF obtenga la nueva actividad de ProFAR isomerasa (al tener una constelación catalítica similar a la de PriA), o si es necesario un rearreglo adicional del sitio activo (como el intercambio de asas).

Los residuos catalíticos de TrpF son la Cys7 (base general) y el Asp126 (ácido general) (*18*) y de PriA son el Asp11 (base general) y el Asp130 (ácido general) (*20*) (Figura 1). Adicionalmente en PriA se ha observado que el residuo Thr166 es indispensable tanto para la actividad de PRA isomerasa como la de ProFAR isomerasa (*20*). Por otro lado, la enzima HisA utiliza los residuos catalíticos Asp11 y Asp130 al igual que PriA y se ha sugerido que el residuo Thr166 es importante para mantener una buena actividad de ProFAR isomerasa (*18*). En la Figura 1 se muestra la reacción enzimática catalizada por estas enzimas. Basados en esto, decidimos que el rediseño del sitio activo de TrpF además de tomar en cuenta los dos aspárticos catalíticos de PriA también debía de incluir el residuo Thr166. Este rediseño del sitio activo se abordó desde diferentes estrategias (Figura 2).

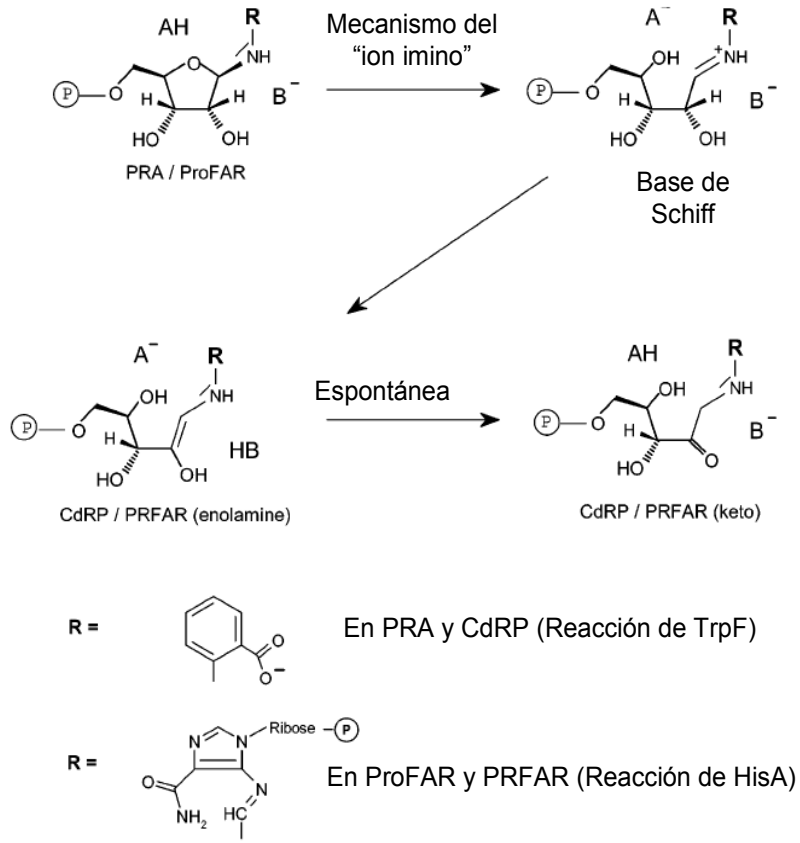

Figura 1. Mecanismo propuesto del rearreglo de Amadori que cataliza TrpF y HisA. El ácido general AH es el Asp126 (TrpF) o el Asp130 (HisA y PriA). La base general B^- es la Cys7 (TrpF) o el Asp11 (HisA y PriA). El mecanismo de reacción más aceptado involucra la catálisis general ácido-base seguida de la isomerización espontánea de una aminoaldosa fosforilada a su correspondiente aminocetosa. En el primer paso, el oxígeno del anillo de la furanosa de PRA/ProFAR es protonado por el ácido general AH, formando como intemediario una base de Schiff. Posteriormente el protón es abstraído del C2´ de la ribosa por la base general B^- para formar la enolamina de CdRP/PRFAR. La conversión subsecuente de la enolamina de CdRP/PRFAR a la forma ceto ocurre espontáneamente. Para PriA el grupo R puede ser PRA/ProFAR, debido a que esta enzima cataliza ambas reacciones enzimáticas.

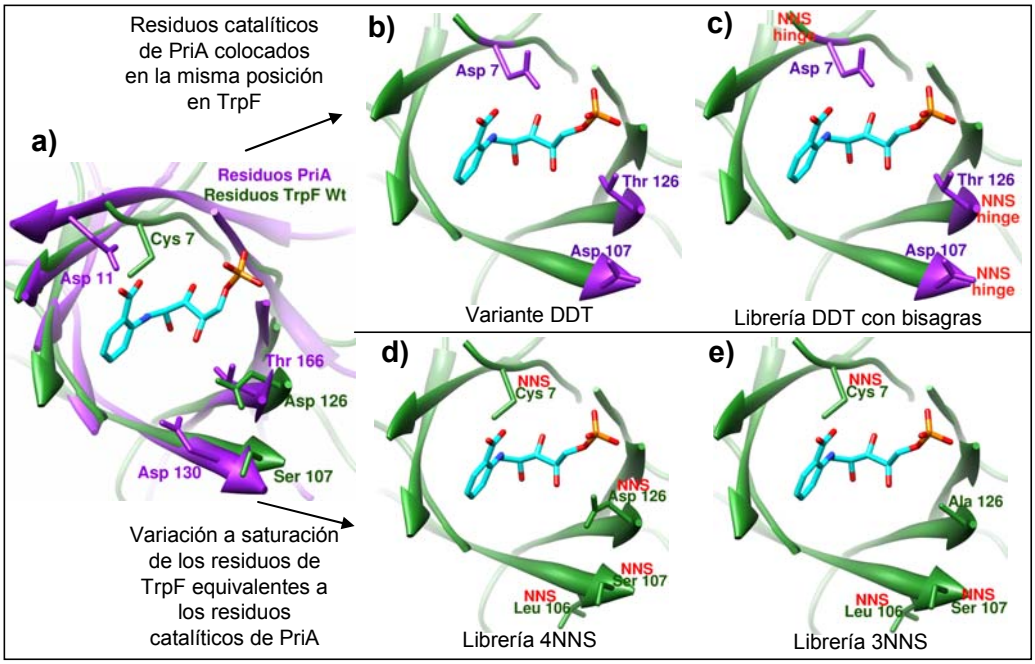

Figura 2. Descripción de las estrategias usadas para el rediseño del sitio activo de TrpF. **a)** Sobreposición estructural de TrpF (PDB: 1pii) y PriA (PDB: 2vep) donde se muestran los residuos implicados en la actividad catalítica de ambas enzimas. **b)** Modelo que ilustra la posición de los residuos Asp 7, Asp107 y Thr 126 de la variante DDT. **c)** Modelo que además de mostrar las posiciones de los residuos Asp 7, Asp107 y Thr 126 también ilustra los residuos bisagra ("hinge") que fueron mutados a saturación con el codon NNS. **d)** Ubicación estructural de los 4 residuos mutados a saturación de la librería 4NNS: Cys 7, Leu 106, Ser 107 y Asp 126. **e)** Ubicación estructural de los 3 residuos mutados a saturación de la librería 3NNS: Cys 7, Leu 106 y Ser 107. Todos los modelos se hicieron utilizando como templado la estructura de TrpF del PDB 1pii. También se muestra el análogo del producto de la reacción de PRA isomerasa (rCDRP), el cual se tomó de TrpF de *Thermotoga maritima* (PDB: 1lbm).

La primera estrategia consistió en colocar los residuos involucrados en la catálisis de PriA en la misma posición en TrpF, construyéndose la variante DDT (Figura 2b). Dado que para mantener la actividad de TrpF puede ser necesario un rearreglo de estos residuos y de las asas β/α que se ubican sobre ellos (*16, 27*), decidimos mutar a saturación los residuos correspondientes a cada una de las bisagras ("hinges") amino de sus respectivas asas 1, 5 y 6, construyéndose la librería DDT con bisagras (Figura 2c). En la segunda estrategia optamos por mutar a saturación los residuos catalíticos Cys7 y Asp126 los cuales están ubicados en una posición equivalente al Asp11 y Thr166 de PriA, respectivamente (Figura 2a). Además, mutamos los residuos Leu106 y

Ser107, que aunque no están involucrados en la catálisis de TrpF, serían equivalentes a la posición estructural del residuo Asp130 de PriA (Figura 2a). De esta manera se diseñó la librería 4NNS (Figura 2d). Como este diseño permite que la combinación de los residuos originales vuelva a ser seleccionada en la posiciones 7 y 126, decidimos construir otra librería que no permitiera dicha selección. La nueva librería se denominó 3NNS y sólo difiere de la 4NNS en que se cambió al Asp126 por alanina y no a saturación (Figura 2e). De esta manera este diseño ya no permite la selección del residuo catalítico Asp126 de TrpF.

Análisis de la actividad de PRA isomerasa

Las cuatro construcciones/librerías descritas en la Figura 2 se evaluaron por su capacidad para complementar la auxotrofia del gen de TrpF en la cepa JM101Δ*trpF*, dado que se ha observado que una cinética de crecimiento de esta cepa en medio mínimo complementada con la enzima TrpF, o con mutantes de la misma, constituye un indicador confiable de la existencia de actividad catalítica de PRA isomerasa (*18, 27*). De la variante DDT (Figura 2b) y la librería DDT con bisagras (Figura 2c) no se obtuvieron mutantes que complementaran dicha actividad.

Para la librería 4NNS encontramos 8 mutantes que complementaron la actividad de PRA isomerasa (Tabla 1). Todas ellas tienen de nuevo los residuos originales en las posiciones Cys7 y Asp126 y sólo presentan variabilidad en las otras dos posiciones que fueron mutadas (Leu106 y Ser107). A pesar de que los residuos Leu106 y Ser107 no funcionan como catalíticos, nuestros resultados indican que sí tienen una ligera influencia en la actividad enzimática de TrpF, esto debido a que las mutantes seleccionadas complementaron más lentamente la auxotrofía que la enzima silvestre. Probablemente estos residuos influyen en la correcta orientación de los dos residuos catalíticos (Figura 2a).

Tabla 1. Secuencias de las mutantes de la librería 4NNS que complementaron la auxotrofía del gen de TrpF.

No. de variante	Residuo en la posición Cys7	Residuo en la posición Leu106	Residuo en la posición Ser107	Residuo en la posición Asp126
1831	*Cys*	*Leu*	*Ser*	*Asp*
1832	Cys	Gln	Pro	Asp
1833	Cys	Pro	Arg	Asp
1834	*Cys*	*Leu*	*Ser*	*Asp*
1448	*Cys*	*Leu*	*Ser*	*Asp*
1449	*Cys*	*Leu*	*Ser*	*Asp*
5660	Cys	Val	Ala	Asp
5661	Cys	Val	Ala	Asp

De la librería 3NNS encontramos sólo una variante que complementó la auxotrofía de TrpF y por lo tanto mantiene la actividad de PRA isomerasa (Figura 3c). Esta variante fue denominada CDA (Figura 3b).

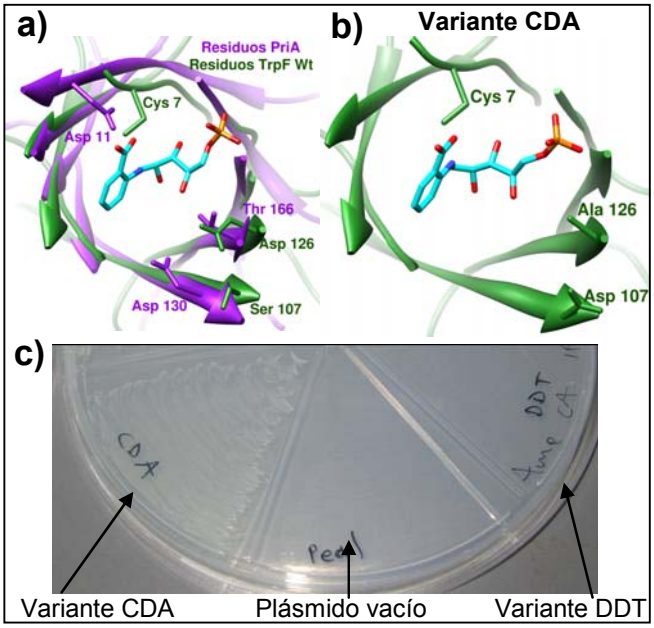

Figura 3. Características de la variante CDA. **a)** Empalme estructural entre TrpF (PDB: 1pii) y PriA (PDB: 2vep). **b)** Ubicación de los residuos seleccionados en la variante CDA (Cys7, Asp107 y Ala126). **c)** Complementación de la auxotrofía del gen de TrpF en medio mínimo sólido por la variante CDA, el plásmido vacío y la variante DDT.

Rediseño de la variante CDA

Como previamente habíamos sugerido la importancia de la Thr166 para la actividad de PRA isomerasa de PriA (*20*) decidimos colocar este residuo en la posición equivalente en la variante CDA. De esta manera hicimos la mutación Ala126Thr obteniendo la variante CDT (Cys7, Asp107 y Thr126). Al comparar la velocidad de

complementación de CDT con CDA encontramos que CDT complementa más rápidamente la actividad de PRA isomerasa (Figura 4a), a pesar de que ambas se encuentran solubles en cantidades muy similares a las de la enzima silvestre (Figura 4b). En la Figura 4a y 4b también se muestra que la nula complementación de la variante DDT no se debió a problemas con su solubilidad.

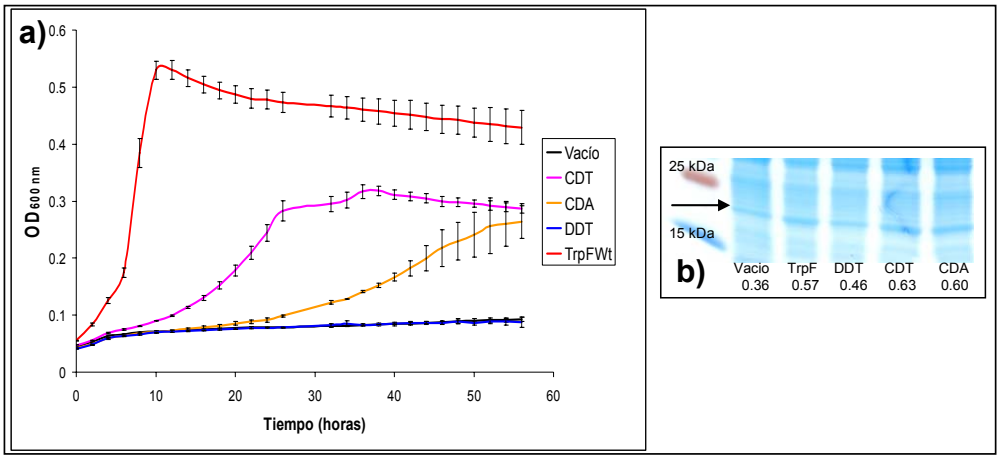

Figura 4. Análisis de complementación y niveles de solubilidad de diferentes variantes. **a)** Cinéticas de crecimiento de la cepa MC1061*ΔtrpF* en medio mínimo líquido complementada con diferentes variantes. **b)** Extracto total soluble de las variantes analizadas en la cinética de crecimiento. La flecha indica la región de las proteínas en análisis y el valor equivale a las unidades de intensidad de la banda correspondiente a cada proteína. El plásmido vacío tiene 1.3 veces mayor cantidad de proteína total que las demás muestras.

Análisis de los posibles residuos funcionales de las variantes CDA y CDT

Para demostrar el rol funcional de los nuevos residuos se hizo una mutagénesis por alanina de cada uno de ellos construyéndose las mutantes descritas en la Tabla 2. De las tres mutaciones hechas en la variante CDT sólo la mutante ADT perdió totalmente la actividad mientras que las mutantes CDA y CAT se vieron medianamente afectadas (Figura 5a), aún cuando ambas son igual o más solubles que la variante CDT (Figura 5c).

Tabla 2. Mutaciones puntuales de las variantes CDT y CDA

Variante original	Descripción de la mutación	Nombre de la mutante
CDT	Cys7Ala	ADT
	Asp107Ala	CAT
	Thr126Ala	CDA
CDA	Cys7Ala	ADA
	Asp107Ala	CAA

Al hacer la mutación Cys7Ala en la variante CDA (mutante ADA) encontramos que se perdió totalmente la actividad de PRA isomerasa (Figura 5b), lo cual también se observó al hacer la misma mutación en la variante CDT (Figura 5a, mutante ADT). Al comparar la mutante CAA (la cual mantiene la actividad de PRA isomerasa) con las mutantes ADA y ADT (que no mantienen la actividad) se observó que las tres se expresan de forma soluble en cantidades muy similares, lo cual sugiere que la solubilidad no es la responsable de la pérdida de actividad por la mutación Cys7Ala (Figura 5c).

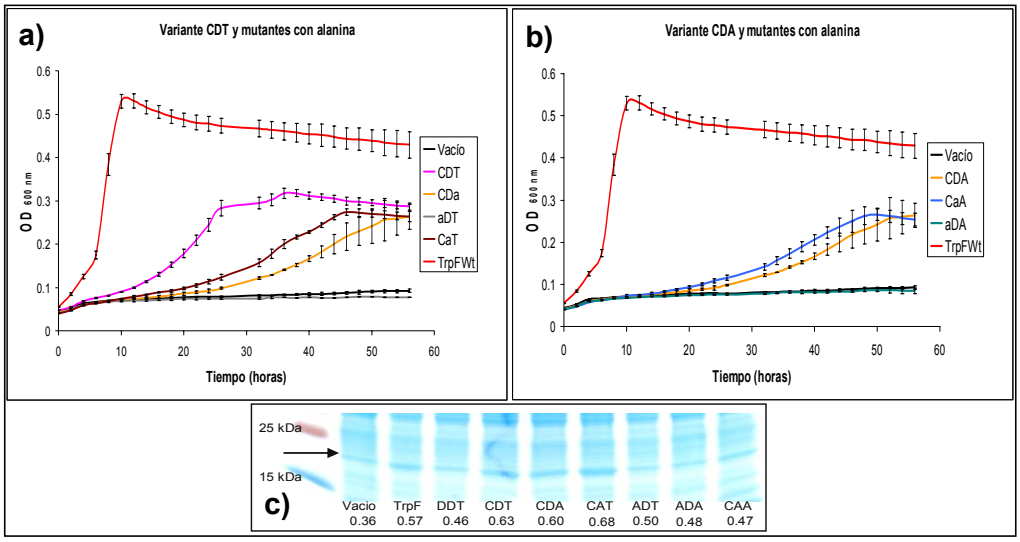

Figura 5. Análisis de complementación y niveles de solubilidad de diferentes variantes. **a y b)** Cinéticas de crecimiento de la cepa MC1061*ΔtrpF* en medio mínimo líquido complementada con diferentes variantes. **c)** Extracto total soluble de las variantes analizadas en la cinética de crecimiento. La flecha indica la región de las proteínas en análisis y el valor equivale a las unidades de intensidad de la banda

correspondiente a cada proteína. El plásmido vacío tiene 1.3 veces mayor cantidad de proteína total que las demás muestras.

Análisis de la función de ProFAR isomerasa

Debido a que las variantes CDA y CDT tienen los residuos catalíticos en posiciones equivalentes a los de PriA, también podrían estar llevando a cabo la actividad de ProFAR isomerasa, por lo cual se evaluó su habilidad para complementar la auxotrofía del gen de HisA en la cepa de selección *E. coli* HFRG6 (*20*). También se evaluaron todas las construcciones descritas en la Figura 2. Los resultados obtenidos indican que ninguna de todas estas variantes/librerías complementa la auxotrofía del gen HisA, descartando así que alguna variante posea la actividad de ProFAR isomerasa.

Mecanismo catalítico propuesto para las variantes CDA y CDT

Con el objetivo de proponer un mecanismo catalítico para las nuevas variantes, se elaboraron los modelos de su estructura tridimensional. Ambos modelos fueron idénticos y sólo cambió la identidad de la cadena lateral del residuo 126 ya sea por alanina (CDA) o por treonina (CDT), por lo que decidimos ejemplificar los resultados experimentales con el modelo de la variante CDT (Figura 6b). Como era de esperarse, en el modelo de la estructura de la variante CDT se puede observar que la Cys7 mantiene la misma posición que en la estructura original de TrpF y se encuentra en una posición equivalente a la del Asp11 de PriA (Figuras 6a y 6b). La pérdida de actividad causada al hacer la mutación Cys7Ala tanto en la variante CDT como en la CDA sugiere que este residuo sigue siendo la base general como lo era en la enzima original (*18*).

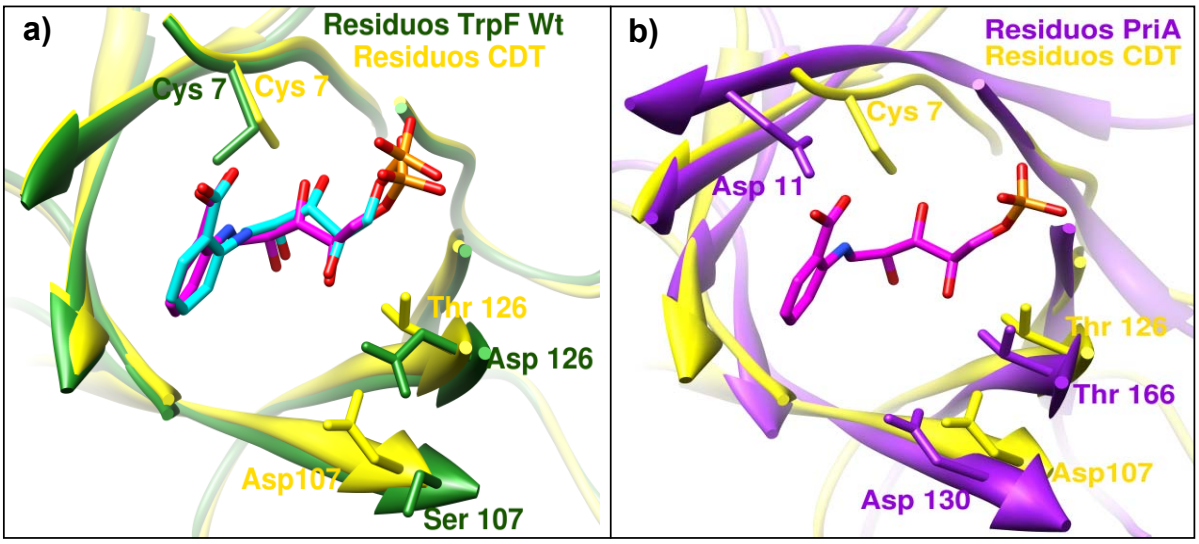

Figura 6. Modelos estructurales de la variante CDT. Los modelos se construyeron utilizando el paquete UCSF Chimera (*44*) y se optimizaron con el servidor RosettaDesign (*36*). Los modos vibracionales de baja frecuencia se obtuvieron utilizando el servidor ElNemo (*45*). En las figuras se muestra la sobreposición del modelo obtenido para la variante CDT con: **a)** TrpF (PDB: 1pii) y **b)** PriA (PDB: 2vep).

Asignar qué residuo funge como el ácido general es más complicado debido a que tanto la mutación Asp107Ala en la variante CDA como las mutaciones Asp107Ala y Thr126Ala en la variante CDT redujeron la actividad de PRA isomerasa pero no la eliminaron. Esto sugiere que ambos residuos están cumpliendo un rol importante pero que no son absolutamente necesarios. Lo que si es importante señalar es que la actividad enzimática parece estarse viendo favorecida al mantener la combinación del Asp107 con la Thr126 (variante CDT), mostrando una complementación cercana a la observada para TrpF pero lejos de las demás mutantes (Figura 5a y b). Es más factible que el Asp107 sea el ácido general, debido a que es poco probable que el grupo hidroxilo de la cadena lateral de la Thr126 pueda proveer un protón para la catálisis (*18*). Además, la ubicación de los residuos Asp107 y Thr126 en el modelo de la variante CDT es muy similar a la posición que en PriA ocupan el Asp130 y la Thr166 respectivamente (Figura 6b). Esto sugiere que las variantes CDT y CDA pueden tener un mecanismo catalítico similar al de PriA, en donde el Asp107 (equivalente al Asp130 de PriA) estaría cumpliendo el rol del ácido general. Mientras que el rol funcional de la Thr126 de la variante CDT (equivalente a la Thr166 de PriA) se encuentra aún sin aclararse, pero como también ha sido observado en PriA (*20*), parecería que este

residuo está involucrado de alguna manera en la catálisis y que aunque en la variante CDT no es esencial como lo es en PriA, si es importante para mejorar la actividad de PRA isomerasa de CDT, lo cual la hace diferente de la variante CDA. Adicionalmente, el análisis de modos vibracionales de baja frecuencia de la estructura de TrpF obtenido por el servidor de ElNemo (*45*) muestra que la hebra β5 (que es donde se encuentra el Asp107) posee una tendencia a moverse hacia el centro del barril, lo cual favorecería el acercamiento de este residuo y su participación como el nuevo ácido general tanto en la variante CDA como en CDT (datos no mostrados).

Pero ¿como explicar que está pasando con las mutantes que no tienen el Asp107 (CAA y CAT) que sería el ácido general y que siguen siendo activas? La complementación de actividad de estas mutantes podría estarse debiendo a un fenómeno de catálisis asistida por sustrato. Este fenómeno ha sido previamente sugerido para variantes de las enzimas PriA y HisA las cuales carecen del residuo que funge como el ácido general y en las que se argumenta que la catálisis puede estarse dando mediante la ayuda del grupo carboxilato de PRA el cual estaría supliendo la ausencia del ácido general (*20*) y por lo tanto mantienen una actividad vestigial. En este contexto en TrpF de *Thermotoga maritima* se encontró que aún después de mutarle el Asp126 (ácido general) mantenía una actividad vestigial (*18*), lo cual también sugiere un fenómeno de catálisis asistida por sustrato. Con base en estas observaciones podemos sugerir que la actividad encontrada en las mutantes CAA y CAT puede ser explicada por una catálisis asistida por sustrato.

Mediante la estrategia de imitar el sitio activo de PriA en TrpF no es posible obtener la actividad de ProFAR isomerasa pero si es posible mantener la actividad original de PRA isomerasa. Tomando en cuenta estos resultados decidimos explorar si al intercambiar algunas asas de TrpF por las de PriA se puede lograr este cambio de actividad. También decidimos que nuestro diseño de intercambio de las asas 1, 5 y 6 de TrpF por las equivalentes de PriA debe hacerse sobre dos moldes de TrpF: uno que mantenga la posición original de sus residuos catalíticos y el otro que posea los residuos equivalentes de PriA (como se analizó en este capítulo). De esta manera estamos dando oportunidad de que tanto los residuos catalíticos al estilo TrpF como al estilo PriA puedan ser explorados funcionalmente en conjunto con el intercambio de

asas, todo ello con el objetivo final de explorar como obtener la actividad de ProFAR isomerasa en TrpF.

Capítulo 4. **Intercambio de las asas β/α 1, 5 y 6 de TrpF por las asas equivalentes de PriA**

Diseño de la estrategia del intercambio de asas

Con el objetivo de explorar *in vivo* la actividad de ProFAR isomerasa en TrpF y a la par de mantener su actividad de PRA isomerasa mediante la estrategia del intercambio y adaptación de las asas β/α, decidimos intercambiarle las asas 1, 5 y 6 a TrpF por las equivalentes de PriA. Se seleccionaron estas asas por ser las que tienen las principales diferencias estructurales entre PriA y HisA, por lo que pueden estar relacionadas tanto en la adquisición de la actividad de PRA isomerasa como en el mantenimiento de la actividad de ProFAR isomerasa de PriA (*20*).

Las tres asas de TrpF se intercambiaron en conjunto con la introducción de cuatro sitios de variabilidad NNS: dos en la bisagra amino del asa 1 y uno para cada una de las bisagras amino del asa 5 y 6. De los dos sitios NNS introducidos para el asa 1, uno corresponde a la posición equivalente al residuo catalítico Cys7 y el segundo a la bisagra de esta asa (Figura 7).

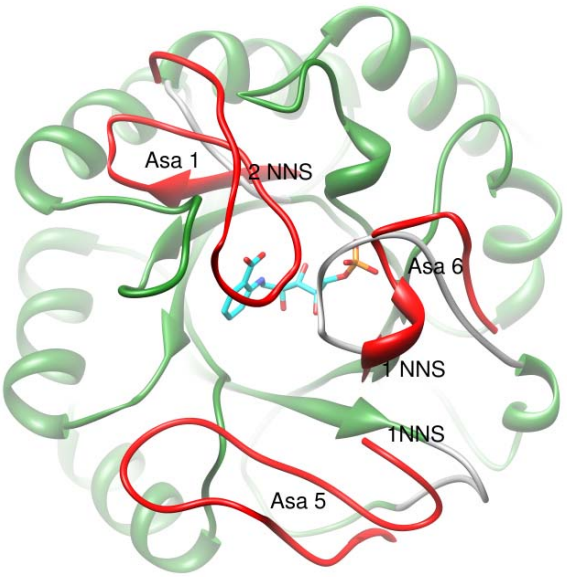

Figura 7. Estructura de TrpF (verde) con la sobreposición de las asas 1, 5 y 6 de PriA (rojo). Las asas 1, 5 y 6 de TrpF están señaladas en color gris. También se muestra la ubicación de los sitios de introducción de variabilidad (NNS).

Tomando en cuenta los resultados obtenidos al rediseñar el sitio activo de TrpF (Capítulo 3) decidimos insertar las 3 asas en conjunto con la variabilidad de las bisagras en los dos moldes de TrpF (Figura 8). De esta manera para cada molde de TrpF se construyó una librería

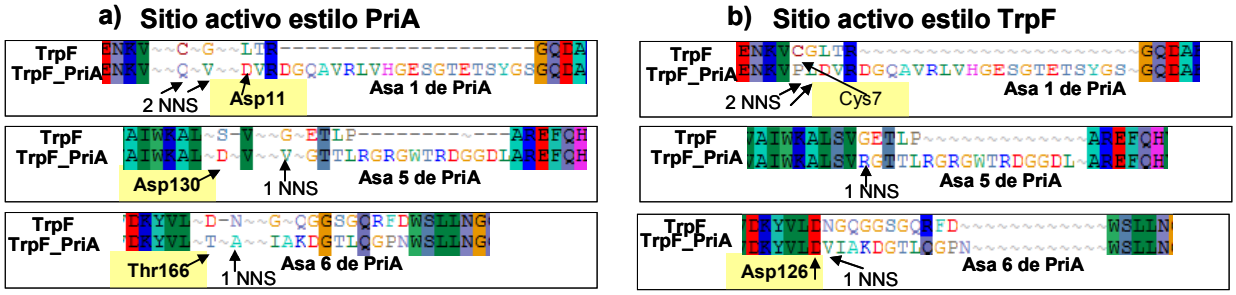

Figura 8. Ilustración a nivel de secuencia primaria del intercambio de asas y los cambios de secuencia realizados en los dos moldes de TrpF. **a)** TrpF con los residuos catalíticos al estilo PriA (Asp11, Asp130 y Thr166). **b)** TrpF con los residuos catalíticos al estilo TrpF (Cys7 y Asp126).

Análisis de las actividades de PRA y de ProFAR isomerasa

La capacidad funcional de las dos librerías se analizóa *in vivo* por la complementación de la auxotrofía tanto de la actividad de PRA isomerasa como de la actividad de ProFAR isomerasa en las respectivas cepas de selección. De la librería con los residuos catalíticos al estilo PriA (Figura 8a), la cual tiene un tamaño teórico de 1.04 x 10^6 variantes, se aislaron tres variantes complementando las diferentes auxotrofías: New4p y NewJUN con ambas actividades (PRA y ProFAR isomerasa) y NewF con una sola actividad (PRA isomerasa), mientras que de la librería con los residuos catalíticos al estilo TrpF (Figura 8b), cuyo tamaño es el mismo que la anterior, se aislaron cuatro variantes: C8-6 y Carol 3 con actividad de PRA y de ProFAR isomerasa y C8-2 y Gajos con actividad sólo de PRA isomerasa. Para analizar de manera cuantitativa la capacidad de complementación de cada una de estas variantes se hicieron cinéticas de crecimiento en medio mínimo líquido adicionando diferentes concentraciones de triptófano o histidina (según corresponda la cepa de selección). Inicialmente analizamos el control negativo (plásmido vacío) en presencia de las diferentes condiciones de medio de cultivo. Este análisis mostró que existe una correlación directa entre la concentración

de triptófano o de histidina adicionada al medio de cultivo y la densidad celular producida en dicha condición (Figura 9).

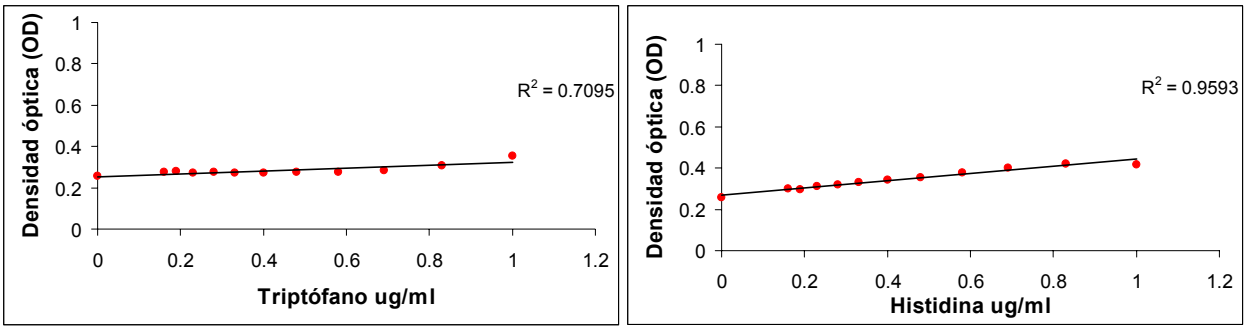

Figura 9. Correlación de la densidad celular con respecto a la concentración del aminoácido adicionado al medio del cultivo. Los puntos rojos representan el promedio del cuadruplicado de la máxima densidad celular del control negativo (plásmido vacío) a las diferentes concentraciones de los aminoácidos triptófano o histidina.

Encontramos que existe una correlación directa entre la densidad celular y la concentración del respectivo aminoácido en el medio de cultivo (Figura 9). Esta correlación nos sirvió para identificar que si la densidad celular de la variante en análisis sobrepasa a la mostrada por el control negativo (plásmido vacío) es señal de que existe una mayor producción del respectivo aminoácido y para poder producirlo dicha variante debe de estar complementando la actividad de PRA o de ProFAR isomerasa (según corresponda la selección). La Figura 10 muestra los resultados obtenidos para la actividad de PRA isomerasa y la Figura 11 para la actividad de ProFAR isomerasa. Debido a la cantidad de datos procedentes de las 1176 cinéticas de crecimiento decidimos mostrar de manera resumida solo las condiciones en las cuales las respectivas variantes presentan o no complementación de actividad.

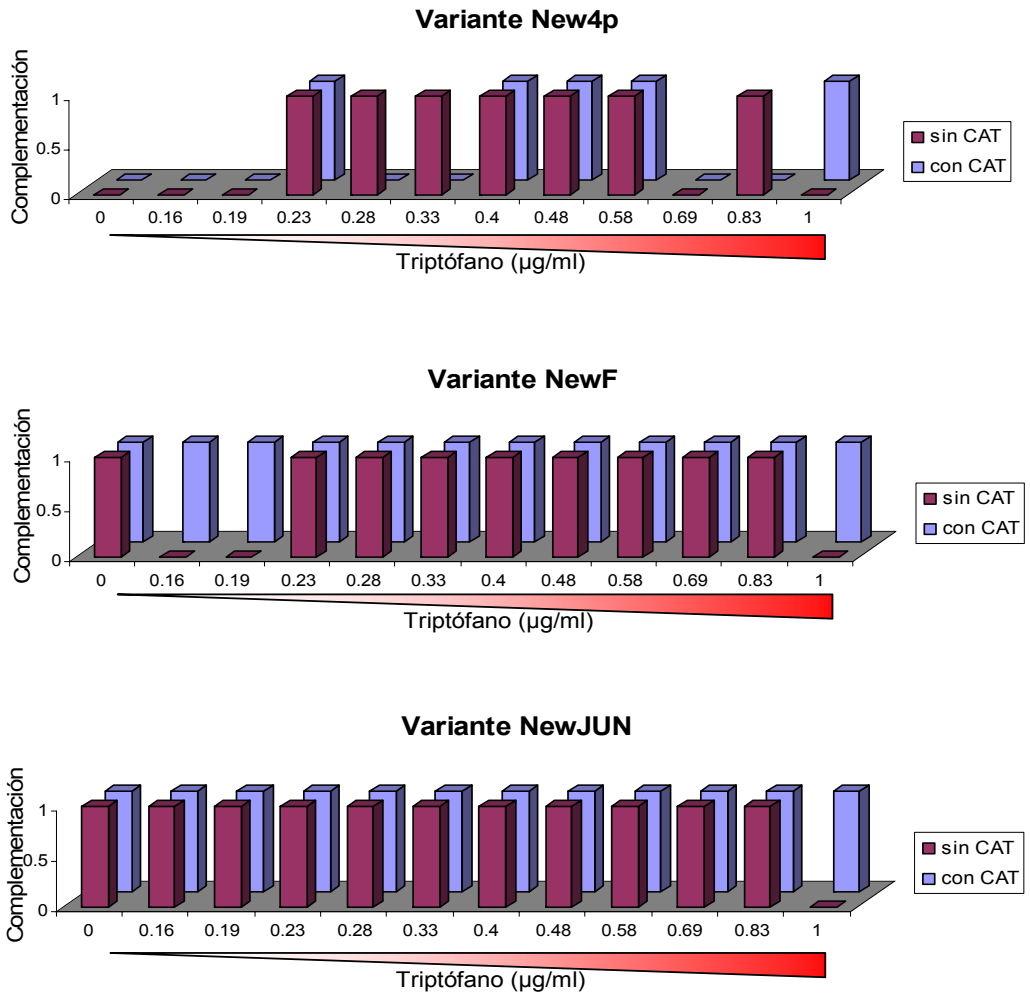

Figura 10 (continúa en la siguiente página). Análisis de complementación de la auxotrofía del gen de TrpF (actividad de PRA isomerasa) por las variantes ganadoras. El valor de 1 representa la condición estadísticamente significativa en la cual el promedio del valor máximo de OD de la variante (medido en unidades de densidad óptica a una Abs_{600nm}) fue mayor que el promedio del control negativo (plásmido vacío). Los datos fueron obtenidos a partir de curvas de crecimiento en medio mínimo y para cada variante se analizaron 4 réplicas por cada concentración de triptófano. De esta manera se analizaron 48 cinéticas de crecimiento por variante.

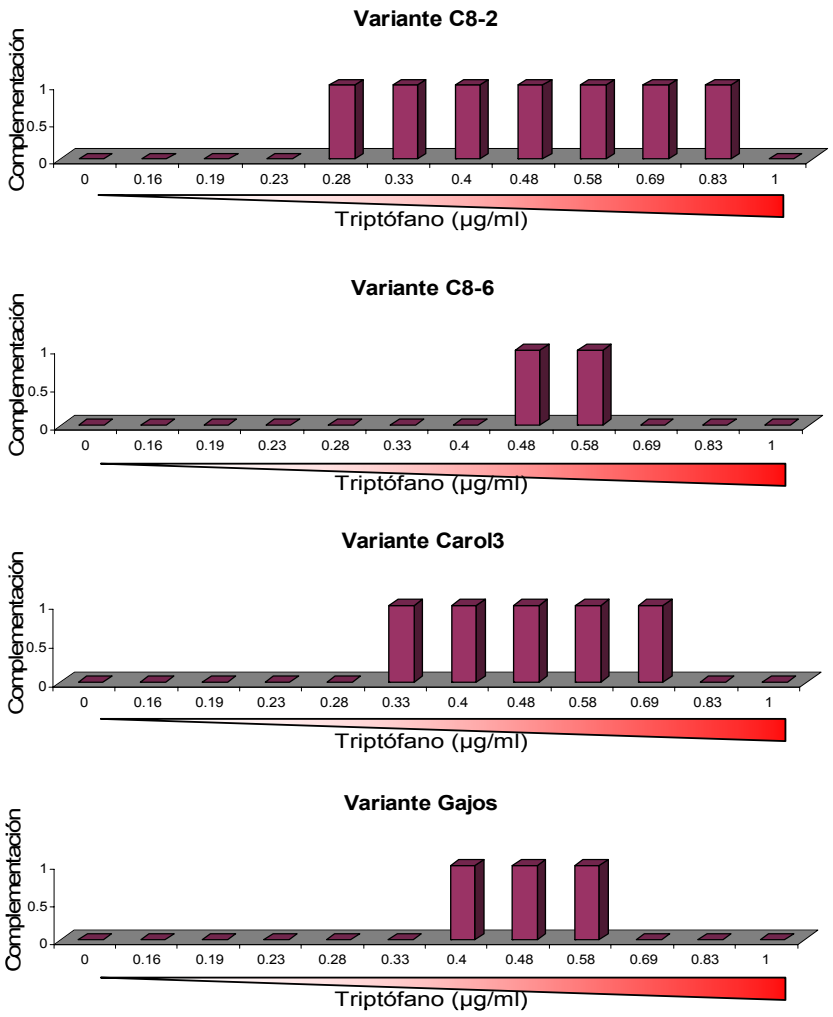

Figura 10. (CONTINUACION).

La variante New4p presentó complementación de la actividad de PRA isomerasa en el intervalo de 0.23 a 0.83 µg/ml de concentración de triptófano, mientras que las variantes NewF y NewJUN presentan crecimiento en casi todas las concentraciones de triptófano (Figura 10). Dado que teníamos el antecedente de que la complementación de la actividad de PRA isomerasa de TrpF se mejora cuando se encuentra fusionada con CAT, subclonamos las variantes New4p, NewF y NewJUN con este reportero de plegamiento y analizamos el papel de dicha fusión en la complementación de ambas funciones enzimáticas. Para el caso de la actividad de PRA isomerasa observamos que las variantes NewF y NewJUN mejoran la capacidad de complementación de esta actividad al estar fusionadas con CAT, mientras que para New4p el efecto es contrario

(Figura 10). Por otro lado las variantes C8-2 y Carol 3 complementan muy similar a New4p (en el intervalo de 0.23 a 0.83 µg/ml de concentración de triptófano para la primera y de 0.33 a 0.69 µg/ml para la segunda), mientras que las variantes C8-6 y Gajos sólo complementan en un intervalo muy limitado de concentración de triptófano: de 0.40 a 0.58 µg/ml (Figura 10).

La complementación de la actividad de ProFAR isomerasa fue menos exitosa en comparación con la de PRA isomerasa. Sólo las variantes Carol 3, C8-6, New4p, New4p-CAT y NewJUN-CAT complementaron la auxotrofía del gen de HisA y dicha complementación sólo fue observada en pocas condiciones del medio de cultivo (Figura 11). Estos resultados sugieren, por ejemplo, que la sobreexpresión obtenida por el aumento de la concentración de IPTG para la variante Carol 3 es un factor importante para obtener complementación de la actividad pero también es necesario adicionar pequeñas concentraciones de histidina (0.5 y 1 µg/ml). Contrario a este comportamiento fue el de la variante New4p, en donde encontramos que cuando se encuentra sin CAT la complementación depende de la sobreexpresión obtenida con 1mM de IPTG y también de la adición de trazas de histidina (0.69, 0.83 y 1 µg/ml) y cuando está en fusión con CAT también depende de la sobreexpresión con 1 mM de IPTG pero de una menor cantidad de histidina (0 y 0.5 µg/ml). Por otro lado las variantes C8-6 y NewJUN-CAT sólo complementaron en una condición de todas las probadas: para C8-6 en 1 mM de IPTG y 1 µg/ml de histidina, y para NewJUN-CAT en 1mM de IPTG y 0.69 µg/ml de histidina.

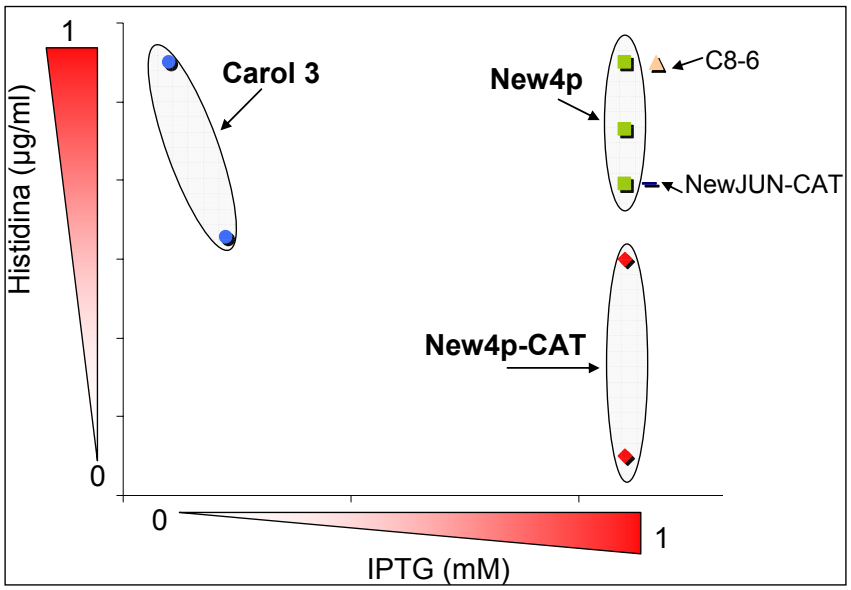

Figura 11. Resumen del análisis de complementación de la auxotrofía del gen de HisA (actividad de ProFAR isomerasa) por las variantes ganadoras. En la figura solo se presentan las variantes y las condiciones estadísticamente significativas en las cuales el promedio del valor máximo de OD (medido en unidades de densidad óptica a una

Abs_{600nm}) es mayor que el promedio del control negativo (plásmido vacío). Los datos fueron obtenidos a partir de curvas de crecimiento en medio mínimo y para cada variante se analizaron 4 réplicas por concentración de IPTG e histidina. De esta manera se analizaron 48 cinéticas de crecimiento por variante.

Para confirmar que los cultivos de las variantes que están complementando las respectivas cepas de selección son resultado del plásmido que tiene el gen de cada variante y no de otras mutaciones adaptativas en el cromosoma de la bacteria, se esparcieron colonias en cajas de LB y se extrajo de nuevo el plásmido. Este plásmido se utilizó para transformar células nuevas con la misma auxotrofía y estas células se estriaron en LB y en medio mínimo confirmando nuevamente que la complementación es producto de los plásmidos que contienen las diferentes variantes. Se secuenciaron estos plásmidos y encontramos que las secuencias son idénticas a las variantes originales. Además, para asegurarnos de que la complementación se debe a los genes que codifican para cada variante y no a algunas mutaciones adicionales adquiridas durante la construcción del plásmido, se aislaron los fragmentos de 700 pb correspondientes a los genes de cada variante y se subclonaron en vector nuevo y se confirmó nuevamente la complementación. Esto también se observó cuando fusionamos algunas variantes con CAT en se encontró que nuevamente rescatábamos la complementación.

La adición de triptófano o de histidina al medio mínimo se hizo para minimizar el estrés celular causado por la deficiencia de ese aminoácido y por la falta de la completa vía metabólica correspondiente. Los resultados obtenidos muestran que la complementación de las actividades por nuestras variantes es muy sensible a diferentes concentraciones de estos aminoácidos lo cual es indicativo de que las actividades enzimáticas que tenemos son vestigiales y requieren de múltiples efectos o conexiones dentro del citoplasma celular para poder que la actividad enzimática se vea reflejada en el crecimiento celular. Además también reflejan la baja sensibilidad del sistema cuando las actividades enzimáticas a seleccionar son vestigiales.

7. Conclusiones

Con el método para el intercambio sistemático de asas catalíticas SCLE (del inglés: Systematic Catalytic Loop Exchange) se encontró que es posible intercambiar sistemáticamente las asas β/α 2, 4 y 6 de la enzima TrpF-LoxP sin afectar su estabilidad, a pesar de que esta enzima es un barril $(\beta/\alpha)_8$ modificado (*16*). Además encontramos que las dos posiciones de las hebras β que preceden a cada asa intercambiada y que fueron mutadas a saturación son importantes para seleccionar proteínas plegadas pero sin función.

El análisis de la adaptabilidad de la relación estructura-función de las asas β/α en TrpF mostró que es posible medir la adaptabilidad funcional de estas asas mediante el valor de SFLA (*24*). Además encontramos que es posible trasladar el componente funcional implícito en las asas pero siempre y cuando se diseñen sitios de variabilidad en los residuos bisagra los cuales la unirán con el nuevo andamiaje, ayudando a su adaptabilidad funcional y estructural.

Los reemplazos de asas hechos en TrpF modificaron drásticamente el arreglo original del sitio activo y como consecuencia afectaron los parámetros catalíticos (*24*). Dada la posición estructural y la importancia del asa 6 en la unión del sustrato y la actividad catalítica de esta enzima, todas las nuevas variantes podrían ser seleccionadas para diferentes especificidades de sustrato así como para otras actividades catalíticas. Además nuestra estrategia demostró su viabilidad imitando al sistema de generación de variabilidad natural en el asa H3 de los anticuerpos (*24*)**.**

El conocer parte de la historia natural de HisA, PriA y TrpF (*4, 20*) nos permitió abordar el rediseño de los residuos catalíticos de TrpF desde diversas estrategias, con las cuales se obtuvo información importante sobre la función de esta enzima, la cual no hubiera sido posible adquirir de otra manera. Basado en los resultados experimentales y en los modelos construidos para las variantes CDA y CDT se sugirió un mecanismo catalítico, en donde la Cys7 mantiene su rol como base general, mientras que el Asp107 puede ser el nuevo ácido general. La Thr126 de la variante CDT parece estar cumpliendo un rol importante al mejorar la actividad catalítica de CDT, tal y como ha sido observado en PriA para el residuo equivalente, la Thr166 (*20*). El mecanismo de reacción propuesto parece funcionar como una combinación de los residuos catalíticos

de TrpF y PriA, en donde de TrpF estaría conservando la identidad y la posición de la base general (Cys7) y de PriA estaría imitando la ubicación e identidad del ácido general (Asp107) y la ubicación y el rol funcional de la Thr166 (ahora Thr126 en TrpF).

La introducción de los sitios de variabilidad NNS en el intercambio de las asas 1, 5 y 6 de TrpF por las correspondientes de PriA fue importante para obtener proteínas con función de PRA y/o de ProFAR isomerasa que demuestran actividad *in vivo*. Prueba de ello es que a partir de dos librerías de un millón de variantes cada una sólo se obtuvieron siete variantes con alguna función, cuyas diferencias radican básicamente en los residuos seleccionados en las posiciones mutagenizadas con NNS. Es importante señalar que la exploración de ambas actividades *in vivo* pudo llevarse a cabo en los dos moldes de los sitios activos de TrpF, corroborando con ello la adaptabilidad del rediseño del sitio activo. Aunque cabe señalar que de la librería con el sitio activo al estilo de TrpF fue de donde se obtuvieron más variantes y sin mutaciones adicionales en las asas intercambiadas, mientras que de la librería construida con el sitio activo al estilo de PriA dos de de las tres variantes ganadoras presentaron mutaciones adicionales en las asas intercambiadas. Esto también refleja la fuerte presión de selección sobre la secuencia de estas asas así como la dependencia de su secuencia y adaptación para lograr la exploración de nuevas funciones enzimáticas.

Uno de los objetivos finales del proyecto de doctorado fue explorar *in vivo* las nuevas funciones enzimáticas adquiridas mediante la utilización de la estrategia de intercambio asas y variabilidad en las bisagras. La evolución y caracterización *in vitro* de estas variantes es parte de otro proyecto del laboratorio, pero podemos concluir que es posible obtener (*in vivo*) la nueva actividad de ProFAR isomerasa en algunas de las variantes de TrpF con las asas 1, 5 y 6 de PriA. Incluso algunas como New4p, NewJUN-CAT, Carol 3 y C8-6 mantuvieron además la actividad original de PRA isomerasa, lo que las hace ser bifuncionales. Dado que las actividades enzimáticas que estamos explorando parecen ser muy bajas, fue necesario aplicar un análisis masivo de complementación para encontrar el intervalo de condiciones en las cuales fuese posible asignar una complementación real de la actividad enzimática *in vivo*.

Las asas β/α de TrpF pueden ser intercambiadas sin afectar la estabilidad de la proteína y sin perder la función original. Además con el diseño de bisagras se puede

adaptar en otras enzimas la relación estructura-función codificada en ellas. Todas las estrategias de intercambio de asas, introducción de variabilidad en sus bisagras y del rediseño del sitio activo, presentadas en su conjunto en la presente tesis son un camino prometedor para diseñar nuevas funciones enzimáticas en proteínas con el plegamiento de barril $(\beta/\alpha)_8$.

8. Perspectivas

A continuación enunciamos algunos de los experimentos que consideramos más inmediatos para el futuro avance del proyecto presentado en esta tesis de investigación. El orden de los mismos es acorde a los capítulos presentados.

- Sobre-expresar y purificar las variantes CDA y CDT así como las mutantes puntuales de alanina. Con esto se puede llevar a cabo el análisis de la actividad enzimática de PRA isomerasa *in vitro*, lo cual ayudaría a describir con detalle el mecanismo de reacción del sitio activo rediseñado de TrpF y el rol funcional de cada uno de los residuos que participan en la catálisis.
- Implementar diferentes estrategias de evolución dirigida en las variantes CDA y CDT para lograr que el sitio activo rediseñado funcione con los mismos parámetros cinéticos que el original y analizar donde ocurrieron las mutaciones que lo mejoraron.
- Resolver la estructura tridimensional de las variantes que tienen rediseñado el sitio activo de TrpF y con ello explicar con detalle el posible mecanismo de reacción.
- Mejorar la actividad enzimática de las variantes ganadoras con las tres asas de PriA intercambiadas en TrpF mediante la identificación de mutaciones estabilizadoras que les permitan mantener la función y sobre ellas probar nuevos ciclos de evolución dirigida.
- Probar diferentes técnicas de evolución dirigida para mejorar la actividad y estabilidad de las variantes ganadoras e incluso hacerlo con la fusión a diversos reporteros de plegamiento, lo que puede ayudar a encontrar mutantes estabilizadas por las fusiones.
- Probar la selección de actividades de PRA y de ProFAR isomerasa en cepas que tengan la co-expresión de chaperonas como GroEL y GroES.
- Dado que en algunas variantes mantenemos la actividad original de PRA isomerasa, se podría mejorar la actividad de ProFAR isomerasa a través de selección por “neutral drift”. Purificar las variantes ganadoras y hacer los respectivos ensayos de actividad enzimática *in vitro.*

9. Bibliografía

1. Vasserot AP, et al. (2003) Optimization of protein therapeutics by directed evolution. *Drug Discov Today 8*: 118-126.
2. Rubin-Pitel SB, Zhao H (2006) Recent advances in biocatalysis by directed enzyme evolution. *Comb Chem High Throughput Screen 9*: 247-257.
3. Kazlauskas R, Lutz S (2009) Engineering enzymes by 'intelligent' design. *Curr Opin Chem Biol 13*: 1-2.
4. Barona-Gómez FO-L, A. Soberón, X. (2008) in *Advances in Protein Physical Chemistry.* (García-Hernández, E. F.-V., A., Ed.) pp 407-438, Transworld Research Network, Kerala.
5. Jaeger KE, Eggert T, Eipper A, Reetz MT (2001) Directed evolution and the creation of enantioselective biocatalysts. *Appl Microbiol Biotechnol 55*: 519-530.
6. Yuan L, Kurek I, English J, Keenan R (2005) Laboratory-directed protein evolution. *Microbiol Mol Biol Rev 69*: 373-392.
7. Orencia MC, et al. (2001) Predicting the emergence of antibiotic resistance by directed evolution and structural analysis. *Nat Struct Biol 8*: 238-242.
8. Cohen N, Abramov S, Dror Y, Freeman A (2001) In vitro enzyme evolution: the screening challenge of isolating the one in a million. *Trends Biotechnol 19*: 507-510.
9. Rothlisberger D, et al. (2008) Kemp elimination catalysts by computational enzyme design. *Nature 453*: 190-195.
10. Siegel JB, et al. (2010) Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science 329*: 309-313.
11. Bornscheuer U (2009) Meeting report: Protein Design and Evolution for Biocatalysis. *Biotechnology Journal 4*: 443-445.
12. Venkitakrishnan RP, et al. (2004) Conformational changes in the active site loops of dihydrofolate reductase during the catalytic cycle. *Biochemistry 43*: 16046-16055.
13. Heinis C, et al. (2006) Evolving the substrate specificity of O6-alkylguanine-DNA alkyltransferase through loop insertion for applications in molecular imaging. *ACS Chem Biol 1*: 575-584.
14. Wang Y, Berlow RB, Loria JP (2009) Role of loop-loop interactions in coordinating motions and enzymatic function in triosephosphate isomerase. *Biochemistry 48*: 4548-4556.
15. Sterner R, Hocker B (2005) Catalytic versatility, stability, and evolution of the (betaalpha)8-barrel enzyme fold. *Chem Rev 105*: 4038-4055.
16. Ochoa-Leyva A, et al. (2009) Protein design through systematic catalytic loop exchange in the (beta/alpha)8 fold. *J Mol Biol 387*: 949-964.
17. Eberhard M, et al. (1995) Indoleglycerol phosphate synthase-phosphoribosyl anthranilate isomerase: comparison of the bifunctional enzyme from Escherichia coli with engineered monofunctional domains. *Biochemistry 34*: 5419-5428.
18. Henn-Sax M, et al. (2002) Two (betaalpha)(8)-barrel enzymes of histidine and tryptophan biosynthesis have similar reaction mechanisms and common strategies for protecting their labile substrates. *Biochemistry 41*: 12032-12042.

19. Barona-Gomez F, Hodgson DA (2003) Occurrence of a putative ancient-like isomerase involved in histidine and tryptophan biosynthesis. *EMBO Rep 4*: 296-300.
20. Wright H, et al. (2008) The structure/function relationship of a dual-substrate (betaalpha)8-isomerase. *Biochem Biophys Res Commun 365*: 16-21.
21. Noda-Garcia L, et al. (2010) Identification and analysis of residues contained on beta --> alpha loops of the dual-substrate (beta alpha)8 phosphoribosyl isomerase A specific for its phosphoribosyl anthranilate isomerase activity. *Protein Sci 19*: 535-543.
22. Glasner ME, Gerlt JA, Babbitt PC (2006) Evolution of enzyme superfamilies. *Curr Opin Chem Biol 10*: 492-497.
23. O'Brien PJ, Herschlag D (1999) Catalytic promiscuity and the evolution of new enzymatic activities. *Chem Biol 6*: R91-R105.
24. Ochoa-Leyva A, et al. (2011) Exploring the Structure-Function Loop Adaptability of a (beta/alpha)8-barrel enzyme through loop replacement and hinge variability. In press*. Journal of Molecular Biology*.
25. Tawfik DS (2006) Biochemistry. Loop grafting and the origins of enzyme species. *Science 311*: 475-476.
26. Park HS, et al. (2006) Design and evolution of new catalytic activity with an existing protein scaffold. *Science 311*: 535-538.
27. Ochoa-Leyva A, et al. (2010) Understanding structure-function adaptability of loops using variable hinges in a (beta/alpha)8-barrel enzyme: implications for the development of novel enzymes. *Submitted to Journal of Molecular Biology*.
28. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem 72*: 248-254.
29. Hommel U, Eberhard M, Kirschner K (1995) Phosphoribosyl anthranilate isomerase catalyzes a reversible amadori reaction. *Biochemistry 34*: 5429-5439.
30. Ochoa-Leyva A. (2008) in *Instituto de Biotecnología* pp 82, Universidad Nacional Autónoma de México, Cuernavaca.
31. Saab-Rincon G, et al. (2005) Generation of variability by in vivo recombination of halves of a (beta/alpha)8 barrel protein. *Biomol Eng 22*: 113-120.
32. De Genst E, et al. (2005) Strong in vivo maturation compensates for structurally restricted H3 loops in antibody repertoires. *J Biol Chem 280*: 14114-14121.
33. Weill JC, Reynaud CA (1996) Rearrangement/hypermutation/gene conversion: when, where and why? *Immunol Today 17*: 92-97.
34. Tonegawa S (1983) Somatic generation of antibody diversity. *Nature 302*: 575-581.
35. Gerstein M, Lesk AM, Chothia C (1994) Structural mechanisms for domain movements in proteins. *Biochemistry 33*: 6739-6749.
36. Liu Y, Kuhlman B (2006) RosettaDesign server for protein design. *Nucleic Acids Res 34*: W235-238.
37. Wilmanns M, et al. (1991) Structural conservation in parallel beta/alpha-barrel enzymes that catalyze three sequential reactions in the pathway of tryptophan biosynthesis. *Biochemistry 30*: 9161-9169.

38. Jurgens C, et al. (2000) Directed evolution of a (beta alpha)8-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc Natl Acad Sci U S A 97*: 9925-9930.
39. Claren J, Malisi C, Hocker B, Sterner R (2009) Establishing wild-type levels of catalytic activity on natural and artificial (beta alpha)8-barrel protein scaffolds. *Proc Natl Acad Sci U S A 106*: 3704-3709.
40. Babbitt PC, Gerlt JA (1997) Understanding enzyme superfamilies. Chemistry As the fundamental determinant in the evolution of new catalytic activities. *J Biol Chem 272*: 30591-30594.
41. Gerlt JA, Babbitt PC (2001) Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu Rev Biochem 70*: 209-246.
42. Panchenko AR, Madej T (2005) Structural similarity of loops in protein families: toward the understanding of protein evolution. *BMC Evol Biol 5*: 10.
43. Peng T, Zintsmaster JS, Namanja AT, Peng JW (2007) Sequence-specific dynamics modulate recognition specificity in WW domains. *Nat Struct Mol Biol 14*: 325-331.
44. Pettersen EF, et al. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem 25*: 1605-1612.
45. Suhre K, Sanejouand YH (2004) ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res 32*: W610-614.
46. Sarkar, G., and Sommer, S.S. (1990). The "megaprimer" method of site-directed mutagenesis. Biotechniques *8*, 404-407.

10. Anexos

Otras publicaciones generadas en este trabajo:

1. Wright, H., Noda-Garcia, L., Ochoa-Leyva, A., Hodgson, D.A., Fulop, V., and Barona-Gomez, F. (2008). The structure/function relationship of a dual-substrate (betaalpha)8-isomerase. Biochem Biophys Res Commun *365*, 16-21.

2. Barona-Gómez, F.O.-L., A. Soberón, X. (2008). Advances and perspectives in protein engineering: From natural history to directed evolution of enzymes. In Advances in Protein Physical Chemistry., E.F.-V. García-Hernández, A., ed. (Kerala: Transworld Research Network), pp. 407-438.