



Universidad Nacional Autónoma de México
Maestría y doctorado en Ciencias bioquímicas

Diseño y caracterización de un dominio Rossmann consenso

TESIS

Que para optar por el grado de:

Maestra en Ciencias

PRESENTA:

Claudia Iris Bravo Bonilla

TUTOR PRINCIPAL:

Dr. Lorenzo Patrick Segovia Forcella

Instituto de Biotecnología, UNAM

MIEMBROS DEL COMITÉ TUTOR:

Dr. Ernesto Ortiz Suri

Instituto de Biotecnología, UNAM

Dr. Christian Sohlenkamp

Maestría y doctorado en Ciencias bioquímicas

Cuernavaca, Morelos | Abril 2015



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

El presente trabajo se realizó bajo la dirección del Dr. Lorenzo Segovia en el Instituto de Biotecnología de la Universidad Nacional Autónoma de México. Durante la realización del mismo se recibió una beca del Consejo Nacional de Ciencia y Tecnología. Asimismo fue financiado por los proyectos PAPIIT IN213511 y CONACYT 132580.

Para mi mamá: Gracias por todo.

AGRADECIMIENTOS

Al Dr. Lorenzo Segovia por la oportunidad, el apoyo, la paciencia y las palabras exactas.

A las siguientes personas por su asesoría técnica y valiosos comentarios: el Dr. Ernesto Ortiz, el Dr. Christian Sohlenkamp, la Dra. Gloria Saab, la Dra. Claudia Martínez, la Dra. Marcela Ayala, el Dr. Miguel Costas, la Lic. Diana Martínez.

A los miembros del laboratorio 12 por su amistad, conocimientos, comidas y conversaciones: Mario, Gustavo, Miguel, Fernando, Banda, Dago, Joaquín, Mayra, Estefa.

A mis amigos de la maestría: Liz, Tona, Caro y Diana. Gracias por su apoyo, risas, consejos y demás.

A mis amigos que me ayudaron a sobrevivir: Matilde, Gemma, Pável.

A mi hermosa familia.

A todos aquellos que no menciono pero que siempre estuvieron.

Gracias. Gracias. Gracias.

Título original: *Diseño y caracterización de un dominio Rossmann consenso*

Escrito por: Claudia Iris Bravo Bonilla

Diseño editorial por: ©Thésika · Diseño de tesis

contacto@thesika.com.mx | www.thesika.mx

Impreso en México DF durante 2015

Composición y diseño: J. Martín Rejón (*Thésika*)

Diseño de cubierta: J. Martín Rejón (*Thésika*)

Encuadernación: Sandy Pamela de Jesús Coronado R. & J. Martín Rejón(*Thésika*)

Corrección ortográfica: Sandy Pamela de Jesús Coronado R.

Tesis



*Diseño y Caracterización
de un Dominio
Rossmann Consenso*



CLAUDIA
IRIS
BRAVO
BONILLA



RESUMEN

Hemos diseñado tres dominios Rossmann consenso a partir de diferentes alineamientos de enzimas *shikimato deshidrogenasas*, la cual cataliza el cuarto paso de la vía del *shikimato*. En cada uno de los consensos se conservaron los aminoácidos que co-evolucionan, y se usó como secuencia guía la secuencia de AroE de *Escherichia coli*. Solamente uno de los consensos se expresó: Con1, el cual se generó a partir de un conjunto de 47 secuencias muy similares a AroE, del que se determinaron 15 aminoácidos consenso. La proteína consenso Con1 se presentó en la fracción soluble con un rendimiento superior al de la proteína nativa AroE. Los experimentos de dicroísmo circular muestran que el contenido de estructura secundaria de Con1 es similar al de AroE. Los datos de desnaturalización química con urea de Con1 se ajustaron a un modelo de transición de dos estados y los resultados obtenidos sugieren que el desplegamiento de Con1 es cooperativo. El perfil de calorimetría de Con1 mostró una única transición térmica correspondiente al comportamiento de un monómero. El valor de T_m^{app} de Con1 fue de 53.2 °C, solamente 2.3 °C mayor que la determinada para AroE en el laboratorio. Con1 mostró actividad *in vivo* similar a la nativa, y de forma *in vitro* es capaz de usar NADP⁺ y NAD⁺ como cofactor.

ÍNDICE DE CONTENIDO

Resumen | 09

Índice de figuras | 11

Índice de tablas | 14

Lista de abreviaturas | 14

Introducción | 17

Dominio Rossmann | 19

Modelo de estudio | 21



Antecedentes | 25

Antecedentes del laboratorio | 26

Antecedentes de ingeniería de consensos | 26

Hipótesis | 31

Objetivos | 33

Estrategia experimental | 35

Determinación de la secuencia consenso | 36

Diseño y síntesis del gen con la secuencia consenso | 36

Expresión | 40

Purificación | 41

Caracterización | 41

Ensayos de actividad | 43



Resultados y Discusión | 45

Determinación de la secuencia consenso | 46

Diseño y síntesis del gen con la secuencia consenso | 48

Expresión | 54

Purificación | 57

Caracterización | 59

Ensayos de actividad | 68

Conclusiones | 73

Perspectivas | 77

Bibliografía | 79

Anexos | 85

Anexo 1 | 86

Anexo 2 | 87

Anexo 3 | 87

Anexo 4 | 88

ÍNDICE DE FIGURAS

Figura 1. <i>Dominio Rossmann</i> . Las flechas designan hojas β y los rectángulos hélices α , los círculos representan los residuos de glicina conservados. Imagen tomada de (BOTTONS <i>et al.</i> , 2002)	20
Figura 2. <i>Reacción catalizada por shikimato deshidrogenasa</i> (EC 1.1.1.25). La enzima <i>shikimato</i> deshidrogenasa, codificada por el gen <i>AroE</i> en <i>E. coli</i> , cataliza la reducción reversible del 3-dehidroshikimato a <i>shikimato</i> en la presencia de NADPH como donador de electrones. Imagen tomada de KEGG: http://www.genome.jp/dbget-bin/www_bget?rn:R02413	22
Figura 3. <i>Representación tridimensional de AroE</i> . En amarillo se muestran las hebras β , en verde las hélices α y en azul las asas. Imagen tomada de (MICHEL <i>et al.</i> , 2003)	23
Figura 4. <i>Vector PMK-RQ de Life Technologies</i> . Plásmido multicopia con un sitio de origen Col E1 y un gen de resistencia a kanamicina (<i>kanR</i>), donde <i>Con</i> puede ser cualquiera de los tres genes consenso	37
Figura 5. <i>Vector de expresión pT4</i> . Plásmido multicopía con un gen de resistencia a kanamicina (<i>kanR</i>) y un promotor <i>trc</i> fuerte, donde <i>EBBE</i> es una enzima quimérica.	37
Figura 6. <i>pT4 consenso</i> . Plásmido multicopia que presenta un gen de resistencia a kanamicina (<i>kan(R)</i>), un promotor <i>trc</i> fuerte. Donde <i>Con</i> puede cualquiera de las tres secuencias consenso.	39
Figura 7. <i>Dendograma correspondiente al dominio Rossmann de las 1637 secuencias</i> . En azul se muestran las secuencias que componen el primer conjunto (Consenso1) junto con las rojas componen el conjunto dos (Consenso2), mientras que en verde se muestran las del tercer conjunto (Consenso 3). Imagen realizada con el programa MEGA 6.0 (TAMURA, STECHER, PETERSON, FILIPSKI, & KUMAR, 2013)	47
Figura 8. <i>Representación tridimensional de la estructura secundaria del dominio Rossmann de Con1</i> . En azul se muestran la posición de los aminoácidos mutados, en rojo claro se muestran los aminoácidos que se encuentran en contacto con el cofactor (estructura de palitos roja). Imagen realizada con Pymol 1.5.	49
Figura 9. <i>Representación tridimensional de la estructura secundaria del dominio Rossmann de Con2</i> . En verde se muestran la posición de los aminoácidos mutados, en rojo claro se muestran los aminoácidos que se encuentran en contacto con el cofactor (estructura de palitos roja). Imagen realizada con Pymol 1.5	50
Figura 10. <i>Representación tridimensional de la estructura secundaria del dominio Rossmann de Con3</i> . En naranja se muestran la posición de los aminoácidos mutados, en rojo claro se muestran los aminoácidos que se encuentran en contacto con el cofactor (estructura de palitos roja). Imagen realizada con Pymol 1.5	51
Figura 11. <i>Alineamiento de las secuencias de AroE y los diferentes consensos</i> . La estructura secundaria se muestra en la parte superior. En los cuadros rojos se muestran los aminoácidos que son los mismos en todas las secuencias. En letras rojas	52

se muestran los aminoácidos que son similares en todas las secuencias. En letras negras se muestran los aminoácidos que no presentan similitud en las secuencias. La x representa todos los aminoácidos que no fueron determinados en el consenso. El modelo fue generado con ESPript (ROBERT & GOUET, 2014)

- Figura 12. *Ligación*. Carril 1: marcador de peso molecular. Carril 2: vector ϵ BE.His digerido y purificado. Carril 3: inserto Con1 purificado. Carril 4: inserto Con2 purificado. Carril 5: inserto Con3 purificado. | 53
- Figura 13. Gel SDS-PAGE de la fracción soluble de diferentes cultivos a 37 °C. Carril 1: marcador de peso molecular. Carril 2: ρ T4 vacío como control negativo. Carril 3: AroE como control positivo. Carril 4: Con1. Carril 5: Con2. Carril 6: Con 3. En rojo se muestra la altura a la cual se espera observar la expresión de las proteínas consenso. | 55
- Figura 14. Gel SDS-PAGE de cultivos a diferentes temperaturas. Carril 1: marcador de peso molecular. Carril 2: ρ T4 vacío a 37 °C. Carril 3: AroE a 37 °C. Carril 4: Con1 a 37 °C. Carril 5: Con2 a 37 °C. Carril 6: Con3 a 37 °C. Carril 7: Con2 a 30 °C. Carril 8: Con3 a 30 °C. Carril 9: Con2 a 25 °C. Carril 10: Con3 a 25 °C. En rojo se muestra la altura a la cual se espera observar la expresión de las proteínas consenso. | 55
- Figura 15. *Western Blot del gel mostrado en la figura anterior*. Carril 1: marcador de peso molecular. Carril 2: ρ T4 vacío a 37 °C. Carril 3: AroE a 37 °C. Carril 4: Con1 a 37 °C. Carril 5: Con2 a 37 °C. Carril 6: Con3 a 37 °C. Carril 7: Con2 a 30 °C. Carril 8: Con3 a 30 °C. Carril 9: Con2 a 25 °C. Carril 10: Con3 a 25 °C. En rojo se muestra la altura a la cual se espera observar la expresión de las proteínas consenso. | 56
- Figura 16. *Gel SDS-PAGE de la purificación de Con1*. Carril 1: marcador de peso molecular. Carril 2: fracción insoluble. Carril 3: fracción soluble después de entrar a la columna. Carril 4: fracción correspondiente a 20 mM de imidazol. Carril 5: fracción correspondiente a 35 mM de imidazol. Carril 6: fracción correspondiente a 75 mM de imidazol. Carril 7: fracción correspondiente a 100 mM de imidazol. Carril 8: fracción correspondiente a 150 mM de imidazol. Carril 9: fracción correspondiente a 200 mM de imidazol. Carril 10: fracción correspondiente a 250 mM de imidazol. En rojo se muestra la altura a la cual se espera observar la expresión de las proteínas consenso. | 57
- Figura 17. *Purificación de Con1*. Carril 1: marcador de peso molecular. Carril 2: fracción soluble. Carril 3: fracción correspondiente a 100 mM de imidazol. Carril 4: muestra concentrada usando el Vivaspin. Carril 5: primera diálisis. Carril 6: muestra purificada después de la segunda diálisis y filtración. En rojo se muestra la altura a la cual se espera observar la expresión de las proteínas consenso. | 58
- Figura 18. *Purificación de AroE*. Carril 1: marcador de peso molecular. Carril 2: fracción insoluble. Carril 3: fracción soluble después de entrar a la columna. Carril 4: fracción correspondiente a 20 mM de imidazol. Carril 5: fracción correspondiente a 35 mM de imidazol. Carril 6: fracción correspondiente a 75 mM de imidazol. | 59

Carril 7: fracción correspondiente a 100 mM de imidazol. Carril 8: fracción correspondiente a 150 mM de imidazol. Carril 9: fracción correspondiente a 200 mM de imidazol. Carril 10: fracción correspondiente a 250 mM de imidazol. En rojo se muestra la altura a la cual se espera observar la expresión de las proteínas consenso.

- Figura 19. Purificación de AroE. Carril 1: marcador de peso molecular. Carril 2: fracción soluble. Carril 3: fracción correspondiente a 100 mM de imidazol. Carril 4: muestra concentrada usando el Vivaspin. Carril 5: muestra purificada después de la segunda diálisis y filtración. En rojo se muestra la altura a la cual se espera observar la expresión de las proteínas consenso. | 59
- Figura 20. *Estructura secundaria*. Con1 y AroE muestran un alto contenido de hélices α . Las elipticidades molares residuales (θ) de ambas fueron medidas desde 190-260 nm. Los porcentajes de estructura secundaria de Con1 fueron determinados usando el algoritmo de deconvolución *CDsstr* | 60
- Figura 21. *Desnaturalización química de Con1*. (A) Valores obtenidos durante el experimento de desnaturalización química de Con1, en gris se observa el modelo obtenido con Savuka. (B) Fracción aparente de la desnaturalización química de Con1, donde 0 indica el estado nativo y 1 es el estado desplegado; en gris se observa el modelo obtenido con Savuka. (C) Valores residuales obtenidos al realizar el ajuste de los datos de Con1 con el modelo de dos estados. Resultados similares fueron obtenidos de tres experimentos independientes. | 63
- Figura 22. *Desnaturalización térmica de AroE y Con1*. El proceso de renaturalización deber ser superponible a los datos obtenidos durante la desnaturalización para poder afirmar que el proceso de desnaturalización térmica es reversible. Resultados similares fueron obtenidos de tres experimentos independientes. | 64
- Figura 23. *Modelo de la desnaturalización térmica de AroE y Con1*. En la parte superior se muestra la fracción aparente de AroE (círculos abiertos) y Con1 (triángulos abiertos), en gris se muestra el ajuste al modelo de dos estados obtenido con Savuka. En la parte inferior se muestran los valores residuales obtenidos al realizar el ajuste de los datos de AroE y Con1 con el modelo de dos estados. | 65
- Figura 24. *Perfiles de calorimetría de AroE (wild type) y Con1 (consenso)*. El valor de T_m^{app} de AroE fue de 50.9 °C (324.1 K), mientras que el de Consenso 1 fue de 53.2 °C (326.4 K). Resultados similares fueron obtenidos de tres experimentos diferentes. | 66
- Figura 25. *Espectrometría de masas de Con1*. El gráfico de la izquierda corresponde a una shikimato deshidrogenasa modificada (Con1) | 67
- Figura 26. *Complementación en medio rico (LB con antibiótico) y medio mínimo m9*. En medio rico (gráfico izquierdo) todas las muestras presentan crecimiento desde las primeras horas. En medio mínimo m9 (gráfico derecho), AroE (control positivo) y Con1 muestran crecimiento después de las veinte horas. Con1 es capaz de emular la actividad *in vivo* de AroE. Resultados similares fueron obtenidos de tres experimentos independientes. | 68

Figura 27. *Michaelis-Menten* de *Con1*. (A) *Shikimate* como el sustrato que varía mientras NADP^+ se mantuvo constante. (B) NADP^+ como cofactor que varía mientras que *shikimate* se mantuvo constante. (C) NAD^+ como cofactor que varía mientras que *shikimate* se mantuvo constante.

ÍNDICE DE TABLAS

Tabla 1. *Comparación de parámetros cinéticos de diferentes enzimas shikimate deshidrogenasas.* Los parámetros cinéticos fueron determinados a pH 9.0 y 25°C al variar la concentración de sustrato o cofactor mientras el otro se mantenía a concentraciones fijas. Los parámetros cinéticos de AroE y YdiB fueron determinados por (MICHEL *et al.*, 2003).

LISTA DE ABREVIATURAS

ADN: ácido desoxirribonucleico.

Ala: alanina.

Arg: arginina.

AroD: enzima deshidratasa 3-dehidroquinato.

AroE: enzima shikimate deshidrogenasa.

Asn: asparagina.

Asp: ácido aspártico.

B. subtilis: *Bacillus subtilis*.

BLA: enzima β -lactamasa.

BLAST: *Basic Local Alignment Search Tool*.

CCM: mutagénesis consenso combinatoria.

Con1: enzima shikimate deshidrogenasa con el dominio Rossmann consenso 1.

Con2: enzima shikimate deshidrogenasa con el dominio Rossmann consenso 2.

Con3: enzima shikimate deshidrogenasa con el dominio Rossmann consenso 3.

Cys: cisteína.

DC: dicroísmo circular.

DSC: calorimetría diferencial de barrido.

E. coli: *Escherichia coli*.

EBE.His: proteína quimérica que presenta el dominio N-terminal de AroE de *E. coli*, dominio C-terminal de AroE de *B. subtilis* y las dos últimas alfa hélices del dominio C-terminal de la AroE de *E. coli*, y una etiqueta de seis histidinas.

EDTA: ácido etildiaminotetraacético.

EMBOSS: *The European Molecular Biology Open Software Suite*.

ExPASy: *Expert Protein Analysis System*.

FAD: flavín adenín dinucleótido.

Fapp: fracción aparente.

Gln: glutamina.
Glu: ácido glutámico.
Gly: glicina.
His: histidina.
IBT: Instituto de Biotecnología.
Ile: isoleucina.
K_{cat}: número de recambio.
K_M: constante de Michaelis.
LB: medio Luria Bertani.
Leu: leucina.
Lys: lisina.
mAG: proteína monomérica verde Azami.
Met: metionina.
MRE: *Mean Residue Ellipticity*.
NAD: nicotinamida adenina dinucleótido.
NADP: nicotinamida adenina dinucleótido fosfato.
Ni: níquel.
NRMSD: desviación del valor cuadrático medio.
NTA: ácido nitrilotriácetico.
PBS: buffer fosfato salino.
PCR: reacción en cadena de la polimerasa.
PDB: *Protein Data Bank*.
Phe: fenilalanina.
PMSF: fluoruro de fenilmetilsulfonilo.
Pro: prolina.
PVDF: fluoruro de polivinilideno.
SA: ácido shikímico.
SCA: análisis de acoplamiento estadístico.
SDS-PAGE: electroforesis en gel de poliacrilamida con dodecilsulfato sódico.
Ser: serina.
SVD: descomposición en valores singulares.
TBST: buffer Tris salino.
Thr: treonina.
TIM: triosafosfato isomerasa.
Tm: temperatura de desnaturalización.
Tm^{app}: temperatura aparente de desnaturalización.
tRNA: ácido ribonucleico de transferencia.
Trp: triptófano.
Tyr: tirosina.
UNAM: Universidad Nacional Autónoma de México.
UV: ultravioleta.
Val: valina.
V_{max}: velocidad máxima.
YdiB: enzima shikimato/quinato deshidrogenasa

SECCIÓN
01



INTRODUCCIÓN



*Diseño y Caracterización
de un Dominio
Rossmann Consenso*



CLAUDIA IRIS BRAVO BONILLA

INTRODUCCIÓN

Las enzimas son excelentes catalizadores debido a que llevan a cabo diferentes reacciones bioquímicas, son altamente selectivas y específicas, y son una alternativa amable con el medio ambiente. Estos atributos han llevado al uso de enzimas en diversas industrias para la síntesis de moléculas complejas y de alto valor (SCHMID *et al.*, 2001).

La mayoría de las enzimas no son aptas para su uso a nivel industrial debido a que las condiciones de los procesos catalíticos pueden causar la inactivación o desnaturalización de la enzima (DALBY, 2007). La estabilización de proteínas es importante para mejorar su vida media y funcionalidad bajo las condiciones adversas presentes en las aplicaciones biotecnológicas.

Se han reportado muchos casos exitosos en los cuales se lograron mejorar ciertas características de las enzimas: incremento de actividad en nuevos sustratos (GLIEDER, FARINAS, & ARNOLD, 2002), modificación de enantio-selectividad (CARR *et al.*, 2003), aumento de la estabilidad a temperaturas extremas (MERZ *et al.*, 2000), pH (CHERRY, 2000) y el uso de cofactores alternativos (RÖTHLISBERGER *et al.*, 2008).

Existen tres enfoques para el mejoramiento de enzimas: ingeniería de proteínas, evolución dirigida e ingeniería de consensos (MORÁN GARCÍA, 2008).

La ingeniería de proteínas o diseño racional consiste en el diseño de mutaciones en los sitios donde se cree que puedan tener un efecto sobre una función (actividad catalítica). Este enfoque depende del conocimiento *a priori* de una gran cantidad de información estructural de la molécula y conocimientos de evolución molecular. En este enfoque se debe tomar en cuenta la relación existente entre secuencia, estructura y función, ya que por una parte se tiene la información de la variabilidad de cada residuo y por el otro, la posición tridimensional y la interacción con otros residuos en el espacio (SEGOVIA & PEIMBERT, 2010). Este conocimiento está basado en análisis bioinformáticos de la secuencia de la proteína, reglas generales derivadas de caracterizar los efectos de las mutaciones sobre las propiedades de la enzima y por la implementación de modelamientos moleculares que nos permiten predecir los efectos de las mutaciones sobre la estructura (DALBY, 2007).

En la evolución experimental o dirigida se busca emular el proceso evolutivo natural como la mutagénesis al azar y la recombinación sexual. Este tipo de enfoque ha demostrado mejorar marcadamente la estabilidad, expresión y actividad de las enzimas sin necesidad de depender de información *a priori*. En este caso, los métodos computacionales pueden ser usados para obtener información útil de la secuencia de proteínas y estructuras que puedan guiar a la mutagénesis al azar de proteínas, haciendo la evolución dirigida más eficiente (RÖTHLISBERGER *et al.*, 2008).

Finalmente, la ingeniería de consensos permite incrementar la estabilidad en las proteínas sin tener que recurrir al análisis de una gran cantidad de información estructural o a la generación de bancos de miles de variantes. Este enfoque semi-racional permite predecir mutaciones estabilizantes mediante el análisis de las secuencias de aminoácidos que en cada posición tiene el residuo más común en un alineamiento múltiple, por lo tanto, el respectivo aminoácido consenso contribuye en mayor medida a la estabilidad de la proteína que el aminoácido no consenso (LEHMANN & WYSS, 2001). Este enfoque toma ventaja de la selección negativa que la naturaleza ha aplicado a todas las secuencias presentes en un alineamiento múltiple (STEIPE, SCHILLER, PLÜCKTHUN, & STEINBACHER, 1994). Las mutaciones al azar desestabilizantes son altamente probables y son purgadas por la naturaleza si la estabilidad general cae bajo cierto umbral. Como resultado, los residuos que estabilizan las proteínas tienden a ser más prevalentes que otros aminoácidos en cualquier otra posición en la familia de la proteína (STEIPE *et al.*, 1994).

En este trabajo buscamos generar dominios más estables usando la ingeniería de consensos, debido a que existen varios casos exitosos en los cuales las enzimas modificadas usando este enfoque mostraron un mejoramiento en diferentes características, como la actividad o su tolerancia a la temperatura.

DOMINIO ROSSMANN

El plegamiento Rossmann fue inicialmente identificado como una proteína de unión a dinucleótidos y permanece como uno de los plegamientos de unión a dinucleótidos más estudiados (ROSSMANN, MORAS, & OLSEN, 1974).

Los dominios Rossmann se encuentran en proteínas con una amplia gama de actividades catalíticas como oxidoreductasas, hidrolasas, liasas e isomerasas. El dominio Rossmann de unión a NAD(P) es uno de los plegamientos más antiguos, ocurre 87 veces en la clasificación de dominios y 42 veces en la clasificación de familias (MA *et al.*, 2008). Y se encuentra en el 1 % de las proteínas eucarióticas y en el 3 % de las procarióticas reportadas hasta la fecha (PERSSON, KALLBERG, OPPERMAN, & JÖRNVAL, 2003).

La superfamilia Rossmann es un ejemplo de las pocas que se combinan con muchas otras superfamilias, ya que el dominio Rossmann se encuentra en combinación con siete diferentes superfamilias de dominios catalíticos. En todos los casos donde los dominios Rossmann están unidos a miembros de una superfamilia catalítica, los dominios están en la misma posición relativa, sus interfaces tienen la misma estructura y sus regiones de enlace presentan el mismo

patrón. Estos resultados indican que la razón por la cual los miembros de una superfamilia de dominios catalíticos conservan un orden secuencial en sus conexiones a los dominios Rossmann es que cada pareja surgió de un único evento de recombinación y la superfamilia surgió de las duplicaciones subsecuentes y divergencia de este par (BASHTON & CHOTHIA, 2002).

Se pueden definir cuatro tipos diferentes de conexiones entre los dominios catalíticos y el Rossmann. El dominio Rossmann puede estar unido al extremo N terminal del dominio catalítico (tipo I) o a su extremo C terminal (tipo II); el dominio catalítico puede ser insertado dentro del dominio Rossmann (tipo III), o el dominio Rossmann puede ser insertado dentro del dominio catalítico (tipo IV). Dentro de una familia, el tipo de conexión de dominios es siempre el mismo. Las diferentes familias que forman un superfamilia dada también presentan el mismo tipo de conexiones (BASHTON & CHOTHIA, 2002).

El dominio Rossmann presenta un plegamiento tipo α/β . El plegamiento que une NAD y NADP consiste de dos motivos de unión a mononucleótidos que están estructuralmente relacionados por una pseudo doble rotación, con la mayoría de las hebras N-terminales adyacentes una a la otra (Figura 1). Juntos, los dos motivos $\beta\alpha\beta$ forman un conjunto de seis hebras β paralelas flanqueadas por hélices α , con el orden relativo de las hojas 3-2-1-4-5-6 (BOTTOMS, SMITH, & TANNER, 2002). En el dominio Rossmann más simple, la estructura inicia en la primera hebra β y termina en la sexta hebra β . El cruce de las hojas internas forma una cresta en el extremo C-terminal de la hebra β que conforma el sitio de unión para el NAD⁺. Los dominios Rossmann unen a la coenzima NAD⁺, esta coenzima participa en las reacciones redox en las cuales el anillo de nicotinamida acepta de forma reversible un hidrógeno (BASHTON & CHOTHIA, 2002).

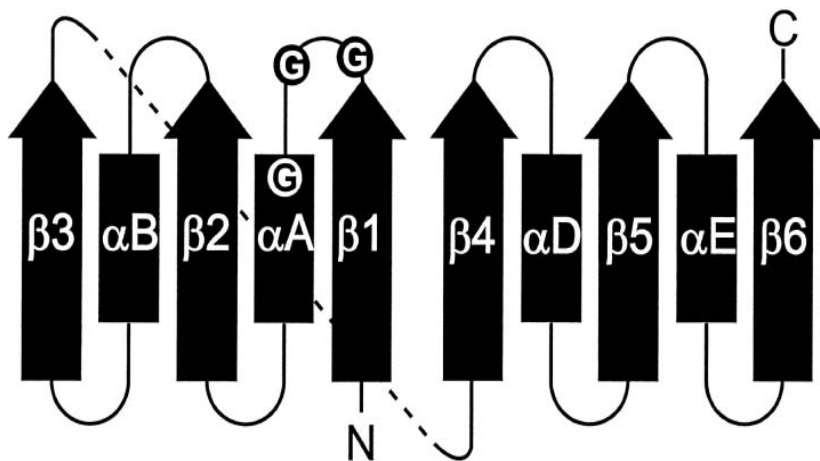


Figura 1. Dominio Rossmann. Las flechas designan hojas β y los rectángulos hélices α , los círculos representan los residuos de glicina conservados. Imagen tomada de (Bottoms et al., 2002).

Existen dos secuencias comunes características en el plegamiento Rossmann. Primero, un bucle rico en glicinas que une fosfatos y que conecta el C-terminal de $\beta 1$ con el N-terminal de αA . Típicamente, este bucle contiene tres residuos de glicina conservados arreglados en patrones como GXGXXG (motivo G) e interactúa con el grupo bifosfato del NAD. La segunda característica es la interacción de una cadena lateral (en la hebra $\beta 2$) con el grupo de la adenina ribosa del cofactor (NAD, NADP o FAD) (BOTTONS *et al.*, 2002).

El dominio Rossmann presenta una característica estructural conservada: una molécula de agua presente en la apoenzima que estabiliza al asa y al cofactor en su conformación extendida y pudiera contribuir entálpicamente a la energía libre de la unión del cofactor (BOTTONS *et al.*, 2002).

MODELO DE ESTUDIO

La enzima *shikimato* deshidrogenasa tiene una gran importancia a nivel industrial: se utiliza como materia prima en la síntesis de inhibidores de neuroaminidasas, como el *oseltamivir*, comercialmente distribuido bajo el nombre de *Tamiflú*, también es de gran interés para la generación de antimicrobiales, herbicidas y antiparasitarios, debido a que la vía del *shikimato* se encuentra solamente en bacterias, arqueas, plantas, hongos y parásitos del orden apicomplexa (MICHEL *et al.*, 2003).

La *shikimato* deshidrogenasa cataliza la reducción reversible del 3-dehidroshikimato a *shikimato* (EC 1.1.1.25) en la presencia de NADPH como donador de electrones (Figura 2). La reacción constituye el cuarto paso en la vía del *shikimato*, en la cual, a partir de fosfoenolpiruvato y 4-eritrosa fosfato, se produce *corismato*, el precursor de los aminoácidos aromáticos y de algunos metabolitos secundarios aromáticos. La vía del *shikimato* une el metabolismo de los carbohidratos con la biosíntesis de los compuestos aromáticos (HERRMANN & WEAVER, 1999). En hongos y levaduras, la *shikimato* deshidrogenasa existe como componente de un complejo pentafuncional mientras que en plantas está presente como una enzima bifuncional: deshidratasa 3-dehidroquinato (AROD, EC 4.2.1.10), y en bacterias como una enzima monofuncional sencilla (S. YE, DELFT, & BROOUN, 2003).

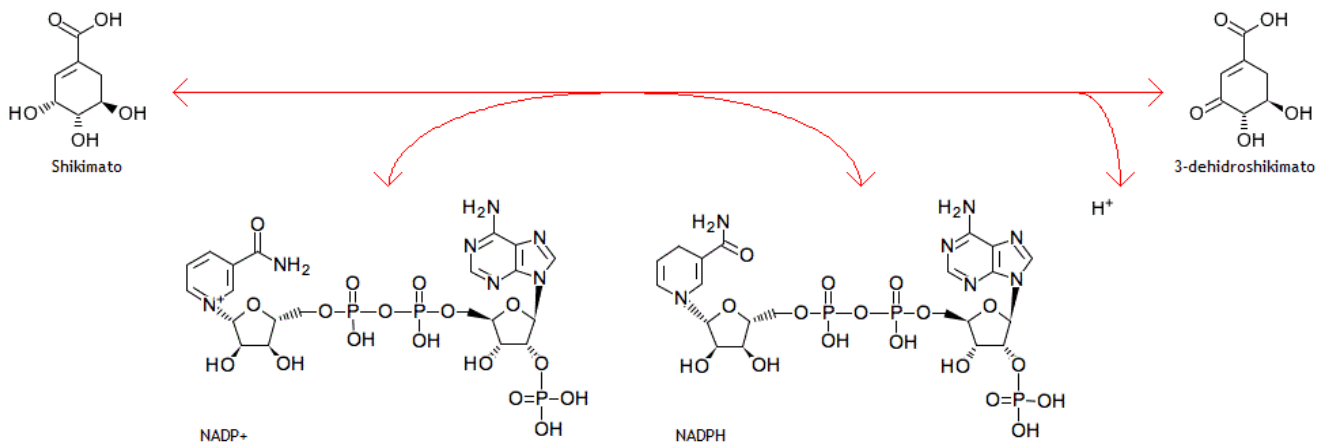


Figura 2. Reacción catalizada por shikimato deshidrogenasa (EC 1.1.1.25). La enzima shikimato deshidrogenasa, codificada por el gen *aroE* en *E. coli*, cataliza la reducción reversible del 3-dehidroshikimato a shikimato en la presencia de NADPH como donador de electrones. Imagen tomada de KEGG: http://www.genome.jp/dbget-bin/www_bget?rn:R02413

La shikimato 5-deshidrogenasa pertenece a la superfamilia de las oxidorreductasas dependientes de NADPH, la cual se encuentra presente en enzimas que participan en rutas catabólicas y anabólicas así como en la detoxificación de xenobióticos. Esta superfamilia usualmente se subdivide en varias familias, incluyendo deshidrogenasas de cadena corta, deshidrogenasas de cadena media, aldo-cetorreductasas y deshidrogenasas de cadena larga. El mecanismo de reacción usado por estas enzimas involucra la transferencia de electrones del cofactor NADPH para estabilizar la acumulación de la carga negativa en el sustrato aceptor (BENACH *et al.*, 2003).

La enzima shikimato deshidrogenasa, codificada por el gen *aroE* en *E. coli*, es un monómero de 271 aminoácidos. Presenta dos dominios que tienen arquitecturas α/β (Figura 3) y que se encuentran conectados por la hélice α_5 y un conector, Asp102-Pro118. El primer dominio está compuesto de dos segmentos discontinuos, Met1-Thr101 y Gly237-Ser271, mientras que el segundo dominio abarca Gly119-Asp236. El arreglo de estos dos dominios a lo largo de las hélices da lugar a un profundo surco en el cual se localiza el cofactor NADP+ (MICHEL *et al.*, 2003).

El dominio N-terminal encargado de la unión del sustrato (y de la dimerización en shikimato deshidrogenasas de algunas especies) es un sándwich $\alpha-\beta-\alpha$. Consiste principalmente de seis hebras β y seis hélices α . El orden de las hebras es 2-1-3-5-6-4, siendo la hebra β_5 antiparalela a las demás. Las tres primeras hebras β siguen una sucesión β/α regular, con las hélices α_1 y α_2 paralelas a las hebras β , flanqueando lados opuestos de la hebra. La siguiente unidad $\alpha/\beta/\alpha$ es irregular, con la hélice α_3 orientada aproximadamente a 45° de acuerdo a la

dirección de la hebra, y una vuelta hélice $\alpha 4$ casi perpendicular a la hebra $\beta 4$. El dominio se completa por una horquilla α helicoidal ($\alpha 9$ y $\alpha 10$), la cual se empaqueta contra la hebra β en el mismo lado de la $\alpha 1$ (MICHEL *et al.*, 2003).

El dominio C-terminal es un dominio Rossmann dependiente de NADP. Está compuesto de seis hebras β paralelas en orden 3-2-1-4-5-6 y hélices α a ambos lados. La tercera y la cuarta hélices α están remplazadas por asas irregulares, por lo cual es de los Rossmann NADP más pequeños. Los dominios interactúan por intercalamiento de cadenas hidrofóbicas: $\alpha 5$ interactúa con $\beta 10$, $\beta 11$ y $\alpha 6$ (numeradas por la posición final en la enzima con ambos dominios), y $\alpha 4$ interactúa con $\alpha 6$ y $\alpha 7$ (MICHEL *et al.*, 2003).

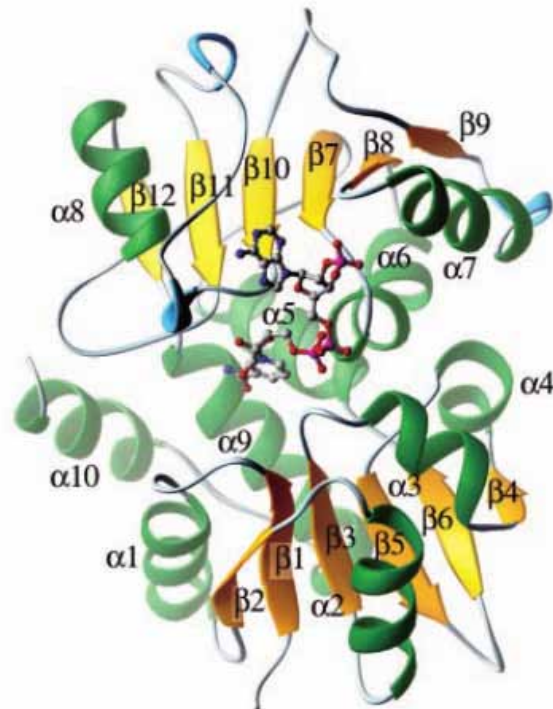


Figura 3. Representación tridimensional de AROE. En amarillo se muestran las hebras β , en verde las hélices α y en azul las asas. Imagen tomada de (MICHEL *et al.*, 2003).

El sitio activo, como en todas las aminoácidos deshidrogenasas, lo constituye una hendidura profunda conformada por dos hélices α en la interfaz entre ambos dominios y en la cual se une el sustrato por residuos principalmente del dominio N-terminal y el cofactor por residuos del dominio C-terminal. Ante la unión del sustrato y el cofactor hay un movimiento de los dominios de una estructura abierta a una cerrada. El movimiento corresponde a una rotación de aproximadamente 25° por el eje que pasa por los residuos Gln26 y Asp102 lo que genera que la punta del dominio N-terminal recorra una distancia de aproximadamente 14 \AA .

Los residuos implicados en la formación de los puentes de hidrógeno que se modifican ante el movimiento conformacional son Asn86, Thr101, Gln244, Thr87 y Asn59. Los tres primeros residuos (Asn86, Thr101 y Gln244) están altamente conservados en la familia de *shikimate* deshidrogenasa, Thr87 en el 92 % de las secuencias y Asn59 en el 60 %. En la conformación abierta, Gln244 forma puentes de hidrógeno con las cadenas laterales de Asn59 y Thr101. Cuando se une el sustrato y se adquiere la conformación cerrada, Gln244 genera nuevas interacciones con Asn86 y mantiene aquellas con Thr101. Una vez formado el sitio activo, el carbono 3 del 3-dehidroshikimate se posiciona a 3 Å de distancia para recibir el hidrógeno del carbono 4 del anillo nitrogenado del NADPH (MICHEL *et al.*, 2003). El mecanismo de acción propuesto plantea que la transferencia de hidruros y la protonación del carbono 3 del dehidroshikimate ocurren de manera concertada. La Lys65 funciona como un ácido/base general. El protón de la Lys65 es recuperado o liberado posteriormente del o hacia el agua presente en el sitio activo por coordinación con otro residuo, Asp102. La molécula de agua, los residuos que unen al sustrato y los que intervienen en la reacción están conservados en toda la familia de *shikimate* deshidrogenasa (ESPINOSA CANTÚ, 2010).



SECCIÓN
02



ANTECEDENTES



*Diseño y Caracterización
de un Dominio
Rossmann Consenso*



CLAUDIA IRIS BRAVO BONILLA

ANTECEDENTES

ANTECEDENTES DEL LABORATORIO

En el laboratorio del Dr. Lorenzo Segovia se trabaja en la construcción y análisis de proteínas quiméricas generadas a partir de la combinación del dominio Rossmann y ambas partes del dominio catalítico de los genes *aroE* (denominado como E) y *ydiB* (denominado como Y) de *Escherichia coli*, y *aroE* de *Bacillus subtilis* (denominado como B).

Los genes *aroE* y *ydiB* de *Escherichia coli*, son xenólogos (SINGH, STAVRINIDES, CHRISTENDAT, & GUTTMAN, 2008), presentan un 25 % de identidad en secuencia, y codifican proteínas con actividad de shikimato deshidrogenasa (AroE, EC 1.1.1.25) y shikimato/quinato deshidrogenasa (YdiB, EC 1.1.1.282) respectivamente, las cuales presentan un 29 % de identidad en secuencia peptídica. Ambas estructuras terciarias están compuestas de un dominio Rossmann de unión a dinucleótidos y un dominio de unión a sustrato (MICHEL *et al.*, 2003). El gen *aroE* de *Bacillus subtilis* presenta un 35 % de identidad con *aroE* de *Escherichia coli* y también presenta un dominio Rossmann de unión a dinucleótidos y un dominio de unión a sustrato.

De todas las proteínas quiméricas construidas sólo la quimera denominada EBE (la cual consiste en el dominio N-terminal de AroE de *E. coli*, dominio C-terminal de AroE de *B. subtilis* y las dos últimas α hélices del dominio C-terminal de AroE de *E. coli*) fue capaz de expresarse, ser soluble, y presentar una actividad *in vitro* e *in vivo* similar a las enzimas parentales.

En el 2008, siguiendo otro enfoque y modelo de estudio diferente: Areli Morán construyó una cistatina consenso realizando un alineamiento de 106 secuencias de fitocistatinas, con un punto de corte establecido del 30 %, cambiando 32 de 106 aminoácidos a lo largo de todo el polipéptido, respetando solamente los residuos responsables de la interacción con el sitio activo de las cisteín-proteasas. Esta cistatina consenso se expresó de forma adecuada en la fracción soluble, mostró un incremento en su actividad proteolítica junto con un aumento de 23 °C en su temperatura media de desnaturalización (MORÁN GARCÍA, 2008).

ANTECEDENTES DE INGENIERÍA DE CONSENSOS

Uno de los primeros trabajos que sugiere un método para mejorar la estabilidad de proteínas sin ningún conocimiento previo sobre su estructura o los principios que gobiernan su estabilidad

es el de Serrano *et al* quienes probaron mediante mutagénesis sitio dirigida de la barnasa (EC 3.1.27.3), el impacto de 17 aminoácidos diferentes entre las secuencias de la barnasa y la binasa (ambas ribonucleasas de 110 aminoácidos que comparten un 85 % de identidad en secuencia) en la estabilidad de la proteína, y observaron efectos entre +1.1 y -1.1 kcal mol⁻¹. Una mutante múltiple que combinaba seis sustituciones mostró una estabilidad incrementada de 3.3 kcal mol⁻¹ sobre la enzima silvestre. Los resultados indicaban que la comparación entre secuencias de enzimas homólogas mesófilas podría contener información importante para incrementar la estabilidad de las proteínas (SERRANO, DAY, & FERSHT, 1993).

Este enfoque comenzó a utilizarse en pequeños motivos o familias pequeñas de proteínas sumamente relacionadas. Nikolova *et al* trabajaron con la proteína supresora de tumores p53, realizando un alineamiento de homólogos de 23 diferentes especies, y se mutaron individualmente 20 residuos no consensos de p53 de humano a su respectivo residuo consenso. Los cambios en la estabilidad fueron desde +1.27 a -1.49 kcal mol⁻¹. Cuatro mutaciones estabilizantes se combinaron en una mutante cuádruple la cual fue estabilizada en 2.65 kcal mol⁻¹ y mostró un incremento de 5.6 °C en la temperatura media de desnaturalización (NIKOLOVA, HENCKEL, LANE, & FERSHT, 1998).

En 1999, Wang *et al* crean dos variantes funcionales y altamente estables de la mini chaperona GroEL, las cuales ayudan en el repliegamiento de proteínas recombinantes, usando un alineamiento de secuencias de 130 miembros de la familia de la chaperonina 60 para designar 37 mutaciones individuales. Se combinaron seis mutaciones estabilizantes que dieron como resultado dos mutantes múltiples con 6.99 y 6.15 kcal mol⁻¹ más estables que la proteína nativa, demostrando que un enfoque basado en la homología es un método simple y eficiente para incrementar la estabilidad de una proteína (WANG, BUCKLE, FOSTER, JOHNSON, & FERSHT, 1999).

Steipe *et al* predijeron sustituciones estabilizantes basadas en la comparación de la frecuencia de aparición de aminoácidos en secuencias de dominios variables V_L y V_H de inmunoglobulinas, asumieron que el aminoácido más frecuente en cualquier posición en un alineamiento de dominios variables V_L o V_H de inmunoglobulinas homólogas contribuía más a la estabilidad del dominio que aquel aminoácido que solo ocurre una vez. Predijeron diez mutaciones estabilizantes para el dominio variable V_H de inmunoglobulina, de las cuales seis fueron estabilizantes y combinadas incrementaban 20 veces el tiempo medio de desnaturalización térmica (WIRTZ & STEIPE, 1999). Estos autores postularon aunque la probabilidad de que las mutaciones al azar sean desestabilizantes es muy alta, son selectivamente neutras siempre que el dominio en su estabilidad total no caiga bajo cierto valor de umbral. A la inversa, las mutaciones al azar que resultaron en una estabilidad térmica

incrementada son altamente improbables en la ausencia de una selección positiva (LEHMANN, PASAMONTES, LASSEN, & WYSS, 2000). Como resultado, los residuos que estabilizan a la proteína tienden a ser más prevalentes que otros aminoácidos en determinadas posiciones.

Es a partir de este trabajo que la ingeniería de consensos se aplica por primera vez a la secuencia de una proteína completa: Lehmann *et al* definieron 18 aminoácidos consenso a partir de 13 secuencias de fitasas (EC 3.1.3.8) de hongos mesofílicos, lo que dio como resultado una primera fitasa consenso, la cual mostraba una temperatura de desplegamiento (T_m) de 78 °C, que es 15 - 22 °C más alta que cualquier valor de T_m de las fitasas parentales usadas para el diseño. Un refinamiento de este enfoque, combinado con experimentos de mutagénesis sitio-dirigida dio como resultado una fitasa consenso optimizada con valores de T_m arriba de 90.4 °C. Estos incrementos en la T_m se deben a la combinación de intercambios múltiples de aminoácidos, los cuales estaban distribuidos en toda la secuencia de la proteína y que afectan principalmente a los residuos expuestos en la superficie o en regiones sin estructura secundaria definida, donde cada sustitución individual tiene un pequeño efecto termo-estabilizador (Lehmann *et al.*, 2000). Cualquiera de estas secuencias consenso finales diferían de las secuencias parentales más cercanas por un 18 % (80 residuos de 444).

La ingeniería de consensos también ha sido aplicada a dominios SH3 (MAXWELL & DAVIDSON, 1998), proteínas con repeticiones ricas en leucina (BINZ, STUMPP, FORRER, AMSTUTZ, & PLÜCKTHUN, 2003), proteínas ankirina (BINZ *et al.*, 2003), dominios WW (JIANG, KOWALSKI, & KELLY, 2001), entre otros. Recientemente, una versión consenso de gp120 fue diseñada y expresada para su uso como vacuna, y aun cuando los datos de estabilidad no se reportaron, tuvo una mejor expresión en comparación con cualquiera de los gp120 parentales (GAO, WEAVER, LU, LI, & LIAO, 2005).

En muchos de estos casos, la estabilidad óptima fue incrementada no solamente al producir secuencias consenso sino al combinar consensos con diseño racional basado en la covariación de aminoácidos, consideraciones de la estructura de la proteína, datos de estabilidad previamente definidos, métodos computacionales o datos de actividad. Por ejemplo, Amin y colaboradores usaron mutagénesis consenso combinatoria (CCM, *combinatorial consensus mutagenesis*) para generar variantes de proteínas de β -lactamasa (BLA) de *Enterobacter cloacae* al alinear 38 secuencias de homólogos de BLA. Una de las variantes que contenía ocho mutaciones consenso mostraba una temperatura media de desnaturalización (T_m) 9.1 °C mayor que la parental (AMIN, LIU, & RAMER, 2004). Binz desarrolló un andamiaje conservado de la proteína ankirina, en el cual se conservaban los residuos importantes para mantener la repetición de la estructura y en el que los aminoácidos potenciales que interactuaban fueran aleatorizados para mejorar las interacciones entre las unidades repetidas de la proteína

ankirina. Esto resultó en un incremento en la expresión, solubilidad y estabilidad de la proteína repetida ankirina (BINZ *et al.*, 2003).

Vázquez *et al* usaron la secuencia consenso en combinación con criterios basados en la estructura para la generación de una glucosa deshidrogenasa (EC 1.1.1.47) termo-estable. Se generaron 24 variantes, de las cuales 11 mostraron una estabilidad térmica superior a la de la enzima nativa. La mejor variante mostraba una vida media de aproximadamente 3.5 días a 65 °C en contraste con la nativa (cuya vida media era de 20 minutos a 25 °C) (VÁZQUEZ-FIGUEROA, CHAPARRO-RIGGERS, & BOMMARIUS, 2007).

La creación *ab initio* de proteínas a partir de la información obtenida por consensos ha generado casos exitosos, como la proteína novel fluorescente por parte de Dai *et al.* Esta proteína se creó a partir de un alineamiento de 31 secuencias de proteína fluorescentes homólogas a la proteína monomérica verde Azami (mAG), la cual fue usada como secuencia guía cuando el porcentaje de conservación de cierta posición del alineamiento fuese menor al 40 %, para eliminar de esta forma los problemas relacionados con la co-variación de los aminoácidos y disminuir las probabilidad de agregar una mutación que causara la pérdida de fluorescencia. La proteína consenso obtenida tuvo una mejor expresión (llegándose a obtener más de 300 mg al final del proceso de purificación), era monomérica, y aunque fue un poco menos estable (térmicamente) que la mAG parental sus propiedades fluorescentes fueron mucho mejores aun cuando sólo difiere por un 10 % de las secuencias parentales (DAI *et al.*, 2007).

Sullivan *et al* diseñaron diferentes variantes de triosafosfato isomerasa (TIM, EC 5.3.1.1) consenso que diferían en un 30 - 40 % de cualquier TIM nativa, pero que entre ellas sólo diferían por un pequeño número de posiciones sin conservar, ya que presentaban un 85 % de identidad en secuencia. La primer TIM consenso era ligeramente activa y presentaba características de glóbulo fundido y era monomérica aun cuando las enzimas parentales son diméricas. La segunda TIM consenso mostró parámetros cinéticos similares a los de las enzimas nativas y una estabilidad térmica incrementada. Aún cuando ambas presentan un 85 % de identidad en secuencia, se originaron a partir de bases de datos de diferente tamaño (639 vs 781 secuencias) pero se cree que esta diferencia no tiene nada que ver con las diferentes características de cada una de las TIM consenso (SULLIVAN, DURANI, & MAGLIERY, 2011).

El éxito de la ingeniería de consensos depende de las secuencias usadas en el diseño, esto lo observaron Jacobs *et al* cuando diseñaron tres dominios FN3 consenso, que presentaban un 40 - 45 % de identidad en secuencia al dominio FN3 más cercano. Los consensos se expresaron en *E. coli*, fueron solubles y altamente estables (78 - 79 °C) y presentaban el mismo plegamiento a pesar de su bajo porcentaje de identidad en secuencia (45 % entre

ellos). El dominio FN3 consenso que presentó menor estabilidad fue el que se derivó de un alineamiento de las secuencias más estables, esto podría deberse a que había pocas secuencias para formar un consenso verdadero: un conjunto mayor de datos produce diseños de consenso más estables que un conjunto pequeño de datos de las moléculas más estables. Otra explicación podría ser que los resultados óptimos para el diseño de consensos depende de conjuntos de secuencias menos estables (JACOBS *et al.*, 2012).

Finalmente, Aerts *et al* realizaron secuencias consenso de la enzima sucrosa fosforilasa (EC 2.4.1.7) usando diferentes alineamientos múltiples de secuencias. Estas enzimas consenso se plegaban en proteínas estables y mostraban actividad similar a la enzima nativa aun cuando diferían de su homólogo más cercano en 70 residuos. Este éxito se debió a que la red de co-evolución se mantuvo intacta y al uso de una secuencia guía para determinar el consenso en regiones muy variables. Estas enzimas consenso presentaban un ligero aumento de 4 °C en su estabilidad térmica, esto puede atribuirse a la diferencia intrínseca de las estabilidades térmicas de las secuencias parentales de los respectivos alineamientos (AERTS *et al.*, 2013).



SECCIÓN
03



HIPÓTESIS



*Diseño y Caracterización
de un Dominio
Rossmann Consenso*



CLAUDIA IRIS BRAVO BONILLA

HIPÓTESIS

El reemplazar varios residuos con el aminoácido consenso correspondiente, incrementará la estabilidad conformacional del dominio Rossmann de una enzima shikimato deshidrogenasa.



SECCIÓN
04



OBJETIVOS



*Diseño y Caracterización
de un Dominio
Rossmann Consenso*



CLAUDIA IRIS BRAVO BONILLA

OBJETIVOS

- Diseño de una enzima *shikimate* deshidrogenasa con un dominio Rossmann consenso.
- Construcción de una enzima *shikimate* deshidrogenasa con un dominio Rossmann consenso.
- Caracterización bioquímica de una enzima *shikimate* deshidrogenasa con un dominio Rossmann consenso.

OBJETIVOS PARTICULARES

- Generar diferentes secuencias consenso al variar los conjuntos iniciales de secuencias homólogas.
- Determinar la localización de los cambios en la estructura secundaria de la enzima *shikimate* deshidrogenasa.
- Determinar el efecto que tiene el conjunto inicial de secuencias en el consenso.
- Caracterizar la estructura secundaria mediante dicroísmo circular.
- Determinación de la actividad *in vitro* e *in vivo* de una enzima *shikimate* deshidrogenasa con un dominio Rossmann consenso.



SECCIÓN
05



ESTRATEGIA EXPERIMENTAL



*Diseño y Caracterización
de un Dominio
Rossmann Consenso*



CLAUDIA IRIS BRAVO BONILLA

ESTRATEGIA EXPERIMENTAL

DETERMINACIÓN DE LA SECUENCIA CONSENSO

Los homólogos de *shikimate* deshidrogenasa fueron identificados a partir de proteínas completas al realizar un BLASTp (J. YE, MCGINNIS, & MADDEN, 2006) en una base de datos no redundantes de secuencias de proteínas, teniendo como semilla la secuencia de AroE de *E. coli* (Anexo 1) y usando un valor de esperanza de $E = 10^{-3}$ (BANDA, 2013).

Usando el programa CD-HIT v4.1 (LI, WEIZHONG, & GODZIK, 2012) el conjunto de datos fue filtrado para reducir la redundancia hasta un valor máximo del 80 %. La región correspondiente al dominio Rossmann se extrajo usando EMBOSS 6.3.1.: extractalign (RICE, P., LONGDEN, I., & BLEASBY, 2000). Estos alineamientos se importaron a ClustalX (THOMPSON, GIBSON, PLEWNIAK, JEANMOUGIN, & HIGGINS, 1997) para realizar un dendograma, a partir de este se obtuvieron los dos primeros conjuntos de datos para obtener las dos primeras secuencias consenso. El tercer conjunto de datos correspondía a las secuencias de arqueas, estas se obtuvieron mediante la información obtenida con *Taxonomic Common Tree* («Taxonomic Common Tree», 2012), a las cuales se les extrajo la región correspondiente al dominio Rossmann. Las secuencias consenso se realizaron con el programa *SeaView* (GOUY, GUINDON, & GASCUEL, 2010), estableciendo un punto de corte del 30 % (Anexo 2 y 3).

DISEÑO Y SÍNTESIS DEL GEN CON LA SECUENCIA CONSENSO

A las tres secuencias consenso se les anexó la secuencia del dominio catalítico de AroE y una etiqueta de seis histidinas. Las secuencias de aminoácidos fueron convertidas a nucleótidos usando el programa *Primo Optimum 3.6 Optimal Gene Synthesis And Expression* de la compañía *Chang Biosciences* (California, USA). Las secuencias obtenidas se sintetizaron de forma comercial por la compañía *Life technologies* (California, USA) (Anexo 4).

De acuerdo a la información proporcionada por la compañía: los genes sintéticos fueron ensamblados a partir de oligonucleótidos sintéticos y/o productos de PCR. Los fragmentos fueron clonados en pMK-RQ (KanR) usando el sitio de clonación *SfiI* (Figura 4). El plásmido de ADN fue purificado y transformado en bacterias y su concentración fue determinada mediante espectroscopía UV. La construcción final fue verificada mediante secuenciación. La congruencia de la secuencia con los sitios de restricción usados fue de 100 %.

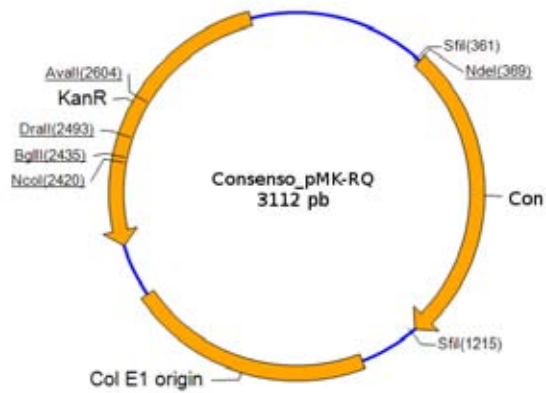


Figura 4. Vector *pMK-RQ* de Life Technologies. Plásmido multicopia con un sitio de origen Col E1 y un gen de resistencia a kanamicina (kanR), donde Con puede ser cualquiera de los tres genes consenso.

A los genes consenso se les adicionó mediante PCR los sitios de corte de las enzimas NcoI y XhoI para su posterior clonación en el vector de expresión pT4 que presenta un gen de resistencia a kanamicina (KanR) y un promotor trc fuerte, usando las respectivas endonucleasas de restricción.

Debido a que el plásmido de expresión presentaba adicionalmente un gen que codifica para una enzima quimérica (EBE.His) fue necesario realizar una reacción de doble digestión para retirar el gen EBE.His y dejar el vector de expresión vacío (Figura 5).

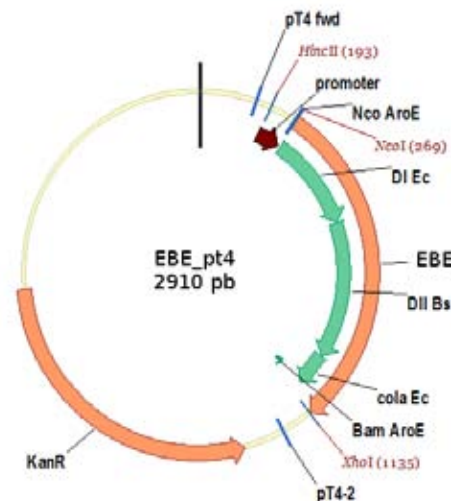


Figura 5. Vector de expresión *pT4*. Plásmido multicopía con un gen de resistencia a kanamicina (kanR) y un promotor trc fuerte, donde EBE es una enzima quimérica.

El vector de expresión se obtuvo al realizar un *maxiprep* de un cultivo saturado de EBE. His-DH5α en 50 mL de medio LB con kanamicina usando el protocolo del kit *Zyppy Plasmid Maxiprep* de *Zymo Research Corp* (California, USA). Se realizó una reacción de doble digestión con las enzimas NcoI y XhoI (New England Biolabs; Massachusetts, USA) para retirar el gen EBE.His y dejar el vector de expresión vacío.

La reacción de doble digestión se realizó en un volumen de 100 μL con 60 unidades de cada una de las enzimas en el buffer adecuado (NEBuffer4, New England Biolabs; Massachusetts, USA) y 2000 ng del vector, con un tiempo de incubación de 4 horas a 37 °C y 10 minutos a 65 °C para desactivar a las enzimas. Se comprobó la escisión del vector en un gel de agarosa y posteriormente se cortó la banda correspondiente al vector vacío para ser purificada usando el protocolo *Purification of DNA fragments from agarose gel* del kit *High pure PCR product purification* de Roche Applied Science (Bavaria, Alemania).

Mediante PCR se insertaron los sitios de corte de las enzimas NcoI y XhoI (en subrayado y cursiva) usando los siguientes oligos: NcoConF (5´ - GAC AGG TCC ATG GAA ACC TAT GCT GT-3´) como oligo directo, y XhoCON1R (5´ - CTC GAC TCG AGT CAG TGG TGG TGG T-3´) como oligo reverso, ambos fueron sintetizados en la Unidad de Síntesis y Secuenciación del Instituto de Biotecnología de la UNAM (Morelos, México).

Las reacciones de PCR se realizaron en volúmenes de reacción de 50 μL con 1 unidad de Taq ADN polimerasa (recombinante) en la presencia del buffer apropiado (Fermentas; Vilnius, Lituania), 200 μM de dNTPS, 15 pmol de cada oligo y 15 ng del ADN templado. Las mezclas de PCR se desnaturalizaron al calentar a 94 °C durante 4 minutos. Se realizaron 20 ciclos de amplificación usando las siguientes condiciones: 30 s a 94 °C, 45 s a 54 °C, 45 s a 72 °C. Y un último ciclo de amplificación de 5 min a 75 °C. Llevado a cabo en el equipo iCycler de Bio-Rad (California, USA).

Se confirmó mediante un gel de agarosa al 0.8 % la presencia de los productos de PCR (con un peso aproximado de 834 pb), se purificaron usando el protocolo *purification of PCR products in solution after amplification* del kit de *Roche Applied Science* (Bavaria, Alemania): *High Pure PCR Product Purification Kit*.

Los productos de PCR purificados fueron ligados en el vector de expresión (pT4) vacío usando la enzima t4 DNA ligasa (Fermentas; Vilnius, Lituania), la reacciones se incubaron durante una hora a temperatura ambiente. Para realizar la ligación se utilizó el siguiente protocolo que consiste en estimar la cantidad de vector e inserto a ligar basándose en la intensidad de las bandas de ADN en un gel de agarosa teñido con bromuro de etidio (Gratchev, s. f.) y usando la siguiente fórmula:

$$L_i = 3 * L_v * \left(\frac{S_i}{S_v} \right) * \left(\frac{I_v}{I_i} \right) * \left(\frac{V_i}{V_v} \right)$$

Donde S_v es el tamaño del vector (en pb), S_i es el tamaño del inserto (en pb), I_v es la intensidad relativa de la banda correspondiente al vector, I_i es la intensidad relativa de la banda correspondiente a los insertos, V_v y V_i son los volúmenes (μL) del vector y el inserto que se cargaron en el gel, L_v es el volumen (μL) del vector que queremos utilizar, y L_i es el volumen (μL) de inserto que se debe usar para cada reacción.

El producto de la ligación fue transformado en células quimio-competentes de *Escherichia coli* XL10-Gold (genotipo: endA1 gyrA96 lac Δ (mcrA)183 Δ (mcrCB- hsdSMR-mrr)173 recA1 thi-1 relA1 supE44 Hte F'[lacI^q lacZ Δ M15 proAB⁺ Tn10(Tet^r) Amy Cam^r] (CASALI, 2003)), la cual es una cepa de células ultracompetentes que permite la transformación de moléculas grandes de ADN con una alta eficiencia y asegura la estabilidad del inserto. Todas las células se hicieron competentes usando el protocolo descrito por Tu *et al* (Tu *et al.*, 2005).

Se seleccionó una colonia de las células con el producto de ligación, se creció en un cultivo líquido de medio LB con kanamicina y se extrajo ADN usando el protocolo *Isolation protocol for small-scale (mini) preparations of purified plasmid ADN* del kit *High pure plasmid isolation* de Roche Applied Science (Vilnius, Lituania). Se mandó a secuenciar en la Unidad de Síntesis y Secuenciación de ADN del Instituto de Biotecnología de la UNAM (Morelos, México), usando los oligos pt4 (GAC ATA TAA ACG GTT CTG GCA) y pt4-2 (GCC AGT GAA TTG TAA TAC GAC) (Figura 6). Las muestras se cuantificaron mediante gel de agarosa y NanoDrop (Delaware, USA).

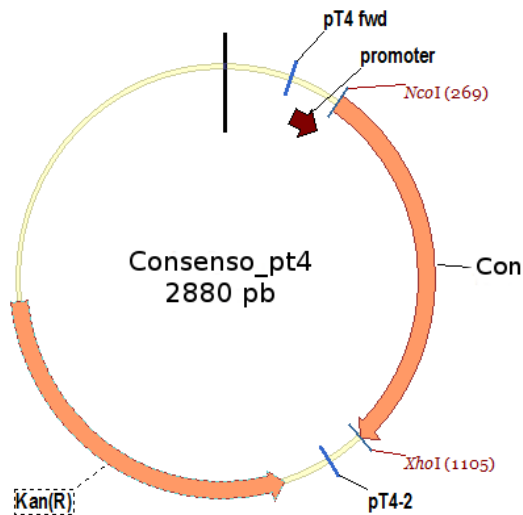


Figura 6. pT4 consenso. Plásmido multicopia que presenta un gen de resistencia a kanamicina (kan^R), un promotor trc fuerte. Donde Con puede cualquiera de las tres secuencias consenso.

Los plásmidos de expresión resultantes fueron secuenciados y clonados en diferentes cepas quimio-competentes de *Escherichia coli*: XL10-Gold, JM101 y BW25113 Δ AroE.

EXPRESIÓN

Para la producción de enzima, 2 % de un cultivo de *Escherichia coli* JM101 transformado con el vector que contenía el gen de interés fue inoculado en 500 mL de medio LB (triptona 10 g/L, NaCl 10 g/L, extracto de levadura 5 g/L) que contenían 25 mg/mL de kanamicina, en un matraz de 3 L e incubado a 37 °C con agitación continua a 200 rpm hasta obtener una DO (a 600 nm) de 6.

La biomasa producida fue recolectada por centrifugación durante 10 minutos a 5000 rpm (3214 x g) y 4 °C, lavada con 100 mL de buffer PBS (137 mM NaCl, 2.7 mM KCl, 10 mM NaH_2PO_4 , 1.8 KH_2PO_4 , pH 7.4), el pellet de células obtenido se disolvió en 100 mL de buffer de lisis (50 mM buffer de fosfatos, 300 mM NaCl, 10 mM imidazol, pH 8.0) suplementado con lisozima (1.5 mg/gramo de pellet) y PMSF (fluoruro de fenilmetilsulfonilo, 0.5 mM). Esta suspensión celular fue incubada en hielo por 30 minutos y sonicada durante 15 minutos con intervalos de 30 s de sonicación y 30 s de descanso, para posteriormente separar las fracciones soluble e insoluble mediante centrifugación a 11000 rpm (15557 x g) durante 15 minutos a 4 °C.

Se confirmó la presencia de la proteína de interés mediante un gel SDS-PAGE (al 15 %, adicionado con tricloroetanol) de la fracción soluble. La electroforesis se llevó a cabo en buffer Tris-glicina. El buffer de carga consistía en Tris-HCl 50 mM pH 6.8, 2 % SDS, 10 % glicerol, 1 % β -mercaptoetanol, 12.5 mM EDTA, 0.02 % bromofenol. El marcador de peso molecular utilizado fue *PageRuler Plus Prestained Protein Ladder* de Thermo Scientific (Massachusetts, USA). Todas las imágenes se tomaron usando el programa *Image Lab 3.0* con el equipo *Gel-Doc EZ* de BioRad (California, USA).

En caso de que no se observase expresión, se realizó un western blot del cultivo para confirmar la presencia de la proteína. El gel SDS-PAGE se extrajo de la cámara y se colocó en 20 mL de buffer de transferencia (14.4 g/L glicina, 3 g/L Trizma base, 200 (ml metanol) durante al menos 1 minuto. Después el gel se colocó en el Trans-Blot Turbo (Bio-Rad; California, USA), se utilizaron 26 V durante 30 minutos para llevar a cabo la transferencia a la membrana de fluoruro de polivinilideno (PVDF) con un tamaño de poro de 0.45 μm . La membrana se incubó con TBST (50 mM Tris-HCl, 150 mM NaCl, 0.05 % Tween 20) adicionado con leche en polvo al 5 % durante 4 horas a 4 °C y posteriormente se lavó con TBST durante 10 minutos por triplicado. La membrana se incubó con Anti-His₆-Peroxidasa (Roche; Basel, Suiza) durante

una hora y después se lavó con TBST durante 10 minutos por triplicado. La membrana fue tratada con Luminol e inmediatamente se analizó en el Chemidoc (Bio-Rad; California, USA). Se lavó la membrana con agua destilada y se agregó Novex HRP Chromogenic Substrate (TMB; Life technologies; California, USA).

PURIFICACIÓN

Una vez confirmada la presencia de la proteína consenso en la fracción soluble, ésta se pasó a través de una columna de níquel (HisTrap FF crude de 5 mL; GE Healthcare, Uppsala, Suecia) para purificar mediante cromatografía de afinidad las proteínas con etiqueta de histidinas. La elución se realizó mediante un gradiente escalonado de imidazol (buffer de fosfatos de sodio 50 mM pH 8.0, 500 mM NaCl, con concentraciones de imidazol: 20, 35, 75, 100, 150, 200, y 250 mM).

La presencia de la proteína con etiqueta de histidinas fue confirmada a través de un gel SDS-PAGE. La fracción que contenía la proteína de interés fue concentrada usando un Vivaspin (corte de peso molecular: 10 kDa; GE Healthcare, Uppsala, Suecia) y después dializada durante 12 horas en buffer de fosfatos de sodio 50 mM, pH 8.0. Las muestras fueron dializadas nuevamente durante 24 horas en buffer desgasificado (fosfatos de sodio 10 mM, pH 8.0) y posteriormente filtradas mediante una unidad de filtración Millex-GS de 0.22 µm (Merck Millipore; Darmstadt, Alemania).

La pureza y concentración de la proteína fueron juzgadas mediante SDS-PAGE y Bradford (Bio-Rad; California, USA), respectivamente.

CARACTERIZACIÓN

Para los experimentos de dicroísmo circular se usó un espectropolarímetro Jasco J-710 (Jasco; Tokio, Japón) con un peltier Jasco (Tokio, Japón) para controlar la temperatura, en una celda de cuarzo de 0.1 cm de ancho (Hellma; Müllheim, Alemania). Los experimentos se realizaron en el laboratorio de la Dra. Gloria Saab en el Instituto de Biotecnología de la UNAM.

Los espectros de uv-lejano para determinar la estructura secundaria fueron realizados por cuatuplicado desde 190 a 260 nm. Los barridos de longitud de onda medían la elipticidad cada nanómetro con un tiempo de integración de 8 s y una velocidad de 20 nm/min a una temperatura constante de 16 °C.

Para la desnaturalización térmica, la elipticidad fue monitoreada a 222 nm con temperaturas desde 10 hasta 88 °C. Los datos fueron colectados en pasos de 1 °C con 8 s para equilibrar la temperatura y una rampa de 30 °C/hora. En el caso de replegamiento, los datos fueron colectados en pasos de 5 °C con 8 s para equilibrar la temperatura y una rampa de 40 °C/hora.

En el caso de la desnaturalización con urea: muestras independientes, conteniendo 0.3 mg de proteína fueron preparadas con concentraciones de urea desde 0 hasta 9 M (con un total de 46 puntos) en buffer de fosfatos 10 mM en un volumen de 1 mL y equilibradas durante 12 horas a 25 °C. Se colectaron los espectros de elipticidad desde 200 hasta 260 nm cada nanómetro. Las muestras se dejaron un tiempo de 3 min para equilibrar la temperatura antes de iniciar el barrido a una velocidad de 20 nm/min con una temperatura constante de 25 °C.

Los experimentos de calorimetría diferencial de barrido (*Differential SCanning Calorimetry*, DSC) fueron realizados en un equipo microcalorímetro VP-DSC (MicroCal; Massachusetts, USA) en el Instituto de Química de la UNAM por el Dr. Miguel Costas. El buffer en el cual la enzima fue dializada se colocó en la celda de referencia. Se realizó un escaneo del buffer para calcular la línea base. Los experimentos se realizaron a una concentración de proteína de 0.4 mg/ml y con diferentes velocidades de lectura (desde 10 K/h hasta 150 K/h).

Muestras de diferentes lotes de Con1 se analizaron mediante espectrometría de masas de alta resolución en el Laboratorio Universitario de Proteómica del Instituto de Biotecnología de la UNAM. Las muestras se corrieron en un gel SDS-PAGE que se tiñó con *Comassie* para poder cortar la banda correspondiente a Con1. Las muestras fueron reducidas con ditiotreitól, alquiladas con iodoacetamida (Sigma-Aldrich, Missouri, USA) y digeridas “*in gel*” con tripsina (Promega, Wisconsin, USA). Los péptidos resultantes fueron aplicados en un sistema LC-MS con bomba de nanoflujo EASY-nLC II (Thermo Fisher, Massachusetts, USA) acoplado a un espectrómetro de masas LTQ-Orbitrap Velos (Thermo Fisher, Massachusetts, USA) con sistema de ionización tipo nano-electrospray. En el sistema de cromatografía de líquidos se utilizó un sistema gradiente de 10-80% de solvente B (agua/acetronitrilo con 0.1% de ácido fórmico) en 120 minutos utilizando una columna capilar hecha en casa (ID 0.75 μ m y 10 cm largo RP-C18). El flujo de sistema LC fue de 300 nanolitros/minuto. Para la fragmentación de los péptidos se utilizó los métodos de CID (*Collision-Induced Dissociation*) y HCD (*High Energy Collision Dissociation*) donde solamente los iones de carga 2+, 3+, y 4+ fueron seleccionados para los eventos de fragmentación. Fueron desconsiderados los iones con cargas 1+, superiores a 5+ y de cargas indefinidas. Todos los espectros fueron adquiridos en modo de detección positivo. La ejecución y captura de los datos de fragmentación fueron realizados de forma dependiente del escaneo total de iones según las cargas pre-determinadas con un ancho de aislamiento de

3.0 (m/z), energía de colisión normalizada de 35 unidades arbitrarias, activación Q de 0.250, tiempo de activación de 10 milisegundos y tiempo máximo de inyección de 10 milisegundos por micro-escaneo. Los datos espectrométricos fueron sometidos a la búsqueda contra la base de datos del NCBIInr a través del programa ProteinProspector, así como también fueron sometidos a la búsqueda restringida contra ODB de UniProt de *E. coli* y no restringida utilizando la base de datos de SwissProt en el *Proteome Discoverer* 1.4.

ENSAYOS DE ACTIVIDAD

Se monitoreó la actividad *in vivo* e *in vitro* de las enzimas consenso. Los ensayos de actividad *in vivo* se realizaron al complementar la cepa BW25113 Δ AroE con las diferentes construcciones consenso y monitorear su crecimiento en medio mínimo M9 (0.4 % glucosa, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, Na_2HPO_4 6.0 g/L, KH_2PO_4 3.0 g/L, NaCl 0.5 g/L, NH_4Cl 1.0 g/L, tiamina 10 mg/mL) adicionado con 25 mg/mL de kanamicina.

Se realizó un cultivo de 3 mL de LB con 25 mg/mL de kanamicina de cada una de la muestras, una vez saturado se centrifugó (5000 rpm, 2 min), el pellet se resuspendió en 1 mL de sales M9 1x (Na_2HPO_4 6.0 g/L, KH_2PO_4 3.0 g/L, NaCl 0.5 g/L, NH_4Cl 1.0 g/L) y se centrifugó nuevamente, este proceso se repitió tres veces. Se tomaron 100 μ L de el pellet resuspendido y se inocularon en 2mL de medio LB o medio mínimo M9, ambos adicionados con 25 mg/mL de kanamicina, se incubaron a 37 °C con agitación continua (250 rpm) y el crecimiento se monitoreó al seguir el cambio en la densidad óptica a 600 nm.

La actividad *in vitro* fue monitoreada siguiendo la producción de NADPH o NADH (Figura 2) a 340 nm (uv-Vis). La medición se llevó a cabo en un espectro Beckman Coulter DU730 con una celda de cuarzo de 10 mm de paso de luz y 8.5 mm de ancho (Hellma; Müllheim, Alemania). Las reacciones se llevaron a cabo por triplicado en un volumen final de 100 μ L (Buffer Tris-HCl 100 mM, pH 9.0) con diferentes concentraciones de NADP⁺ o NAD⁺ (asumiendo un coeficiente de extinción de 6220 $M^{-1} cm^{-1}$), para determinar la preferencia de la enzima por algún cofactor, en la presencia de ácido shikímico (SA). Todos los reactivos excepto la enzima fueron mezclados previamente e incubados a 25 °C durante 5 minutos. La enzima fue agregada al final para iniciar la reacción.

Los valores de K_M y V_{Max} fueron determinados ya sea al variar la cantidad de sustrato o la de cofactor mientras se mantenía una de ellas a niveles de saturación. Se consideraron 5 mM de sustrato o cofactor como la concentración saturante. Para la determinación del valor de

K_m para el ácido shikímico ($K_{M(SA)}$), NADP+ se fijó a 5 mM, y el ácido shikímico se varió desde 100 μ M hasta 3000 μ M. Para la determinación del K_M de NADP+ ($K_{M(NADP)}$), el ácido shikímico se fijó a 5 mM y NADP+ varió desde 100 μ M hasta 4000 μ M. Finalmente para determinar el valor de K_M de NAD+ ($K_{M(NAD)}$), el ácido shikímico se fijó a 5 mM y NAD+ varió desde 100 μ M hasta 10000 μ M.

El análisis de datos se llevó a cabo usando *SigmaPlot*. Los datos obtenidos (bajo las condiciones de velocidad inicial, y al calcular la pendiente de la parte lineal de cada reacción) se ajustaron a la ecuación de Michaelis-Menten.



SECCIÓN
06



RESULTADOS Y DISCUSIÓN



*Diseño y Caracterización
de un Dominio
Rossmann Consenso*



CLAUDIA IRIS BRAVO BONILLA

RESULTADOS Y DISCUSIÓN

DETERMINACIÓN DE LA SECUENCIA CONSENSO

Debido a que en años recientes la cantidad de información de secuencias biológicas ha crecido de una forma muy rápida, el uso de herramientas bioinformáticas para poder analizar esta información se ha vuelto indispensable. Los programas BLAST (*Basic Local Alignment Search Tool*) en sus diferentes variantes son herramientas bioinformáticas ampliamente utilizadas que permiten la búsqueda de secuencias similares en bases de datos de ADN o proteínas (J. YE *et al.*, 2006), permitiéndonos identificar 4280 secuencias homólogas a *aroE*. Esta búsqueda se llevó a cabo en una base de datos no redundante de proteínas que excluía las muestras ambientales de proyectos de secuenciación de genomas completos (BANDA, 2013).

El programa CD-Hit puede manejar de forma eficiente bases de datos grandes. En este caso se utilizó para generar bases de datos de proteínas no redundantes para disminuir los esfuerzos en la búsqueda de datos y mejorar la sensibilidad de detección de homología (LI & GODZIK, 2006). Nos permitió asegurarnos que las secuencias utilizadas no se pareciesen entre sí más allá del valor establecido (80 %), de esa forma la sobre representación de enzimas se eliminó, debido a que un consenso es fácilmente influenciado por la manera en la cual las secuencias son muestreadas y puede ser fácilmente desviado a una secuencia particular si demasiadas secuencias similares son incluidas (DAI *et al.*, 2007). Al final, se recuperaron 1637 secuencias previamente alineadas por BLAST.

La región correspondiente al dominio Rossmann, el cual abarca del aminoácido 102 al 244 de la secuencia de *AroE*, fue extraída usando el paquete de análisis de secuencias de EMBOSS (*The European Molecular Biology Open Software Suite*) (RICE, LONGDEN, & BLEASBY, 2000).

Estos alineamientos correspondientes solo a la región Rossmann fueron importados a ClustalX (LARKIN *et al.*, 2007) para obtener un dendograma, el cual es una representación gráfica en forma de árbol que resume el proceso de agrupación en un análisis de grupos, los objetos similares se conectan mediante enlaces cuya posición en el diagrama está determinada por el nivel de similitud/disimilitud entre los objetos (VILLARDÓN, 2011). A partir de este dendograma se obtuvieron dos conjuntos de secuencias: el primero de ellos consiste en 47 secuencias (Consenso 1) cuyo nivel de similitud con *AroE* es mayor, el segundo conjunto (que engloba al primer conjunto) consiste en 275 secuencias (Consenso 2) cuyo nivel de similitud con *AroE* es alto.

No hay mucha información sobre el número de secuencias necesarias (que van desde 2 hasta más de 200) para determinar un consenso o a qué grupo deben de pertenecer exactamente las secuencias, por lo que se propuso realizar un nuevo conjunto de datos que abarcaran únicamente las secuencias pertenecientes a arqueas, para esto se utilizaron los indicadores de taxonomía de las secuencias filtradas por redundancia para posteriormente usar el buscador de taxonomía del *National Center for Biotechnology Information: Taxonomic Common Tree* («Taxonomic Common Tree», 2012). Esta búsqueda nos arrojó un total de 96 secuencias similares a AroE correspondientes a arqueas (Consenso 3) (Figura 7).

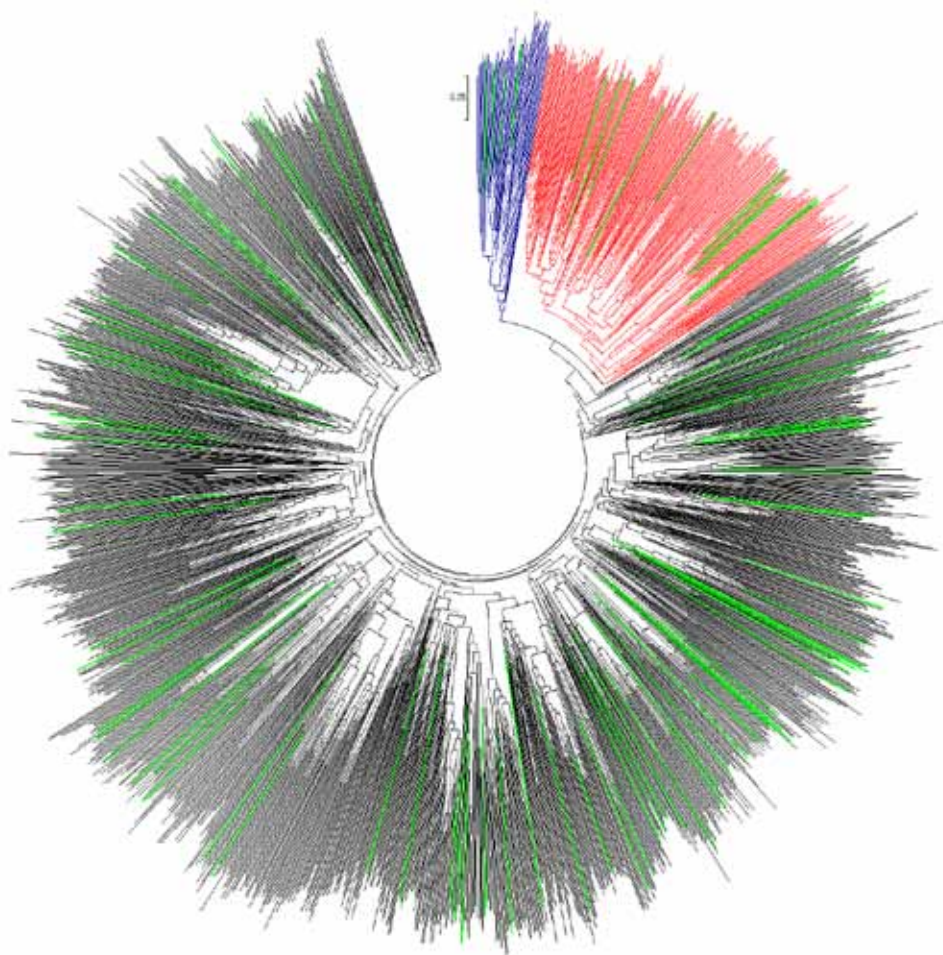


Figura 7: *Dendrograma correspondiente al dominio Rossmann de las 1637 secuencias.*

En azul se muestran las secuencias que componen el primer conjunto (Consenso 1) junto con las rojas componen el conjunto dos (Consenso 2), mientras que en verde se muestran las del tercer conjunto (Consenso 3). Imagen realizada con el programa MEGA 6.0 (TAMURA, STECHER, PETERSON, FILIPSKI, & KUMAR, 2013).

Las secuencias consenso se generaron con el programa multiplataforma *SeaView* (GOUY *et al.*, 2010), este programa facilita los alineamientos múltiples de secuencias moleculares de datos mediante el uso de una interfaz gráfica del usuario. Se estableció un punto de corte del 30 %, ya que se ha reportado que este porcentaje de identidad es un parámetro eficiente para obtener un consenso de mayoría simple (WANG *et al.*, 1999).

Para obtener una secuencia consenso del dominio Rossmann se realizó un alineamiento de secuencias homólogas a AroE. Aun cuando es difícil dar una razón convincente por la cual una secuencia debe ser omitida del alineamiento, se disminuyó la sobre representación de enzimas al filtrar por redundancia, ya que el proceso por el cual el alineamiento principal fue creado está inherentemente sesgado. Cada una de los consensos proviene de alineamientos con diferente número de secuencias y se ha demostrado que la diferencia en el tamaño de secuencias de cada uno de los consensos no tiene algún efecto significativo en la secuencia consenso final (SULLIVAN *et al.*, 2011). Finalmente, se obtuvieron tres secuencias consenso del dominio Rossmann.

DISEÑO Y SÍNTESIS DEL GEN CON LA SECUENCIA CONSENSO

Los aminoácidos determinados en las secuencias consenso se compararon con los datos de análisis de acoplamiento estadístico (SCA) realizados por Jesús Banda durante el 2013, este método se basa en el análisis de co-variación entre residuos de una proteína y por lo tanto la determinación del acoplamiento entre ellos debido a su evolución conjunta (LOCKLESS & RANGANATHAN, 1999). Esta información nos permitió determinar cuáles aminoácidos forman una red de co-evolución que no debe ser interrumpida para así poder mantener características importantes como la actividad, especificidad, plegamiento, estabilidad o solubilidad (STRAFFORD *et al.*, 2012), ya que anteriormente se ha observado que combinar mutaciones consenso individuales solo funciona cuando las correlaciones entre las posiciones son tomadas en cuenta (SULLIVAN *et al.*, 2011).

El análisis de acoplamiento estadístico previamente realizado a ambos dominios de la enzima *shikimate* deshidrogenasa de AroE de *Escherichia coli* (BANDA, 2013), determinó que los residuos muy conservados correspondientes al dominio Rossmann se encontraban en las posiciones: 102, 212, 215, 220, 237, 240, 244. Y al mismo tiempo se determinó usar como secuencia guía la secuencia del dominio Rossmann de la enzima *shikimate* deshidrogenasa de *Escherichia coli* (AroE).

En la secuencia del Consenso 1 se determinaron 111 aminoácidos, entre ellos se encuentran los residuos determinados por SCA. De los aminoácidos consenso sólo 15 son diferentes a la secuencia guía de AroE (ninguno de ellos pertenece a los residuos determinados por SCA), la mayoría de las sustituciones entre AroE y Consenso1 son sustituciones no conservadoras (Leu120Asp, Ser131Ala, Leu141Tyr, Asp142Gly, Val152Phe, Glu156Gln, Ile208Thr, Lys219Leu, Ser231Ala, Ala243Gln), con cinco excepciones (Val104Ile, Arg117Lys, Leu135Ile, Val145Ile, Asp195Glu).

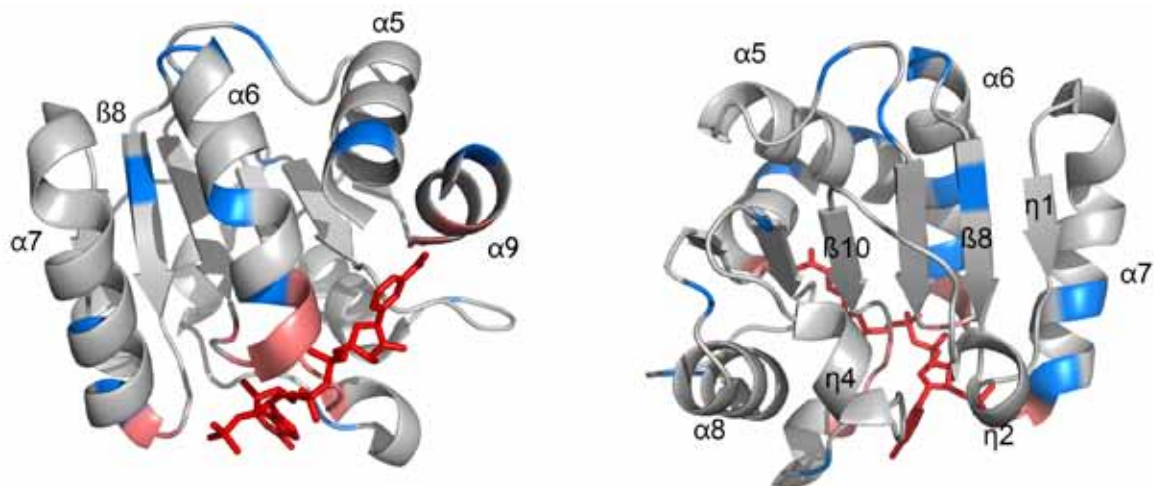


Figura 8. Representación tridimensional de la estructura secundaria del dominio Rossmann de Con1. En azul se muestran la posición de los aminoácidos mutados, en rojo claro se muestran los aminoácidos que se encuentran en contacto con el cofactor (estructura de palitos roja). Imagen realizada con Pymol 1.5.

Las sustituciones de los aminoácidos se encuentran localizadas en regiones particulares del dominio: siete de ellas se encuentran en α hélices (Val104Ile, Ser131Ala, Leu135Ile, Leu141Tyr, Val152Phe, Glu156Gln, Ala243Gln), una en hoja β (Val145Ile) y siete en giros y asas (Arg117Lys, Leu120Asp, Asp142Gly, Asp195Glu, Ile208Thr, Lys219Leu, Ser231Ala).

Las mutaciones en las hélices α se encuentran dispersas: una en la hélice $\alpha 5$ (Val104Ile), tres en la $\alpha 6$ (Ser131Ala, Leu135Ile, Leu141Tyr), dos en la $\alpha 7$ (Val152Phe, Glu156Gln) y una en la $\alpha 9$ (Ala243Gln). En la hoja $\beta 8$ se encuentra la única sustitución (Val145Ile) correspondiente a este tipo de estructura secundaria.

En la secuencia del Consenso 2 se determinaron 108 aminoácidos, entre ellos se encuentran los residuos determinados por SCA. De los aminoácidos consenso sólo 32 son diferentes a la secuencia guía de AroE (ninguno de ellos pertenece a los residuos determinados por SCA), la mayoría de las sustituciones entre AroE y Consenso 2 son sustituciones no

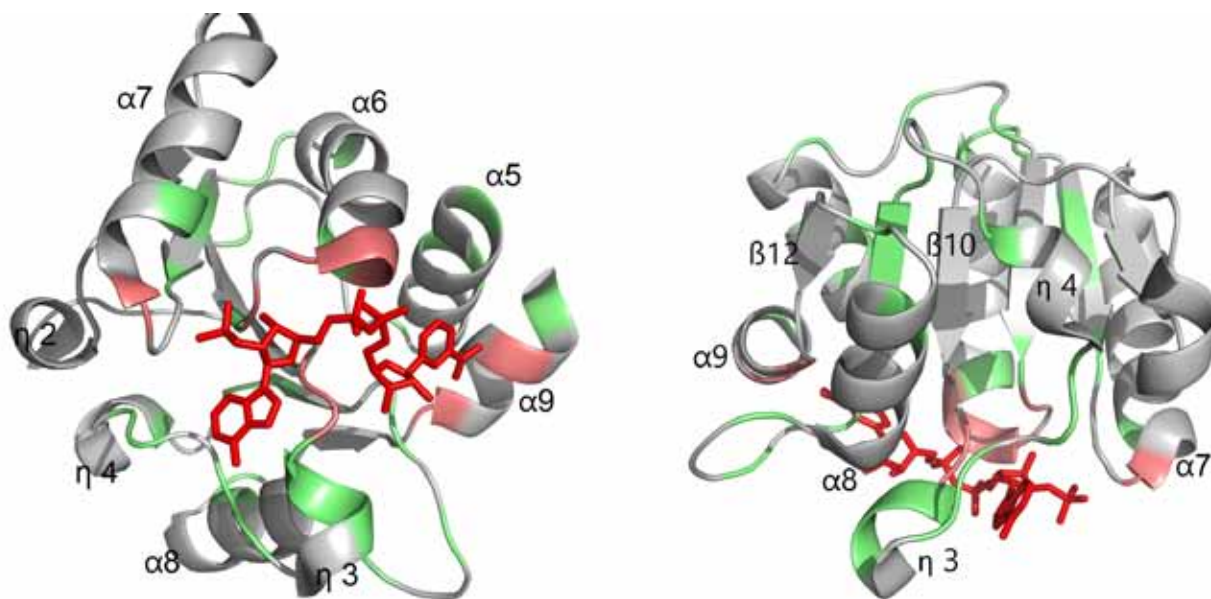


Figura 9. Representación tridimensional de la estructura secundaria del dominio Rossmann de Con2. En verde se muestran la posición de los aminoácidos mutados, en rojo claro se muestran los aminoácidos que se encuentran en contacto con el cofactor (estructura de palitos roja). Imagen realizada con Pymol 1.5

conservadoras (Ser108Arg, Ser114Gly, Arg117Leu, Leu120Lys, Ser131Ala, Asp142Gly, Cys143Pro, Thr146Val, Thr148Ala, Ile204Phe, Ile208Thr, Tyr209Leu, Cys210Ala, Phe214Met, Gln216Gly, Lys219Pro, Ala243Glu), con quince excepciones (Val104Ile, Leu107Val, Ile125Leu, Leu141Ala, Val145Leu, Arg154Lys, Ser190Ala, Gly191Ser, Ile192Leu, Asp195Glu, Ile196Leu, Ala198Pro, Ile199Leu, Cys226Ala, Ser231Ala).

Más de la mitad de las sustituciones de los aminoácidos se encuentran localizadas en las regiones correspondientes a los giros y asas (Ser114Gly, Arg117Leu, Leu120Lys, Asp142Gly, Cys143Pro, Ser190Ala, Gly191Ser, Ile192Leu, Asp195Glu, Ile196Leu, Ala198Pro, Ile199Leu, Ile204Phe, Ile208Thr, Phe214Met, Gln216Gly, Lys219Pro, Ser231Ala), mientras que ocho sustituciones se encuentran en α hélices (Val104Ile, Leu107Val, Ser108Arg, Ser131Ala, Leu141Ala, Arg154Lys, Cys226Ala, Ala243Glu) y solamente seis en hojas β (Ile125Leu, Val145Leu, Thr146Val, Thr148Ala, Tyr209Leu, Cys210Ala).

Las mutaciones en las hélices α se encuentran dispersas: tres en la hélice $\alpha 5$ (Val104Ile, Leu107Val, Ser108Arg), dos en $\alpha 6$ (Ser131Ala, Leu141Ala) y una en $\alpha 7$ (Arg154Lys), $\alpha 8$ (Cys226Ala) y $\alpha 9$ (Ala243Glu). Las mutaciones en las hojas beta corresponden a la hoja $\beta 7$ (Ile125Leu), $\beta 8$ (Val145Leu, Thr146Val, Thr148Ala) y $\beta 11$ (Tyr209Leu, Cys210Ala).

En la secuencia del Consenso 3 se determinaron 79 aminoácidos, entre ellos se encuentran los residuos determinados por SCA. De los aminoácidos consenso sólo 17 son diferentes a la secuencia guía de AroE (ninguno de ellos pertenece a los residuos determinados por SCA), la mayoría de las sustituciones entre AroE y Con3 son conservadoras (Ile122Val, Leu124Ile, Ile125Leu, Gly133Ala, Leu135Ala, Pro137Ala, Ile184Leu, Ile192Met, Ile204Leu, Asn234Thr), con siete excepciones (Leu106Phe, Ser131Ala, Asp142Gly, Cys210Val, Phe214Val, Lys217Pro, Cys226Ala).

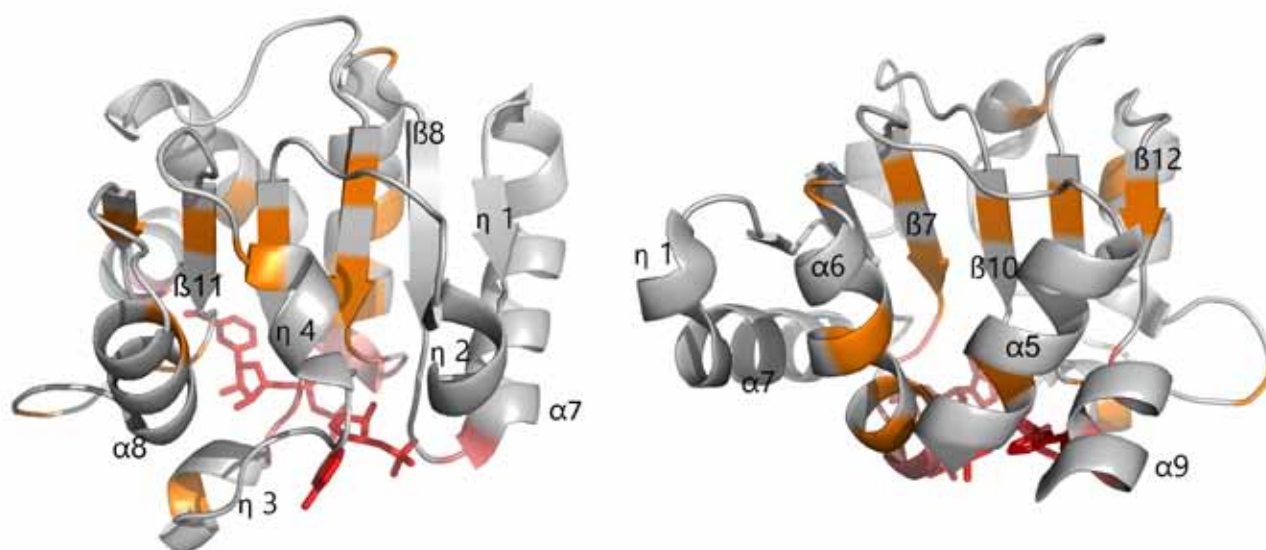


Figura 10. Representación tridimensional de la estructura secundaria del dominio Rossmann de Con3. En naranja se muestran la posición de los aminoácidos mutados, en rojo claro se muestran los aminoácidos que se encuentran en contacto con el cofactor (estructura de palitos roja). Imagen realizada con Pymol 1.5

Las mutaciones se encuentran dispersas a lo largo del dominio: seis corresponden a hélices α (Leu106Phe, Ser131Ala, Gly133Ala, Leu135Ala, Pro137Ala, Cys226Ala), seis a hojas β (Ile122Val, Leu124Ile, Ile125Leu, Ile184Leu, Cys210Val, Asn234Thr) y cinco a giros y loops (Asp142Gly, Ile192Met, Ile204Leu, Phe214Val, Lys217Pro).

Las mutaciones en las hélices α se encuentran principalmente en la hélice $\alpha 6$ (Ser131Ala, Gly133Ala, Leu135Ala, Pro137Ala), mientras que las hélices $\alpha 5$ y $\alpha 8$ sólo presentan una mutación cada una (Leu106Phe y Cys226Ala respectivamente). La mitad de las mutaciones en hojas β se localizan en $\beta 7$ (Ile122Val, Leu124Ile, Ile125Leu), mientras que la otra mitad está dispersa en las siguientes hojas: $\beta 10$ (Ile184Leu), $\beta 11$ (Cys210Val), $\beta 12$ (Asn234Thr).

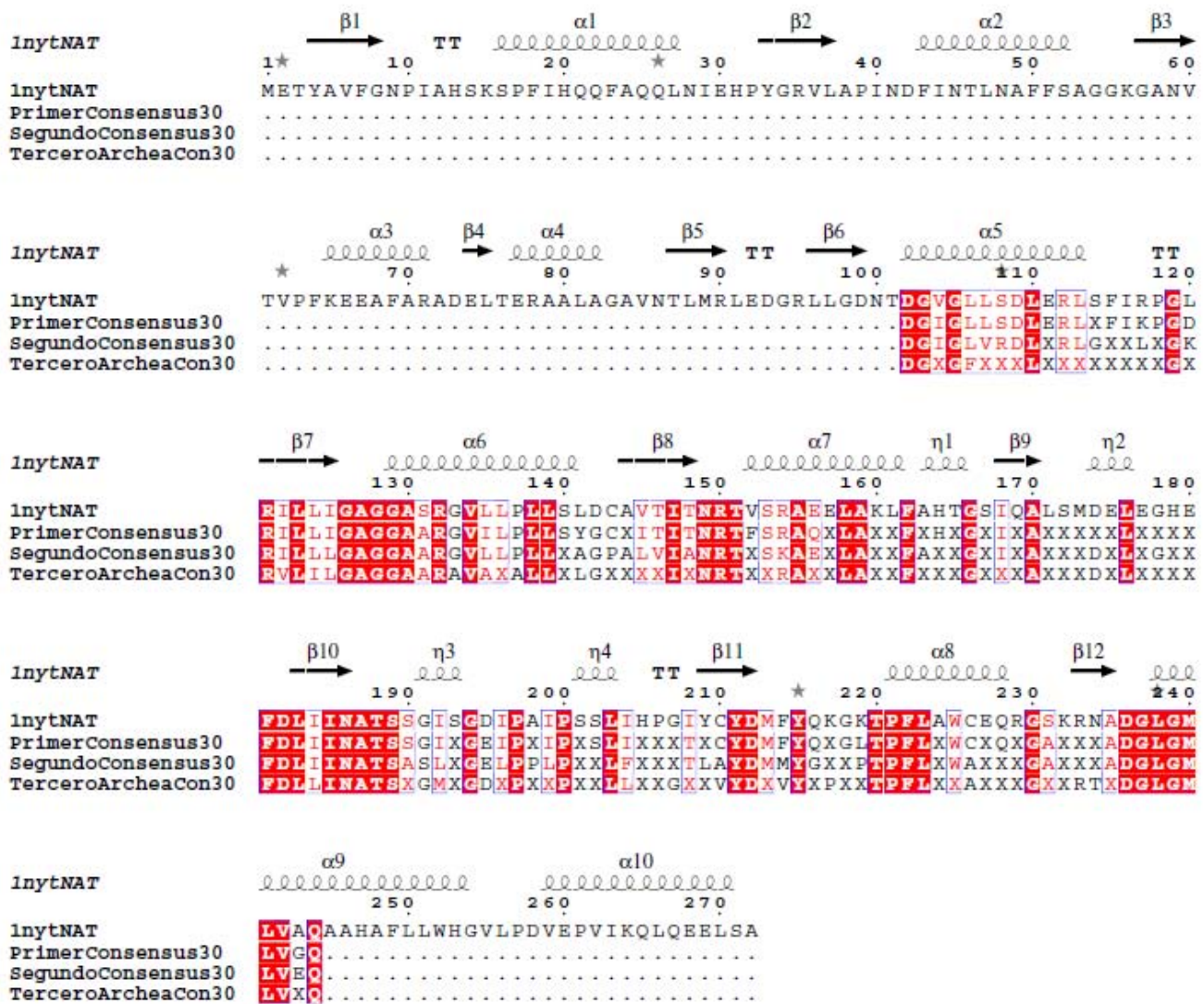


Figura 11: Alineamiento de las secuencias de AroE y los diferentes consensos. La estructura secundaria se muestra en la parte superior. En los cuadros rojos se muestran los aminoácidos que son los mismos en todas las secuencias. En letras rojas se muestran los aminoácidos que son similares en todas las secuencias. En letras negras se muestran los aminoácidos que no presentan similitud en las secuencias. La X representa todos los aminoácidos que no fueron determinados en el consenso. El modelo fue generado con ESPrpt (ROBERT & GOUET, 2014)

Posteriormente a las tres secuencias consenso de los dominios Rossmann se les anexó el dominio catalítico de AroE y una etiqueta de seis histidinas para permitir su posterior purificación. Estas secuencias de aminoácidos fueron retro-traducidas usando el programa PRIMO de *Chang Bioscience* el cual es un programa diseñado para optimizar la expresión de genes de interés en un organismo huésped (en este caso, *E. coli*) disminuyendo el uso de codones raros, ya que la presencia de éstos puede resultar en un estancamiento durante la traducción y en una expresión deficiente (nula expresión, expresión en niveles bajos o como

proteínas truncadas) (CRUZ-VERA, MAGOS-CASTRO, ZAMORA-ROMO, & GUARNEROS, 2004), de la misma forma tampoco resulta conveniente utilizar únicamente los codones de máxima frecuencia, debido al posible secuestro de los aminoacil-tRNAs óptimos que podría afectar el crecimiento de la bacteria (MORÁN GARCÍA, 2008).

Los productos de PCR se ligaron de acuerdo al protocolo anteriormente descrito, el cual consiste en estimar la cantidad de vector e inserto a ligar basándose en la intensidad de las bandas de ADN en un gel de agarosa teñido con bromuro de etidio (GRATCHEV, n.d.), al usar la siguiente fórmula:

$$Li=3 *Lv*\left(\frac{Si}{Sv}\right)*\left(\frac{Iv}{Ii}\right)*\left(\frac{Vi}{Vv}\right)$$

Donde Sv es el tamaño del vector (2044 pb), Si es el tamaño del inserto (834 pb), Iv es la intensidad relativa de la banda correspondiente al vector (con un valor de uno, segundo carril de la Figura 12), Ii es la intensidad relativa de la banda correspondiente a los insertos (con un valor de cinco para Con1 y Con2, y de cuatro para Con3; carril 3, 4 y 5 respectivamente de la Figura 12), Vv y Vi son los volúmenes de el vector y el inserto que se cargaron en el gel, Lv es el volumen del vector que queremos utilizar, y Li es el volumen de inserto que se debe usar para cada reacción.

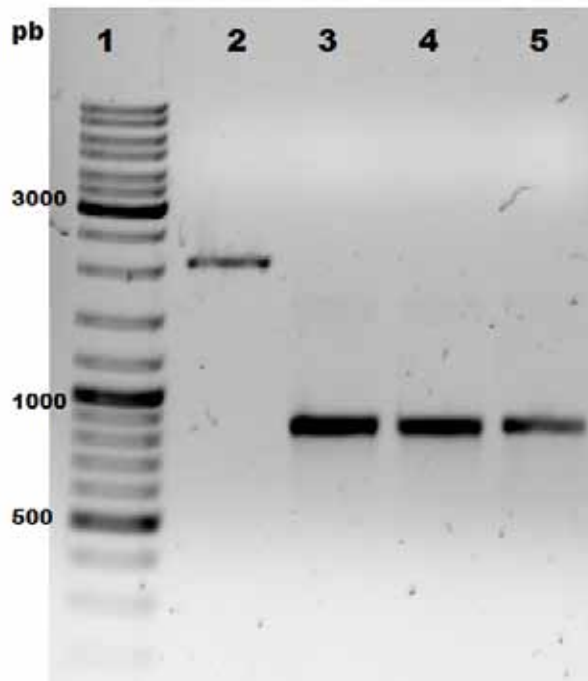


Figura 12: Ligación. Carril 1: marcador de peso molecular. Carril 2: vector EBE.His digerido y purificado. Carril 3: inserto Con1 purificado. Carril 4: inserto Con2 purificado. Carril 5: inserto Con3 purificado.

Las construcciones finales fueron verificadas mediante secuenciación y patrón de restricción. La congruencia de las secuencias con los sitios de restricción usados fue de 100 %, a excepción de la secuencia de Consenso 1 en la cual la etiqueta de histidinas consta de cuatro histidinas. Debido a que con una etiqueta de cuatro histidinas se puede retener más del 90 % de la proteína en una columna de Ni²⁺-NTA (TERPE, 2003) se decidió dejarla así.

Las construcciones fueron clonadas en diferentes cepas quimio-competentes de *E. coli*: XL10-Gold (genotipo: *T et* Δ (*mcrA*)183 Δ (*mcrCB-hsdSMR-mrr*)173 *endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte* [*F' proAB lacI^qZ Δ M15 Tn10(Tet^r) Amy Cam^r*]; JM101 (genotipo: *supE thi-1* Δ (*lac-proAB*) [*F' traD36 proAB lacI^qZ Δ M15*] (CGSC, 2014)) y BW25113 Δ AroE km- (genotipo: *F⁻*, Δ (*araD-araB*)567, Δ *lacZ4787(::rrnB-3)*, λ -, *rph-1*, Δ (*rhaD-rhaB*)568, *hsdR514*, colección Keio (BABA *et al.*, 2006), resistencia eliminada según Datsenko (DATSENKO & WANNER, 2000) y donada por Adelfo Escalante (IBT, UNAM, 2009).

Cada uno de los tres consensos del dominio Rossmann obtenidos presenta diferente número de mutaciones: el primer consenso presenta 15 mutaciones, el segundo 32 mutaciones y el tercero presenta un total de 17 mutaciones. Trabajos anteriores sugieren que la red de co-evolución se mantiene intacta solamente si se realiza diseño de consenso a toda la secuencia (AERTS *et al.*, 2013). En este trabajo hemos visto que el realizar el diseño de consenso a una parte de la enzima, mantiene la red de co-evolución intacta.

EXPRESIÓN

Para la producción de enzima, se usó la cepa de *Escherichia coli* JM101 con las siguientes construcciones: el vector pT4 vacío como control negativo, el vector pT4 con el gen AroE (con etiqueta de histidinas) como control positivo y el vector pT4 con cada una de los genes consenso.

Una vez obtenida las fracciones solubles de cada uno de los cultivos, se realizó un gel SDS-PAGE para determinar si las proteínas consenso se expresaban. Usando la herramienta ExPASy (GASTEIGER, 2003) se calcularon los pesos de cada una de las proteínas: AroE 30.64 kDa, Con1 30.15 kDa, Con2 29.89 kDa y Con3 29.91 kDa. Por lo tanto se espera observar expresión de estas proteínas entre las bandas del marcador correspondientes a 25 y 35 kDa.

En la Figura 13 se puede observar la expresión de AroE (carril 3) y Con1 (carril 4) a 37 °C, pero no se observó expresión de Con2 (carril 5) ni Con3 (carril 6). Se realizó un gel SDS-PAGE de la fracción insoluble y tampoco se observó expresión a 37 °C de Con2 o Con3.

Debido a que Con2 y Con3 no se expresan a 37 °C, se decidió probar con otras dos

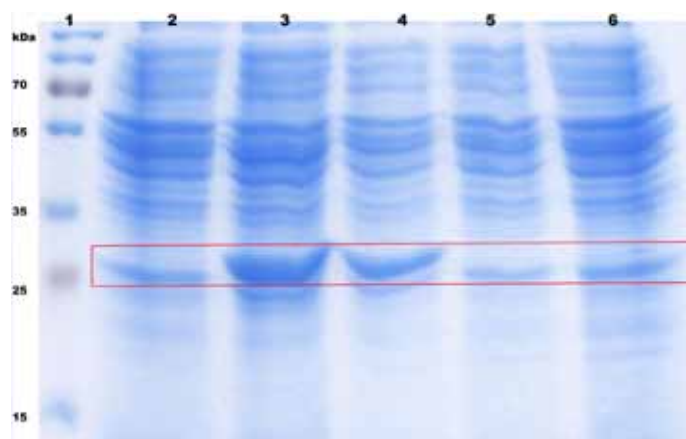


Figura 13: Gel SDS-PAGE de la fracción soluble de diferentes cultivos a 37 °C. Carril 1: marcador de peso molecular. Carril 2: pT4 vacío como control negativo. Carril 3: AroE como control positivo. Carril 4: Con1. Carril 5: Con2. Carril 6: Con 3. En rojo se muestra la altura a la cual se espera observar la expresión de las proteínas consenso.

temperaturas de cultivo: 25 y 30 °C. *Escherichia coli* es una bacteria mesofílica capaz de crecer bien en el rango de temperatura desde 21 hasta 49 °C, con un óptimo a 37 °C; de forma rutinaria se usan las temperaturas sub-óptimas entre 15 y 30 °C para la producción de proteínas recombinantes ya que la expresión de proteínas a bajas temperaturas previene la formación de cuerpos de inclusión y la tasa de traducción disminuye lo cual le da tiempo suficiente al polipéptido naciente para plegarse de forma correcta (SONG, AN, KANG, LEE, & CHA, 2012).

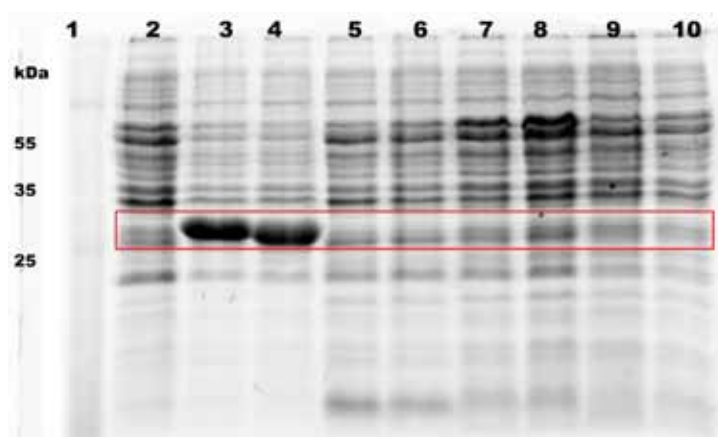


Figura 14: Gel SDS-PAGE de cultivos a diferentes temperaturas. Carril 1: marcador de peso molecular. Carril 2: pT4 vacío a 37 °C. Carril 3: AroE a 37 °C. Carril 4: Con1 a 37 °C. Carril 5: Con2 a 37 °C. Carril 6: Con3 a 37 °C. Carril 7: Con2 a 30 °C. Carril 8: Con3 a 30 °C. Carril 9: Con2 a 25 °C. Carril 10: Con3 a 25 °C. En rojo se muestra la altura a la cual se espera observar la expresión de las proteínas consenso.

El cultivo se centrifugó y el pellet se cargó en un gel SDS-PAGE. En la Figura 14 podemos observar la expresión a diferentes temperaturas: el carril 2 corresponde a el cultivo a 37 °C de pT4 vacío (control negativo), el carril 3 corresponde a el cultivo a 37 °C de AroE (control positivo), en el carril 4 se observa la expresión de Con1 a 37 °C, los cultivos correspondientes a Con2 se observan en el carril 5 (temperatura de cultivo 37 °C), 7 (temperatura de cultivo 30 °C) y 9 (temperatura de cultivo 25 °C), finalmente los cultivos correspondientes a Con3 se observan en el carril 6 (temperatura de cultivo 37 °C), 8 (temperatura de cultivo 30 °C) y 10 (temperatura de cultivo 25 °C).

Aun a diferentes temperaturas de cultivo, no se observan diferencias en la expresión de AroE y Con1. Tampoco se observa expresión de Con2 y Con3 en ninguna de las tres temperaturas de cultivo.

Se realizó un western blot al gel anterior (Figura 15). Debido a que el anticuerpo es específico para una etiqueta de seis histidinas por eso no se observa la señal de Con1 (la secuenciación de pT4-Con1 mostró que presenta un etiqueta de cuatro histidinas). La secuenciación del plásmido pT4-Con2 y pT4-Con3 muestra que ambos presentan una etiqueta de seis histidinas, pero no se observa señal alguna en el western blot, por lo tanto, Con2 y Con3 no pudieron expresarse.

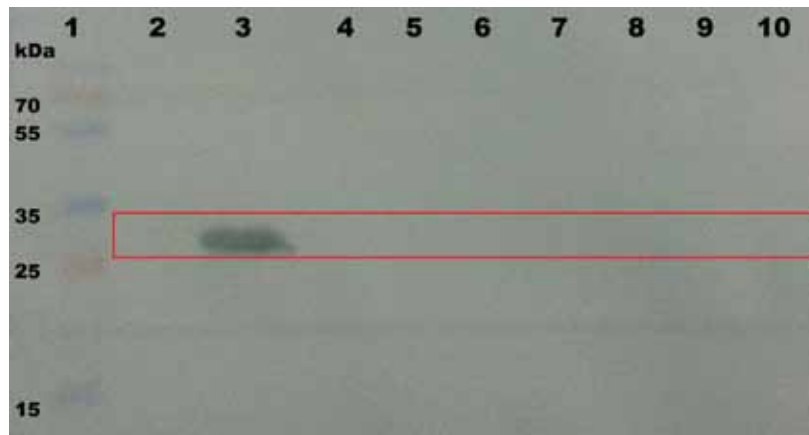


Figura 15: Western Blot del gel mostrado en la figura anterior. Carril 1: marcador de peso molecular. Carril 2: pT4 vacío a 37 °C. Carril 3: AroE a 37 °C. Carril 4: Con1 a 37 °C. Carril 5: Con2 a 37 °C. Carril 6: Con3 a 37 °C. Carril 7: Con2 a 30 °C. Carril 8: Con3 a 30 °C. Carril 9: Con2 a 25 °C. Carril 10: Con3 a 25 °C. En rojo se muestra la altura a la cual se espera observar la expresión de las proteínas consenso.

La proteína consenso Con1, la cual presenta 15 mutaciones respecto a AroE, se encuentra en la fracción soluble. Aun con diferentes temperaturas de cultivo no se observan diferencias en la expresión de Con1. La expresión de Con1 muestra que ninguno de los aminoácidos consensos modificó la solubilidad.

No fue posible expresar ninguna de las otras dos proteínas consenso (Con2 y Con3) en las diferentes temperaturas de cultivo probadas, aun cuando las tres secuencias consenso presentan un 90 % de identidad en secuencia. Ambas secuencias consenso provienen de alineamientos diferentes y con mayor número de secuencias (275 secuencias para Con2 y 96 secuencias para Con3) que el correspondiente a Con1 (47 secuencias).

La falta de expresión de las proteínas consenso Con2 y Con3 podría deberse al uso de una secuencia guía inapropiada, debido a que estos consensos se originaron a partir de bases de datos de diferente origen al de Con1: Con2 abarca un mayor número de secuencias de gamma-proteobacterias mientras que Con3 consiste en secuencias de arqueas. Aunque no se debe descartar el hecho de que los aminoácidos consenso presentes en Con2 y Con3 desestabilizan de alguna forma el plegamiento de la proteína, ocasionando una degradación de estos productos. A partir de esto podemos afirmar que el mantener la red de co-evolución y usar una secuencia guía no garantiza la obtención de un consenso exitoso en todos los casos.

PURIFICACIÓN

Una vez confirmada la presencia de la proteína Con1 en la fracción soluble, ésta se cargó en una columna de níquel y se eluyó mediante un gradiente escalonado de imidazol. Se realizó un gel SDS-PAGE de cada una de las fracciones recuperadas para determinar donde se encontraba

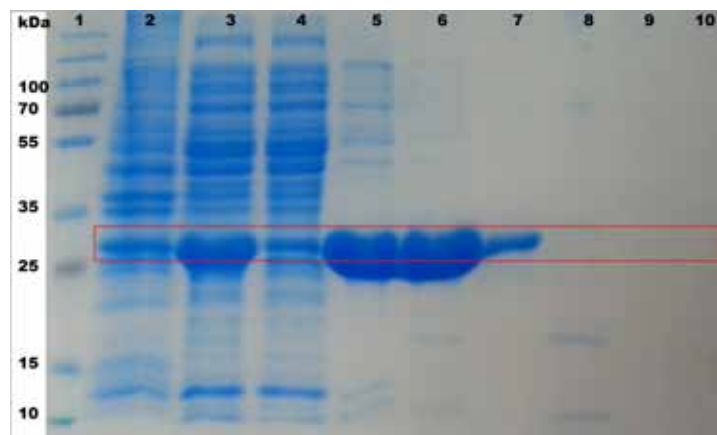


Figura 16: *Gel SDS-PAGE de la purificación de Con1.* Carril 1: marcador de peso molecular. Carril 2: fracción insoluble. Carril 3: fracción soluble después de entrar a la columna. Carril 4: fracción correspondiente a 20 mM de imidazol. Carril 5: fracción correspondiente a 35 mM de imidazol. Carril 6: fracción correspondiente a 75 mM de imidazol. Carril 7: fracción correspondiente a 100 mM de imidazol. Carril 8: fracción correspondiente a 150 mM de imidazol. Carril 9: fracción correspondiente a 200 mM de imidazol. Carril 10: fracción correspondiente a 250 mM de imidazol. En rojo se muestra la altura a la cual se espera observar la expresión de las proteínas consenso.

Con1. Se decidió utilizar la fracción correspondiente a 100 mM de imidazol porque fue la que presentó menor número de proteínas contaminantes (Figura 16). Esta fracción se concentró, y dializó de forma exhaustiva en buffer desgasificado para disminuir la turbidez de las muestras, y poder realizar su posterior caracterización.

La concentración y pureza de la proteína fueron juzgadas mediante SDS-PAGE (Figura 17) y Bradford (Bio-Rad; California, USA). Las muestras presentan un 93-95 % de pureza de acuerdo al análisis realizado con el programa *Image Lab 3.0*. Al final de la purificación se obtienen 65-70 mg de proteína Con1 por cada litro de cultivo.

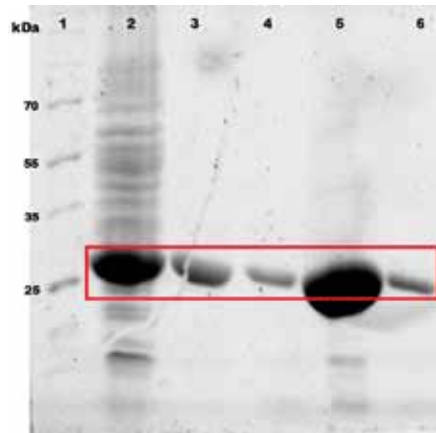


Figura 17: Purificación de Con1. Carril 1: marcador de peso molecular. Carril 2: fracción soluble. Carril 3: fracción correspondiente a 100 mM de imidazol. Carril 4: muestra concentrada usando el Vivaspin. Carril 5: primera diálisis. Carril 6: muestra purificada después de la segunda diálisis y filtración. En rojo se muestra la altura a la cual se espera observar la expresión de las proteínas consenso.

Se purificó AroE como control realizando la misma metodología descrita para Con1. La proteína AroE (también) se encuentra en la fracción soluble y se purificó a partir de la fracción correspondiente a 100 mM de imidazol (Figura 18).

La concentración y pureza de la proteína AroE fueron estimadas mediante SDS-PAGE (Figura 19) y Bradford (Bio-Rad; California, USA). Las muestras presentan un 93-95 % de pureza de acuerdo al análisis realizado con el programa *Image Lab 3.0*. Al final de la purificación se obtienen 20-25 mg de proteína AroE por cada litro de cultivo.

La proteína consenso Con1 se expresó de mejor forma que la proteína parental (AroE). Al final de la purificación se obtienen 65-70 mg de proteína Con1 por cada litro de cultivo, lo cual es un poco más del doble de lo que se obtiene al purificar AroE (20-25 mg).

El proceso de purificación aquí descrito y utilizado, nos permitió purificar proteína con un porcentaje de hasta 95 % de pureza, sin la necesidad de incrementar los pasos en el tren de purificación.

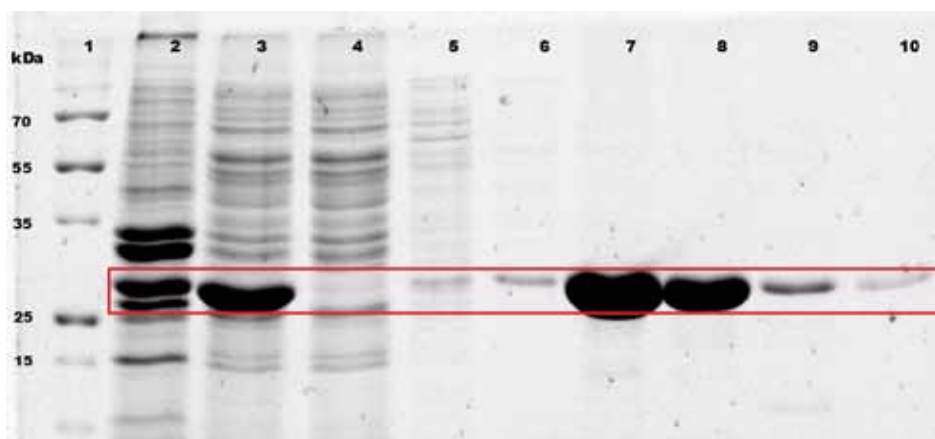


Figura 18: Purificación de AroE. Carril 1: marcador de peso molecular. Carril 2: fracción insoluble. Carril 3: fracción soluble después de entrar a la columna. Carril 4: fracción correspondiente a 20 mM de imidazol. Carril 5: fracción correspondiente a 35 mM de imidazol. Carril 6: fracción correspondiente a 75 mM de imidazol. Carril 7: fracción correspondiente a 100 mM de imidazol. Carril 8: fracción correspondiente a 150 mM de imidazol. Carril 9: fracción correspondiente a 200 mM de imidazol. Carril 10: fracción correspondiente a 250 mM de imidazol. En rojo se muestra la altura a la cual se espera observar la expresión de las proteínas consenso.

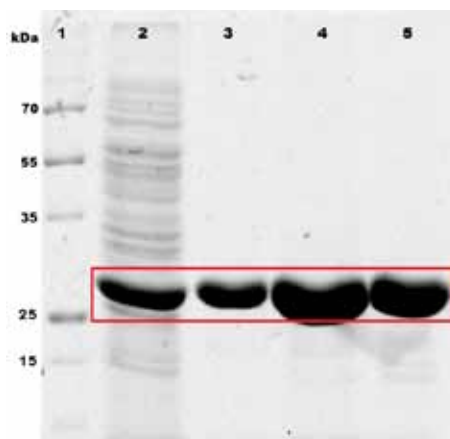


Figura 19: Purificación de AroE. Carril 1: marcador de peso molecular. Carril 2: fracción soluble. Carril 3: fracción correspondiente a 100 mM de imidazol. Carril 4: muestra concentrada usando el Vivaspin. Carril 5: muestra purificada después de la segunda diálisis y filtración. En rojo se muestra la altura a la cual se espera observar la expresión de las proteínas consenso.

CARACTERIZACIÓN

Una vez purificada Con1, se procedió a caracterizarla mediante: difracción circular (CD), y calorimetría.

Mediante uv-lejano se determinó la estructura secundaria de AroE y Con1. La señal de CD se convirtió a MRE (*Mean Residue Ellipticity*, cuyas unidades son $\text{deg cm}^2 \text{dmol}^{-1}$) para

normalizar los datos de diferentes concentraciones de proteínas y tipos (AroE y Con1), y poder ser comparados, usando la siguiente fórmula:

$$[\theta]_{MRE} = \frac{\theta * MRW}{10 * l * c}$$

Donde θ es el valor obtenido en la lectura (mdeg), MRW es el cociente del peso molecular de la proteína (en daltones) entre el número total de aminoácidos que la componen menos uno, l es ancho de la celda en cm y c es la concentración de la muestra en mg/mL.

El dicroísmo circular se observa cuando las moléculas absorben la luz circular polarizada izquierda y derecha en diferentes grados. El cromóforo amida del enlace peptídico en las proteínas domina el espectro de CD abajo de 250 nm. Los espectros de AroE y Con1 muestran dos mínimos en 208 nm y 222 nm, correspondientes a una proteína con contenido alfa-helicoidal (Figura 20). El espectro de la proteína Con1 presenta un mayor contenido de estructura secundaria respecto al espectro de AroE.

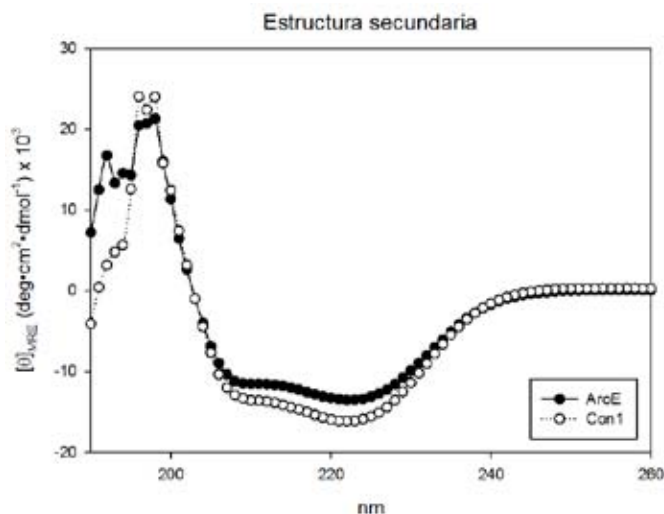


Figura 20: Estructura secundaria. Con1 y AroE muestran un alto contenido de hélices α . Las elipticidades molares residuales (θ) de ambas fueron medidas desde 190-260 nm. Los porcentajes de estructura secundaria de Con1 fueron determinados usando el algoritmo de deconvolución CDsstr.

Para estimar el porcentaje de estructura secundaria de Con1 se utilizó el servidor DICHROWEB (WHITMORE & WALLACE, 2004) y el algoritmo CDsstr (con el set 4 de datos), el cual permite predecir la estructura secundaria de la proteína en cuestión, al realizar una comparación con los datos de un conjunto de proteínas cuyo contenido de estructura secundaria es conocida (SREERAMA & WOODY, 2000). CDsstr usa el método de SVD (*Singular*

Value Decomposition) el cual ha demostrado dar la mejor predicción para las alfa hélices aunque el análisis para las otras estructuras es variable (W. C. JOHNSON, 1999).

Los resultados obtenidos de la deconvolución del espectro de Con1 indican que este presenta un 51 % en estructuras helicoidales y un 23 % en hojas β , con un NRMSD de 0.006. Mientras que los resultados de AroE muestran un 49 % de estructura helicoidal y un 21 % en hojas β (NRMSD: 0.014). Finalmente, la estructura cristalográfica de AroE (código PDB: 1nyt) depositada en el PDB (*Protein Data Bank*) posee un contenido de estructura secundaria de: 44 % helicoidal (13 helices, 121 residuos), 18 % en hojas β (12 hojas, 49 residuos) (Frishman & Argos, 1995).

Al comparar los datos de AroE obtenidos mediante DC con los reportados por el PDB, notamos que existe una diferencia de 5 % en la estructura helicoidal y 3 % en hebras beta, valores que entran dentro del margen de error (10 %) de la técnica. Cuando se comparan los datos obtenidos de la deconvolución de los espectros de AroE y Con1 podemos observar que Con1 presenta un contenido de estructura secundaria superior al de AroE en un 2 %, lo cual indica que las mutaciones introducidas mediante el consenso modificaron de forma ligera la estructura secundaria.

Se realizó una curva de desnaturalización de Con1 con urea, estas curvas son usadas para obtener un estimado de la estabilidad conformacional de las proteínas al medir las diferencias de energía libre entre los estados nativo (plegado) y desnaturalizado (desplegado) (MONERA, KAY, & HODGES, 1994). El perfil de desnaturalización se monitoreó al seguir el cambio en la elipticidad de Con1 a 222 nm.

En la Figura 21 se muestran los valores obtenidos a 222 nm como una función de la concentración de urea. A bajas concentraciones de urea (desde 0 hasta 3.2 M) podemos observar que la señal obtenida es constante y se consideró como la señal del estado plegado (región de pre-transición). La región de transición se observa entre 3.4 y 4.6 M de urea. A altas concentraciones de urea (desde 4.8 hasta 9 M) la señal observada presenta una pendiente constante y se consideró como la señal del estado desplegado (región de post-transición).

Se graficó la fracción aparente del estado desnaturalizado contra la concentración de urea, donde 0 indica que la muestra está en su forma plegada y 1 que la muestra está en su estado desplegado (Figura 21). Para calcular la fracción aparente (F_{app}) se usó la siguiente fórmula:

$$F_{app} = \frac{\theta - [(\theta_T * m_N) + b_N]}{[(\theta_T * m_U) + b_U] - [(\theta_T * m_N) + b_N]}$$

Donde θ es el valor de DC en mdeg, θ_T es el valor que le corresponde al estado nativo a cada condición de temperatura o urea dada, m_N es la pendiente de la señal del estado nativo,

b_N es la ordenada al origen de la señal del estado nativo, m_U es la pendiente de la señal del estado desplegado y b_U es la ordenada al origen de la señal del estado desplegado.

Los datos de desnaturalización química se ajustaron a un modelo de transición de dos estados (CHEN, BAASE, NICHOLSON, & SCHELLMAN, 1992) usando el software *Savuka* modificado por el Dr. Osman Bilsel (OSMAN BILSEL, s. f.). Para un proceso de dos estados con desplegamiento cooperativo si se define a X_N como la fracción de proteína plegada y a X_U como la fracción desnaturalizada:

$$X_N + X_U = 1$$

donde Y es el parámetro de medición obtenido:

$$Y = (Y_N * X_N) + (Y_U * X_U)$$

se define a la fracción de proteína desnaturalizada como:

$$X_U = \frac{Y_N - Y}{Y_N - Y_U}$$

siendo la constante de equilibrio, k :

$$k = \frac{X_U}{1 - X_U} = \frac{X_U}{X_N} = \frac{Y_N - Y}{Y - Y_U}$$

el cambio de energía libre es:

$$\Delta G = -RT \ln(k) = -RT \ln\left(\frac{Y_N - Y}{Y - Y_U}\right)$$

Para determinar la estabilidad conformacional en ausencia del desnaturalizante (ΔG en agua), el método más simple consiste en asumir una correlación lineal continua hasta cero y usar una regresión cuadrada que siga la ecuación:

$$\Delta G = \Delta G_{\text{agua}} - m[\text{desnaturalizante}]$$

Los datos obtenidos muestran que a una concentración de 3.8 M de urea, la mitad de la estructura de Con1 se encuentra desplegada y la otra mitad se encuentra en el estado nativo; la estabilidad conformacional de Con1 en la ausencia de urea es de 7.83 ± 0.75 kcal/mol, y la dependencia del ΔG en la concentración de urea es de 2.05 ± 0.19 kcal/mol/[urea]. Los resultados sugieren que el desplegamiento de Con1 es cooperativo: cualquier desplegamiento parcial de la estructura desestabiliza la estructura restante, lo cual provoca un colapso simultáneo a una estructura desplegada (BINSHENG, 2011).

La estabilidad térmica de la proteína Con1 y AroE se evaluó al inducir un desplegamiento térmico seguido por dicrósmo circular en el uv lejano y por calorimetría (DSC: *Differential SCanning Calorimetry*). Las mediciones de desnaturalización térmica mediante DC son

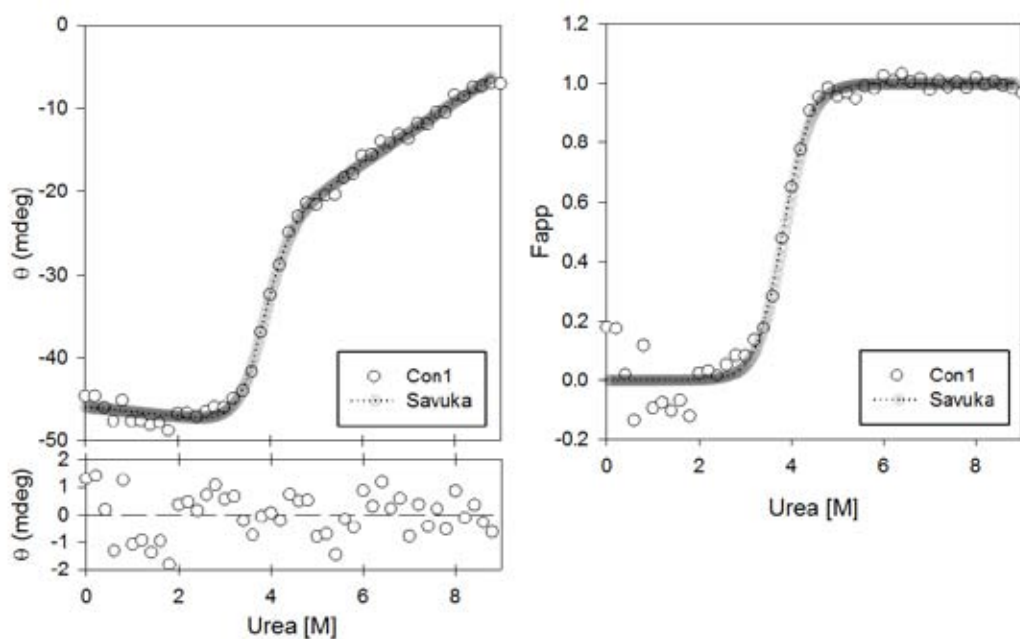


Figura 21: Desnaturalización química de Con1. (A) Valores obtenidos durante el experimento de desnaturalización química de Con1, en gris se observa el modelo obtenido con Savuka. (B) Fracción aparente de la desnaturalización química de Con1, donde 0 indica el estado nativo y 1 es el estado desplegado; en gris se observa el modelo obtenido con Savuka. (C) Valores residuales obtenidos al realizar el ajuste de los datos de Con1 con el modelo de dos estados. Resultados similares fueron obtenidos de tres experimentos independientes.

comúnmente usadas para evaluar la estabilidad de las proteínas y son especialmente usadas para evaluar los efectos de las mutaciones de aminoácidos en la estructura de la proteína (PELTON & MCLEAN, 2000).

No se pudo realizar un análisis termodinámico para determinar las energías libres de plegamiento debido a que la desnaturalización térmica de ambas proteínas fue irreversible (Figura 22). En su lugar, la temperatura aparente de fusión (T_m^{app}) a la cual el 50 % de la señal observada cambia debido a que la mitad de la proteína se encuentra desplegada, fue usada para medir la estabilidad.

En la curva de desnaturalización térmica de AroE (Figura 22) podemos observar una curva de desnaturalización sigmoidea que, por definición, corresponde a un mecanismo de dos estados (SHAW, SCHOLTZ, PACE, & GRIMSLEY, 2009). Mientras que la curva de desnaturalización térmica de Con1 muestra un indicio de otra transición, lo cual podría indicar que el mecanismo de desnaturalización de esta proteína no es de dos estados.

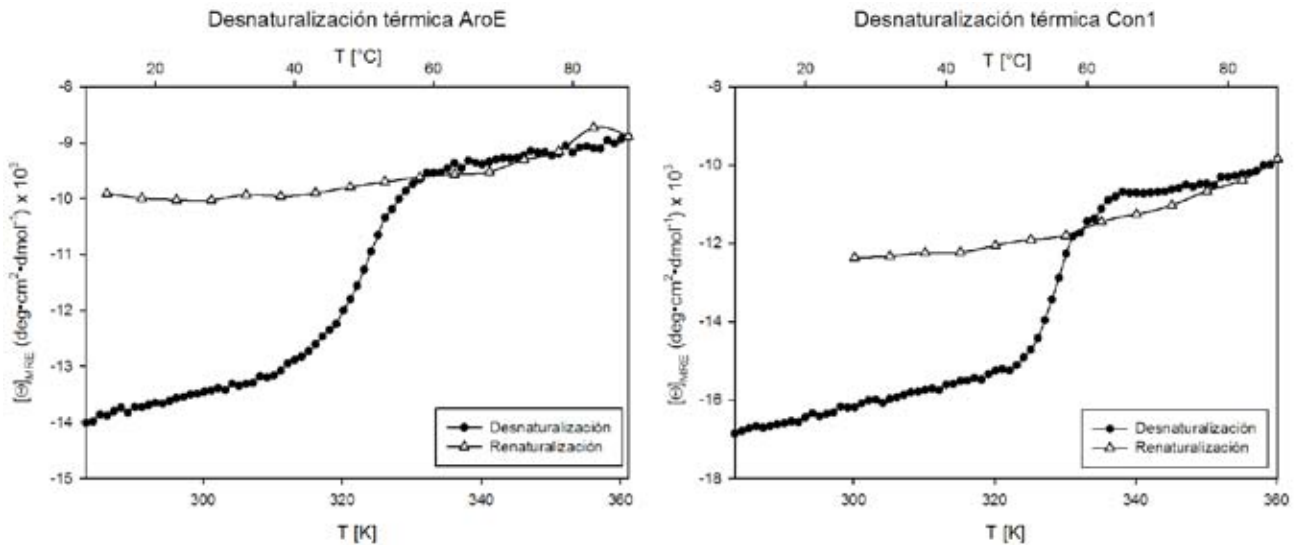


Figura 22: Desnaturalización térmica de AroE y Con1. El proceso de renaturalización deber ser superponible a los datos obtenidos durante la desnaturalización para poder afirmar que el proceso de desnaturalización térmica es reversible. Resultados similares fueron obtenidos de tres experimentos independientes.

Los experimentos de desnaturalización térmica se realizaron al monitorear el cambio de elipticidad a 222 nm al incrementar la temperatura. La elipticidad medida fue normalizada y el punto medio donde se observa una transición de desplegamiento (T_m^{app}) fue determinado usando el software *Savuka*, al ajustar los datos de desnaturalización térmica a un modelo de dos o tres estados (Figura 23).

En la Figura 23 se graficó la fracción aparente del estado desnaturalizado contra la temperatura. En gris se muestra el ajuste al modelo de dos estados obtenido con el programa *Savuka*, las curvas muestran una transición sigmoideal, lo cual es un indicativo de una reacción de desplegamiento cooperativo. Los datos obtenidos sugieren que AroE comienza a perder estructura a partir de los 38 °C y alcanza el estado desplegado a los 57 °C, con una T_m^{app} de 50 °C. Mientras que Con1 comienza a perder estructura a partir de los 52 °C y alcanza el estado desplegado a los 61 °C, con una T_m^{app} de 56 °C. Finalmente no fue posible obtener un ajuste de tres estados para la curva de desnaturalización térmica de Con1.

Aun cuando los resultados obtenidos mediante la desnaturalización química y térmica de AroE y Con1, son consistentes con un mecanismo de dos estados esto no demuestra este modelo con certeza (LUMRY & BILTONEN, 1966).

La calorimetría diferencial de barrido mide la capacidad calorífica de los estados y el exceso de calor asociado con las transiciones que puede ser inducido por el cambio de temperatura. Esta técnica puede ser usada para determinar la estabilidad del equilibrio termodinámico y el mecanismo de plegamiento (C. M. Johnson, 2013).

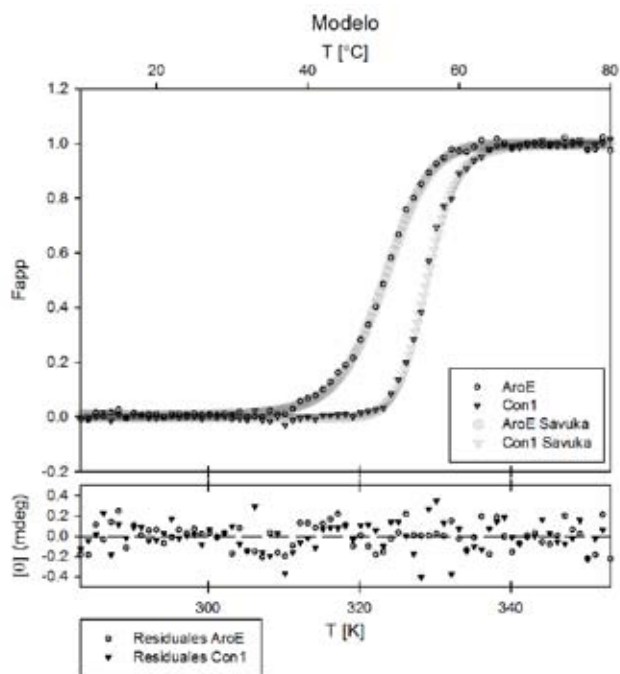


Figura 23: Modelo de la desnaturalización térmica de AroE y Con1.

En la parte superior se muestra la fracción aparente de AroE (círculos abiertos) y Con1 (triángulos abiertos), en gris se muestra el ajuste al modelo de dos estados obtenido con Savuka. En la parte inferior se muestran los valores residuales obtenidos al realizar el ajuste de los datos de AroE y Con1 con el modelo de dos estados.

Para determinar los perfiles de calorimetría de las proteínas AroE y Con1, a los resultados obtenidos se les corrigió la línea base del instrumento y todos los datos se normalizaron de acuerdo a su concentración para obtener los valores de temperatura aparente de fusión (T_m^{app}) debido a que ninguno de los termogramas mostró un proceso reversible.

Los perfiles de calorimetría de las proteínas AroE y Con1 mostraron ambas un comportamiento correspondiente a un monómero ya que no se observó dependencia sobre la concentración de proteína. El perfil de calorimetría de AroE muestra una única transición térmica que comienza a partir de los 42 °C y finaliza alrededor de los 60 °C. Mientras que el perfil de Con1 muestra dos transiciones térmicas, la primera de ellas aparece a los 45 °C y la segunda a los 57 °C (Figura 24). El valor de T_m^{app} de AroE fue de 50.9 °C, mientras que el de Consenso 1 fue de 53.2 °C.

La T_m^{app} de la enzima Con1 es solamente 2.3 °C mayor que la determinada para AroE en el laboratorio, y sólo 0.7 °C menor que la T_m de una shikimato deshidrogenasa de *Haemophilus influenzae* (SINGH *et al.*, 2005) pero 12.7 °C menor que la enzima shikimato deshidrogenasa

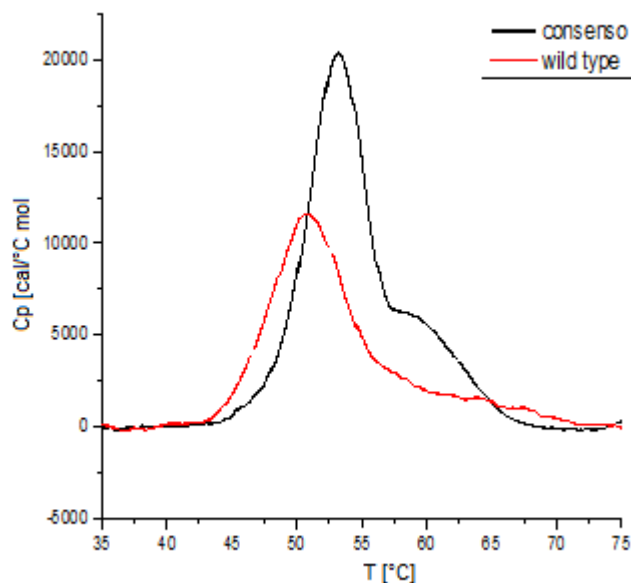
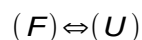


Figura 24: Perfiles de calorimetría de AroE (wild type) y Con1 (consenso). El valor de T_m^{app} de AroE fue de 50.9 °C (324.1 K), mientras que el de Consenso 1 fue de 53.2 °C (326.4 K). Resultados similares fueron obtenidos de tres experimentos diferentes.

de *Mycobacterium tuberculosis* (Arcuri et al., 2008). Aunque es generalmente cierto que las proteínas con una T_m^{app} alta son más estables que aquellas con valores de T_m^{app} bajos, este parámetro no es una medida definitiva de estabilidad (WHITFORD, 2005).

La espectrometría de masas es una técnica analítica sensible, relativamente simple y robusta, que permite identificar proteínas (en cantidades que van desde los picomoles hasta zeptomoles) si la secuencia de su genoma es conocida (GUERRERA & KLEINER, 2005). Esta técnica se basa en la medición de la relación masa/carga (m/z) de especies moleculares, estas determinaciones requieren de la generación de especies cargadas eléctricamente, lo cual se logra mediante la ionización por nano-electrospray (YATES, RUSE, & NAKORCHEVSKY, 2009). La secuenciación demostró que las bandas analizadas contienen la misma proteína: una *shikimate* deshidrogenasa modificada (Figura 25).

La proteína nativa, AroE, se ajusta a un proceso de dos estados, que puede ser representado como:



donde F y U son los estados plegado y desplegado de la proteína. Y en este caso, la única transición observada corresponde a la transformación del estado F al estado U, la cual comienza alrededor de los 40 °C y finaliza antes de los 60 °C.

40-20145egovis #4639-5572 RT: 52.60-60.48 AV: 295 NL: 7.58E7

T: FTMS + cNSI Full ms [300.00-2000.00]

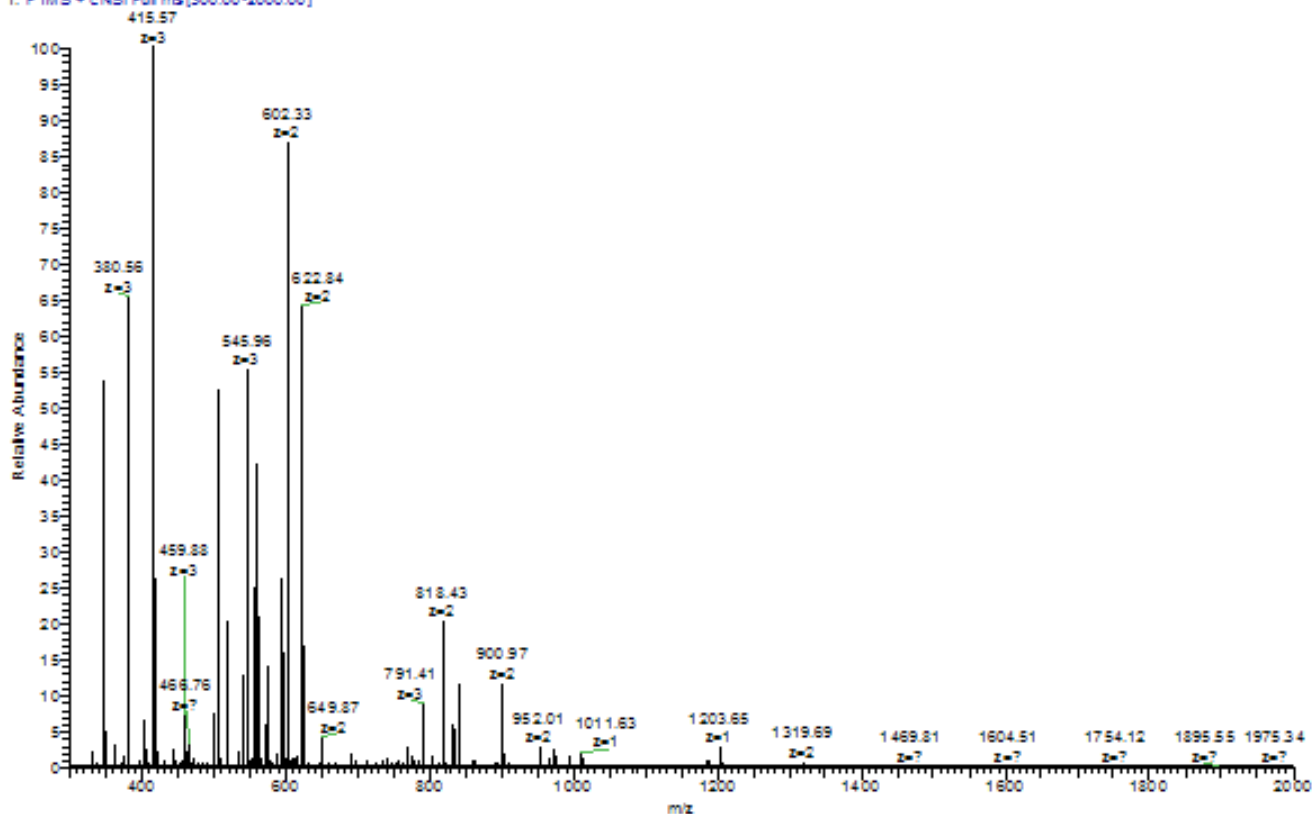
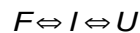


Figura 25: Espectrometría de masas de Con1. El gráfico de la izquierda corresponde a una shikimato deshidrogenasa modificada (Con1)

Mediante dos técnicas independientes, con muestras de lotes diferentes, se observa la presencia de otra transición. Esta transición sugiere la presencia de un núcleo más estable, el cual puede corresponder al dominio Rossmann consenso. Finalmente, se sugiere que el proceso de desplegamiento de Con1 podría corresponder a un proceso de tres estados, que puede ser representado como:



donde F, I e U son los estados plegado, intermediario y desplegado de la proteína, respectivamente. La primera transición, que corresponde a la transformación del estado F al estado I, comienza en 45 °C, con un máximo a 53.2 °C; esta transición puede corresponder al desplegamiento de la fracción no estabilizada de la enzima

La segunda transición, que corresponde al desplegamiento del estado I, la cual se observa a partir de 57 °C, y finalmente termina en el estado U a 70°C, con un máximo en 62 °C, podría corresponder al desplegamiento del núcleo estabilizado por las mutaciones consenso.

En el caso de Con1, las mutaciones consenso introducidas solamente en el dominio Rossmann son capaces de aumentar la estabilidad conformacional de toda la proteína. Mediante dos técnicas independientes, con muestras de lotes diferentes, se observa la presencia de otra transición, que sugiere la presencia de un núcleo más estable.

ENSAYOS DE ACTIVIDAD

Se realizaron ensayos de complementación para determinar la actividad *in vivo* de las enzimas consenso. En medio rico (LB con antibiótico) todas las muestras presentan crecimiento desde las primeras horas. Al monitorear el crecimiento en medio mínimo M9, se puede observar que AroE y Con1 presentan crecimiento después de las veinte horas (Figura 26), esto indica que Con1 es capaz de emular la actividad *in vivo* de AroE. Debido a que Con2 y Con3 no mostraron crecimiento en medio mínimo M9, no se muestran sus datos; estos resultados son congruentes con la observación previa de que estas proteínas no mostraban expresión.

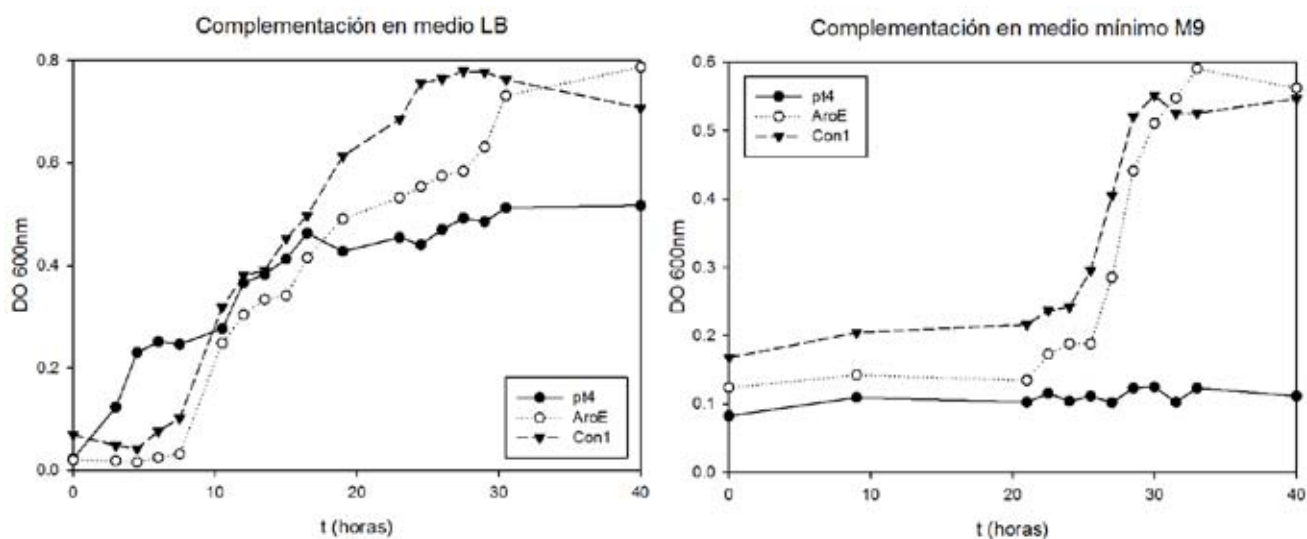


Figura 26: *Complementación en medio rico (LB con antibiótico) y medio mínimo M9.* En medio rico (gráfico izquierdo) todas las muestras presentan crecimiento desde las primeras horas. En medio mínimo M9 (gráfico derecho), AroE (control positivo) y Con1 muestran crecimiento después de las veinte horas. Con1 es capaz de emular la actividad *in vivo* de AroE.

Resultados similares fueron obtenidos de tres experimentos independientes.

La actividad enzimática de Consenso 1 fue determinada a 25 °C al monitorear la reducción de NADP⁺ o NAD⁺ a 340 nm en la presencia de ácido *shikímico*. Los datos obtenidos (bajo las condiciones de velocidad inicial, y al calcular la pendiente de la parte lineal de cada reacción) se normalizaron para poder ajustarse a la ecuación de Michaelis-Menten. El análisis de datos se llevó a cabo usando *SigmaPlot 12.5*.

Para normalizar los datos obtenidos se usó la siguiente fórmula:

$$V_i = \frac{\frac{Abso}{min}}{l * \epsilon} * FD$$

Donde $Abso/min$ es la pendiente que se obtiene en cada reacción, l es el ancho de la celda en centímetros (1 cm), ϵ es el coeficiente de extinción molar del NADP+ o NAD+ ($6220 \text{ M}^{-1} \text{ cm}^{-1}$), y FD es el factor de dilución que se obtiene dividiendo el volumen total de la reacción entre el volumen agregado de enzima.

Una vez obtenido el valor de V_i , se sustituyó en la ecuación de Michaelis-Menten:

$$V_i = \frac{k_{cat} * E * S}{K_m + S}$$

$$\frac{V_i}{E} = \frac{k_{cat} * S}{K_m + S}$$

Donde E es la concentración del stock donde se tomó la enzima. Y se graficó la concentración de cada sustrato o cofactor contra V_i/E (Figura 27).

La capacidad de Con1 para usar ácido shikímico como sustrato se determinó al monitorear la producción de NADPH+ a 340 nm y 25 °C. Los valores de K_M y V_{max} se determinaron al variar las concentraciones de ácido *shikímico* mientras se mantenía el NADP+ en condiciones saturantes. Se observó que Con1 cataliza la oxidación de *shikimato* en la presencia de NADP+ con una k_{cat} de $924.65 \pm 18.40 \text{ s}^{-1}$ y una constante de Michaelis de $231.51 \pm 17.35 \mu\text{M}$, presentando una R^2 de 0.935.

La capacidad de Con1 para usar NADP+ como cofactor se determinó al monitorear la producción de NADP+ a 340 nm y 25 °C en la presencia de ácido *shikímico*. Los valores de K_M y V_{max} se determinaron al variar las concentraciones de NADP+ mientras se mantenía el ácido *shikímico* en condiciones saturantes. Se observó que Con1 cataliza la oxidación de *shikimato* en la presencia de NADP+ con una k_{cat} de $1067.88 \pm 24.04 \text{ s}^{-1}$ y una constante de Michaelis de $279.28 \pm 23.47 \mu\text{M}$. Con una R^2 de 0.932.

También se determinó la capacidad de Con1 para usar NAD+ como cofactor se determinó al monitorear la producción de NADP+ a 340 nm y 25 °C en la presencia de ácido *shikímico*. Los valores de K_M y V_{max} se determinaron al variar las concentraciones de NAD+ mientras se mantenía el ácido *shikímico* en condiciones saturantes. Se observó que Con1 cataliza la oxidación de *shikimato* en la presencia de NAD+ con una k_{cat} de $250.37 \pm 36.65 \text{ s}^{-1}$ y una constante de Michaelis de $9915.99 \pm 2359.49 \mu\text{M}$. Con una R^2 de 0.948.

Los valores obtenidos de los parámetros cinéticos de Con1 se compararon con los reportados para AroE y YdiB (Michel *et al.*, 2003) (Tabla 1).

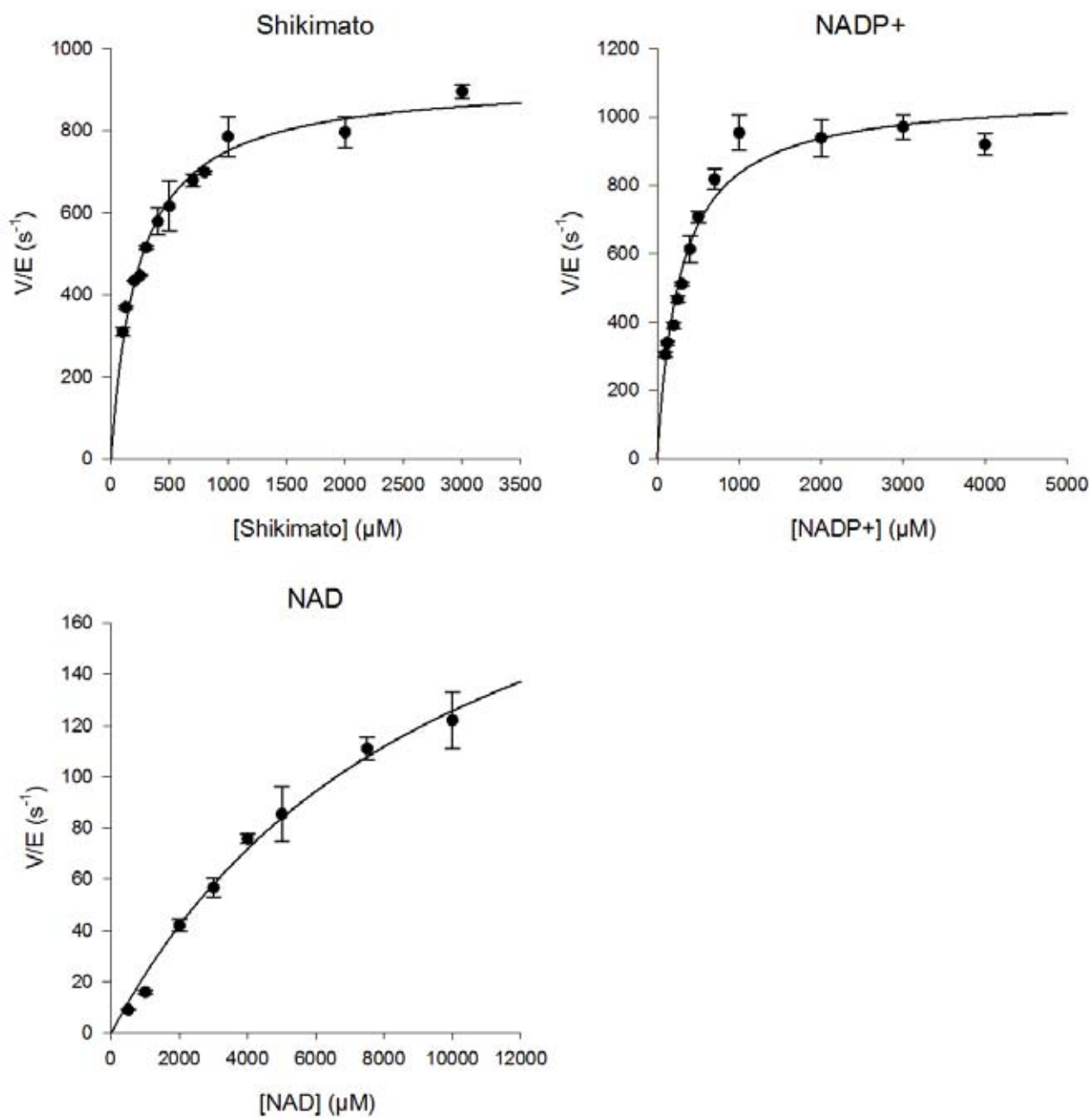


Figura 27: Michaelis-Menten de Con1. (A) Shikimate como el sustrato que varía mientras NADP+ se mantuvo constante. (B) NADP+ como cofactor que varía mientras que shikimate se mantuvo constante. (C) NAD+ como cofactor que varía mientras que shikimate se mantuvo constante.

Tabla 1. Comparación de parámetros cinéticos de diferentes enzimas shikimato deshidrogenasas. Los parámetros cinéticos fueron determinados a pH 9.0 y 25°C al variar la concentración de sustrato o cofactor mientras el otro se mantenía a concentraciones fijas. Los parámetros cinéticos de AroE y YdiB fueron determinados por (Michel et al., 2003).

	Cinética dependiente de sustrato			Cinética dependiente de cofactor					
	Shikimato			NADP+			NAD+		
Enzima	k_{cat} (s ⁻¹)	K_M (μM)	k_{cat}/K_M (s ⁻¹ M ⁻¹)	k_{cat} (s ⁻¹)	K_M (μM)	k_{cat}/K_M (s ⁻¹ M ⁻¹)	k_{cat} (s ⁻¹)	K_M (μM)	k_{cat}/K_M (s ⁻¹ M ⁻¹)
Con1	925	232	4x10 ⁶	1068	279	3.83x10 ⁶	250	9916	2.52x10 ⁴
AroE (b)	237	65	3.65x10 ⁶	237	56	4.23x10 ⁶	NA	NA	NA
YdiB (c)	0.12	120	1x10 ³	0.12	100	1.2x 10 ³	0.05	87	5.74x10 ²

NA: No presenta actividad.

Se ha reportado que las enzimas pertenecientes al grupo SDH tienen una estricta o fuerte preferencia al NADP+ en lugar del NAD+, y que las enzimas que usan este último como cofactor presentan una eficiencia reducida (PEEK & CHRISTENDAT, 2014). Michel *et al* han reportado que AroE presenta una estricta preferencia al NADP+ como cofactor y carece de actividad en la presencia de NAD+; mientras que YdiB presenta actividad con ambos cofactores pero muestra una preferencia fuerte hacia NADP+.

Con1 presenta una eficiencia catalítica similar a la de AroE cuando ocupa NADP+ como cofactor, y al contrario de AroE, presenta actividad con NAD+. Con1 muestra una K_M en presencia de NAD+ 35 veces más alta que la K_M de NADP, lo que sugiere que NADP+ es el cofactor preferido de Con1. En presencia de NAD+ Con1 muestra una eficiencia catalítica mayor (en dos órdenes de magnitud) a la de YdiB.

Estudios anteriores sugieren que una secuencia consenso se supone que debe parecerse al ancestro evolutivo de una familia de enzimas, en cuyo caso, debería de mostrar un comportamiento más promiscuo, esto es menor actividad frente a un rango más amplio de sustratos (BERSHTEIN, GOLDIN, & TAWFIK, 2008). Pero este no ha sido el caso para las fitasas consenso (LEHMANN *et al.*, 2000), la sucrosa forforilasa consenso (AERTS *et al.*, 2013) y la enzima shikimato deshidrogenasa con dominio Rossmann aquí reportada.

SECCIÓN
07



CONCLUSIONES



*Diseño y Caracterización
de un Dominio
Rossmann Consenso*



CLAUDIA IRIS BRAVO BONILLA

CONCLUSIONES

Se construyeron tres enzimas *shikimate* deshidrogenasas con dominios Rossmann consenso a partir de diferentes alineamientos de secuencias múltiples. En cada uno de los consensos se conservaron los aminoácidos que co-evolucionan, y se usó como secuencia guía la secuencia de AroE de *Escherichia coli*.

Sólo uno de los consensos (Con1) se expresó. Con1 se expresó en la fracción soluble, obteniéndose de 65-70 mg de proteína al final de la purificación por cada litro de cultivo, más del doble de lo obtenido con AroE.

No fue posible expresar ninguna de las otras dos proteínas consenso (Con2 y Con3) en las diferentes temperaturas de cultivo probadas, aún cuando las tres secuencias consenso presentan un 90 % de identidad en secuencia. Ambos consensos no expresados presentan un mayor número de mutaciones (32 para Con2 y 17 para Con3), y provienen de alineamientos con mayor número de secuencias (275 secuencias para Con2 y 96 secuencias para Con3) que el correspondiente a Con1 (47 secuencias, 15 mutaciones).

Al determinar la estructura secundaria de Con1 mediante dicroísmo circular se observó que el espectro de la proteína Con1 presenta un contenido de estructura secundaria superior al de AroE en un 2 %, lo cual indica que las mutaciones introducidas mediante el consenso modificaron de forma ligera la estructura secundaria.

Los datos de desnaturalización química con urea de Con1 se ajustaron a un modelo de transición de dos estados usando el software *Savuka*, y los resultados obtenidos sugieren que el desplegamiento de Con1 y AroE es cooperativo. Debido a que la desnaturalización térmica de Con1 fue irreversible no se pudieron determinar las energías libres de desplegamiento. El análisis de los datos con *Savuka* sugiere una T_m^{app} de 55.6 °C, la cual es 5.6 °C mayor a la determinada para AroE (50 °C).

El perfil de calorimetría de Con1 muestra un comportamiento correspondiente a un monómero y que no es dependiente de la concentración. El valor de T_m^{app} de Con1 fue de 53.2 °C, el cual es 2.3 °C mayor que el determinado para AroE (50.9 °C) en el laboratorio. Con1 también muestra dos transiciones térmicas, la primera de ellas aparece a los 45 °C y la segunda a los 57 °C, lo cual indica que el desplegamiento de Con1 podría corresponder a un proceso de tres estados.

Con1 es capaz de emular la actividad *in vivo* de AroE. Al determinar la actividad *in vitro* de Con1, observamos que presenta una eficiencia catalítica similar a la de AroE cuando ocupa NADP⁺ como cofactor, y al contrario de AroE, presenta actividad con NAD⁺. Con1 también

muestra una K_M en presencia de NAD⁺ treinta y cinco veces más alta que la K_M de NADP, lo que sugiere que NADP⁺ es el cofactor preferido de Con1. En presencia de NAD⁺ Con1 muestra una eficiencia catalítica mayor (en dos órdenes de magnitud) a la de YdiB.

El diseño de consensos nos permitió mutar quince aminoácidos del dominio Rossmann de la enzima *shikimate* deshidrogenasa de *E. coli*. Estos aminoácidos consenso incrementaron la estabilidad conformacional de toda la enzima, al generar un núcleo más estable, en al menos 2 °C en su T_m^{app} . Todo esto junto con una mejoría en algunas de sus características: un aumento en sus constantes cinéticas y el hecho de que es capaz de usar otro cofactor para llevar a cabo su actividad catalítica.

Esperamos que este dominio Rossmann consenso nos permita desarrollar proteínas quiméricas estables, solubles y activas, mediante el intercambio de dominios de proteínas homólogas.



SECCIÓN
08



PERSPECTIVAS



*Diseño y Caracterización
de un Dominio
Rossmann Consenso*



CLAUDIA IRIS BRAVO BONILLA

PERSPECTIVAS

Mediante diseño racional buscar mutaciones puntuales que incrementarán de forma significativa la estabilidad del consenso expresado (Con1) y/o la actividad catalítica en presencia de otros cofactores. Realizar un análisis de correlación de mutaciones (CMA, *correlated mutation analysis*) para explorar el rol de las redes de correlación en el consenso y posteriormente examinar si cada una de las mutaciones consensos tiene efectos neutrales, estabilizantes o desestabilizantes.

Es necesario realizar un proceso de purificación más astringente para eliminar toda contaminación posible y descartar que la contaminación sea la causa de la segunda transición observada en calorimetría. Para esto sería necesario realizar separación por punto isoeléctrico, cromatografía de exclusión por tamaño o remover la etiqueta de histidinas de la proteína Con1.

Una vez asegurada la pureza de la proteína Con1 es necesario realizar más experimentos de dicroísmo circular y calorimetría con diferentes condiciones (rampa de temperatura, concentraciones) para determinar si el proceso de desplegamiento corresponde a dos o tres estados. Todos estos experimentos también deben realizarse a la proteína AroE, para poder comparar los resultados.

Finalmente sería buena idea expresar solamente el dominio Rossmann consenso Con1, para poder determinar sus características, y compararlas contra las de otros dominios Rossmann conocidos. Para posteriormente realizar quimeras con otras enzimas que presenten un dominio Rossmann.

En el caso de Con2 y Con3 es necesario determinar si existe transcripción de los genes consenso, por lo que podría realizarse RT-PCR semi-cuantitativa para monitorear los niveles de expresión de los genes. En caso de que no exista transcripción, se podría cambiar la secuencia guía utilizada. Otro de los aspectos que podrían ayudar a la expresión de los dos consensos restantes, sería usar una cepa que presente chaperonas que ayuden al plegamiento de los consensos.



SECCIÓN
09



BIBLIOGRAFÍA



*Diseño y Caracterización
de un Dominio
Rossmann Consenso*



CLAUDIA IRIS BRAVO BONILLA

BIBLIOGRAFÍA

- AERTS, D., VERHAEGHE, T., JOOSTEN, H.-J., VRIEND, G., SOETAERT, W., & DESMET, T. (2013). Consensus engineering of sucrose phosphorylase: the outcome reflects the sequence input. *Biotechnology and Bioengineering*, 110(10), 2563–72. doi:10.1002/bit.24940
- AMIN, N., LIU, A., & RAMER, S. (2004). Construction of stabilized proteins by combinatorial consensus mutagenesis. *Protein Engineering, Design & Selection : PEDS*, 17(11), 787–793. doi:10.1093/protein/gzh091
- ARCURI, H. A., BORGES, J. C., FONSECA, I. O., PEREIRA, J. H., NETO, J. R., BASSO, L. A., ... DE AZEVEDO, W. F. (2008). Structural studies of shikimate 5-dehydrogenase from *Mycobacterium tuberculosis*. *Proteins*, 72(2), 720–30. doi:10.1002/prot.21953
- BABA, T., ARA, T., HASEGAWA, M., TAKAI, Y., OKUMURA, Y., BABA, M., ... MORI, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, 2, 1–11. doi:10.1038/msb4100050
- BANDA, J. (2013). *Análisis y diseño de dominios Rossmann de unión a dinucleótidos*. Universidad Nacional Autónoma de México.
- BASHTON, M., & CHOTHIA, C. (2002). The geometry of domain combination in proteins. *Journal of Molecular Biology*, 315(4), 927–39. doi:10.1006/jmbi.2001.5288
- BENACH, J., LEE, I., EDSTROM, W., KUZIN, A. P., CHIANG, Y., ACTON, T. B., ... HUNT, J. F. (2003). The 2.3-Å crystal structure of the shikimate 5-dehydrogenase orthologue YdiB from *Escherichia coli* suggests a novel catalytic environment for an NAD-dependent dehydrogenase. *The Journal of Biological Chemistry*, 278(21), 19176–82. doi:10.1074/jbc.M301348200
- BINSHENG, Y. (2011). The Stability of a Three-State Unfolding Protein. In *Thermodynamics - Physical Chemistry of Aqueous Systems* (pp. 365–3990).
- BINZ, H. K. K., STUMPP, M. T., FORRER, P., AMSTUTZ, P., & PLÜCKTHUN, A. (2003). Designing Repeat Proteins: Well-expressed, Soluble and Stable Proteins from Combinatorial Libraries of Consensus Ankyrin Repeat Proteins. *Journal of Molecular Biology*, 332(2), 489–503. doi:10.1016/S0022-2836(03)00896-9
- BOTTOMS, C. A., SMITH, P. E., & TANNER, J. J. (2002). A structurally conserved water molecule in Rossmann dinucleotide-binding domains, 2125–2137. doi:10.1110/ps.0213502.citric
- CARR, R., ALEXEEVA, M., ENRIGHT, A., EVE, T. S. C., DAWSON, M. J., & TURNER, N. J. (2003). Directed evolution of an amine oxidase possessing both broad substrate specificity and high enantioselectivity. *Angewandte Chemie (International Ed. in English)*, 42(39), 4807–10. doi:10.1002/anie.200352100
- CASALI, N. (2003). *Escherichia coli* host strains. *Methods in Molecular Biology (Clifton, N.J.)*, 235(4), 27–48. doi:10.1385/1-59259-409-3:27
- CGSC. (2014). *E. coli* Genetic Resources at Yale CGSC, The Coli Genetic Stock Center. Retrieved July 14, 2014, from <http://cgsc.biology.yale.edu/Strain.php?ID=121248>
- CHEN, B. L., BAASE, W. A., NICHOLSON, H., & SCHELLMAN, J. A. (1992). Folding kinetics of T4 lysozyme and nine mutants at 12 .degree.C. *Biochemistry*, 31(5), 1464–1476. doi:10.1021/bi00120a025
- CHERRY, J. R. (2000). Directed evolution of microbial oxidative enzymes. *Current Opinion in Biotechnology*, 11(3), 250–4.

- CRUZ-VERA, L. R., MAGOS-CASTRO, M. A., ZAMORA-ROMO, E., & GUARNEROS, G. (2004). Ribosome stalling and peptidyl-tRNA drop-off during translational delay at AGA codons. *Nucleic Acids Research*, 32(15), 4462–8. doi:10.1093/nar/gkh784
- DAI, M., FISHER, H. E., TEMIROV, J., KISS, C., PHIPPS, M. E., PAVLIK, P., ... BRADBURY, A. R. M. (2007). The creation of a novel fluorescent protein by guided consensus engineering. *Protein Engineering, Design & Selection : PEDS*, 20(2), 69–79. doi:10.1093/protein/gzl056
- DALBY, P. A. (2007). Engineering enzymes for biocatalysis. *Recent Patents on Biotechnology*, 1(1), 1–9.
- DATSENKO, K. A., & WANNER, B. L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America*, 97(12), 6640–5. doi:10.1073/pnas.120163297
- Escherichia coli str. K-12 substr. MG1655 chromosome, complete genome. (2012). Retrieved from http://www.ncbi.nlm.nih.gov/nuccore/NC_000913.2?from=3428042&to=3428860&report=fasta&strand=true
- ESPINOSA CANTÚ, A. (2010). *Obtención de una quimera activa por la recombinación de dominios entre dos shikimate deshidrogenasas*. Universidad Nacional Autónoma de México.
- FRISHMAN, D., & ARGOS, P. (1995). Knowledge-based protein secondary structure assignment. *Proteins*, 23(4), 566–79. doi:10.1002/prot.340230412
- GAO, F., WEAVER, E., LU, Z., LI, Y., & LIAO, H. (2005). Antigenicity and immunogenicity of a synthetic human immunodeficiency virus type 1 group m consensus envelope glycoprotein. *Journal of Virology*, 79(2), 1154–1163. doi:10.1128/JVI.79.2.1154
- GASTEIGER, E. (2003). ExpASY: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, 31(13), 3784–3788. doi:10.1093/nar/gkg563
- GLIEDER, A., FARINAS, E. T., & ARNOLD, F. H. (2002). Laboratory evolution of a soluble, self-sufficient, highly active alkane hydroxylase. *Nature Biotechnology*, 20(11), 1135–9. doi:10.1038/nbt744
- GOUY, M., GUINDON, S., & GASCUEL, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, 27(2), 221–4. doi:10.1093/molbev/msp259
- GRATCHEV, A. (n.d.). Ligating a DNA fragment into a plasmid vector. Retrieved from http://www.methods.info/Methods/RNA_DNA/ligation_simple.html
- HERRMANN, K. M., & WEAVER, L. M. (1999). the Shikimate Pathway. *Annual Review of Plant Physiology and Plant Molecular Biology*, 50, 473–503. doi:10.1146/annurev.arplant.50.1.473
- JACOBS, S., DIEM, M., LUO, J., TEPLYAKOV, A., OBMOLOVA, G., MALIA, T., ... O'NEIL, K. (2012). Design of novel FN3 domains with high stability by a consensus sequence approach. *Protein Engineering, Design & Selection : PEDS*, 25(3), 107–117. doi:10.1093/proeng/gzr064
- JIANG, X., KOWALSKI, J., & KELLY, J. (2001). Increasing protein stability using a rational approach combining sequence homology and structural alignment: Stabilizing the WW domain. *Protein Science*, 1454–1465. doi:10.1101/ps.640101.based.
- JOHNSON, W. C. (1999). Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins*, 35(3), 307–12.
- LARKIN, M. A., BLACKSHIELDS, G., BROWN, N. P., CHENNA, R., MCGETTIGAN, P. A., MCWILLIAM, H., ... HIGGINS, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)*, 23(21), 2947–8. doi:10.1093/bioinformatics/btm404
- LEHMANN, M., PASAMONTES, L., LASSEN, S. F., & WYSS, M. (2000). The consensus concept for thermostability engineering of proteins. *Biochimica et Biophysica Acta*, 1543(2), 408–415.

- LEHMANN, M., & WYSS, M. (2001). Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. *Current Opinion in Biotechnology*, 12(4), 371–5.
- LI, W., & GODZIK, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics (Oxford, England)*, 22(13), 1658–9. doi:10.1093/bioinformatics/btl158
- LI, WEIZHONG, & GODZIK, A. (2012). CD-HIT. Retrieved from <http://weizhong-lab.ucsd.edu/cd-hit/>
- LOCKLESS, S. W., & RANGANATHAN, R. (1999). Evolutionarily conserved pathways of energetic connectivity in protein families. *Science (New York, N.Y.)*, 286(5438), 295–9.
- LUMRY, R., & BILTONEN, R. (1966). Validity of the “two-state” hypothesis for conformational transitions of proteins. *Biopolymers*, 4(8), 917–44. doi:10.1002/bip.1966.360040808
- MA, B.-G., CHEN, L., JI, H.-F., CHEN, Z.-H., YANG, F.-R., WANG, L., ... ZHANG, H.-Y. (2008). Characters of very ancient proteins. *Biochemical and Biophysical Research Communications*, 366(3), 607–11. doi:10.1016/j.bbrc.2007.12.014
- MAXWELL, K. L., & DAVIDSON, A. R. (1998). Mutagenesis of a buried polar interaction in an SH3 domain: sequence conservation provides the best prediction of stability effects. *Biochemistry*, 37(46), 16172–82. doi:10.1021/bi981788p
- MERZ, A, YEE, M. C., SZADKOWSKI, H., PAPPENBERGER, G., CRAMERI, A., STEMMER, W. P., ... KIRSCHNER, K. (2000). Improving the catalytic activity of a thermophilic enzyme at low temperatures. *Biochemistry*, 39(5), 880–9.
- MICHEL, G., ROSZAK, A. W., SAUVÉ, V., MACLEAN, J., MATTE, A., COGGINS, J. R., ... LAPTHORN, A. J. (2003). Structures of shikimate dehydrogenase AroE and its Paralog YdiB. A common structural framework for different activities. *The Journal of Biological Chemistry*, 278(21), 19463–72. doi:10.1074/jbc.M300794200
- MONERA, O. D., KAY, C. M., & HODGES, R. S. (1994). Protein denaturation with guanidine hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions. *Protein Science : A Publication of the Protein Society*, 3(11), 1984–91. doi:10.1002/pro.5560031110
- MORÁN GARCÍA, A. DEL C. (2008). *Diseño y Construcción de una Cistatina Consenso*. Universidad Nacional Autónoma de México.
- NIKOLOVA, P. V, HENCKEL, J., LANE, D. P., & FERSHT, A R. (1998). Semirational design of active tumor suppressor p53 DNA binding domain with enhanced stability. *Proceedings of the National Academy of Sciences of the United States of America*, 95(25), 14675–80.
- PELTON, J. T., & MCLEAN, L. R. (2000). Spectroscopic methods for analysis of protein secondary structure. *Analytical Biochemistry*, 277(2), 167–76. doi:10.1006/abio.1999.4320
- PERSSON, B., KALLBERG, Y., OPPERMANN, U., & JÖRNVALL, H. (2003). Coenzyme-based functional assignments of short-chain dehydrogenases/reductases (SDRs). *Chemico-Biological Interactions*, 143-144, 271–8.
- RICE, P., LONGDEN, I., & BLEASBY, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends in Genetics*.
- RICE, P., LONGDEN, I., & BLEASBY, A. (2000). EMBOSS 6.3.1.: extractalign. Moby portal. Retrieved from <http://mobyle.pasteur.fr/cgi-bin/portal.py#forms::extractalign>
- ROBERT, X., & GOUET, P. (2014). Deciphering key features in protein structures with the new ENDScript server. *Nucleic Acids Research*, 42(April), 320–324. doi:10.1093/nar/gku316

- ROSSMANN, M. G., MORAS, D., & OLSEN, K. W. (1974). Chemical and biological evolution of a nucleotide-binding protein. *Nature*, 250(5463), 194–199. doi:10.1038/250194a0
- RÖTHLISBERGER, D., KHERSONSKY, O., WOLLACOTT, A. M., JIANG, L., DECHANCIE, J., BETKER, J., ... BAKER, D. (2008). Kemp elimination catalysts by computational enzyme design. *Nature*, 453(7192), 190–5. doi:10.1038/nature06879
- ROY, A., KUCUKURAL, A., & ZHANG, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, 5(4), 725–38. doi:10.1038/nprot.2010.5
- SCHMID, A., DORDICK, J. S., HAUER, B., KIENER, A., WUBBOLTS, M., & WITHOLT, B. (2001). Industrial biocatalysis today and tomorrow. *Nature*, 409(6817), 258–68. doi:10.1038/35051736
- SEGOVIA, L., & PEIMBERT, M. (2010). Ingeniería de proteínas y evolución dirigida. *Mensaje bioquímico*, xxxiv, 135–141.
- SERRANO, L., DAY, A., & FERSHT, A. (1993). Step-wise mutation of barnase to binase: a procedure for engineering increased stability of proteins and an experimental analysis of the evolution of protein stability. *Journal of Molecular Biology*, 305–312.
- SINGH, S., KOROLEV, S., KOROLEVA, O., ZAREMBINSKI, T., COLLART, F., JOACHIMIAK, A., & CHRISTENDAT, D. (2005). Crystal structure of a novel shikimate dehydrogenase from *Haemophilus influenzae*. *Journal of Biological Chemistry*, 280(17), 17101–8. doi:10.1074/jbc.M412753200.Crystal
- SINGH, S., STAVRINIDES, J., CHRISTENDAT, D., & GUTTMAN, D. S. (2008). A phylogenomic analysis of the shikimate dehydrogenases reveals broadSCALE functional diversification and identifies one functionally distinct subclass. *Molecular Biology and Evolution*, 25(10), 2221–32. doi:10.1093/molbev/msn170
- SONG, J. M., AN, Y. J., KANG, M. H., LEE, Y.-H., & CHA, S.-S. (2012). Cultivation at 6–10°C is an effective strategy to overcome the insolubility of recombinant proteins in *Escherichia coli*. *Protein Expression and Purification*, 82(2), 297–301. doi:10.1016/j.pep.2012.01.020
- STEIPE, B., SCHILLER, B., PLÜCKTHUN, A., & STEINBACHER, S. (1994). Sequence Statistics Reliably Predict Stabilizing Mutations in a Protein Domain. *Journal of Molecular Biology*.
- STRAFFORD, J., PAYONGSRI, P., HIBBERT, E. G., MORRIS, P., BATH, S. S., STEADMAN, D., ... DALBY, P. A. (2012). Directed evolution to re-adapt a co-evolved network within an enzyme. *Journal of Biotechnology*, 157(1), 237–45. doi:10.1016/j.jbiotec.2011.11.017
- SULLIVAN, B. J., DURANI, V., & MAGLIERY, T. J. (2011). Triosephosphate isomerase by consensus design: dramatic differences in physical properties and activity of related variants. *Journal of Molecular Biology*, 413(1), 195–208. doi:10.1016/j.jmb.2011.08.001
- TAMURA, K., STECHER, G., PETERSON, D., FILIPSKI, A., & KUMAR, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, 30(12), 2725–9. doi:10.1093/molbev/mst197
- Taxonomic Common Tree. (2012). Retrieved from <http://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi>
- TERPE, K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology*, 60(5), 523–33. doi:10.1007/s00253-002-1158-6
- THOMPSON, J. D., GIBSON, T. J., PLEWNIAC, F., JEANMOUGIN, F., & HIGGINS, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, 25(24), 4876–82.
- TU, Z., HE, G., LI, K. X., CHEN, M. J., CHANG, J., CHEN, L., ... WU, X. (2005). An improved system for competent cell preparation and high efficiency plasmid transformation using different *Escherichia coli* strains. *Electronic Journal of Biotechnology*, 8(1). doi:10.2225/vol8-issue1-fulltext-8

- VÁZQUEZ-FIGUEROA, E., CHAPARRO-RIGGERS, J., & BOMMARIUS, A. S. (2007). Development of a thermostable glucose dehydrogenase by a structure-guided consensus concept. *Chembiochem: A European Journal of Chemical Biology*, 8(18), 2295–301. doi:10.1002/cbic.200700500
- VILLARDÓN, J. L. V. (2011). *El Análisis de Clusters* (p. 22).
- WANG, Q., BUCKLE, A. M., FOSTER, N. W., JOHNSON, C. M., & FERSHT, A. R. (1999). Design of highly stable functional GroEL minichaperones. *Protein Science : A Publication of the Protein Society*, 8(10), 2186–93. doi:10.1110/ps.8.10.2186
- WHITFORD, D. (2005). *Proteins: Structure and Function* (p. 542). England: John Wiley & Sons Ltd.
- WHITMORE, L., & WALLACE, B. A. (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Research*, 32(Web Server issue), W668–73. doi:10.1093/nar/gkh371
- WIRTZ, P., & STEIPE, B. (1999). Intrabody construction and expression III: engineering hyperstable V(H) domains. *Protein Science : A Publication of the Protein Society*, 8(11), 2245–50. doi:10.1110/ps.8.11.2245
- YE, J., MCGINNIS, S., & MADDEN, T. L. (2006). BLAST: improvements for better sequence analysis. *Nucleic Acids Research*, 34(Web Server issue), W6–9. doi:10.1093/nar/gkl164
- YE, S., DELFT, F. VON, & BROOUN, A. (2003). The crystal structure of shikimate dehydrogenase (AroE) reveals a unique NADPH binding mode. *Journal of Bacteriology*. doi:10.1128/JB.185.14.4144



SECCIÓN
10



ANEXOS



*Diseño y Caracterización
de un Dominio
Rossmann Consenso*



CLAUDIA IRIS BRAVO BONILLA

ANEXOS

ANEXO 1: SECUENCIA DE AMINOÁCIDOS DE LA ENZIMA SHIKIMATO DESHIDROGENASA (ARO E) DE ESCHERICHIA COLI

La secuencia correspondiente a AroE se obtuvo de la base de datos de NCBI («Escherichia coli str. K-12 substr. MG1655 chromosome, complete genome», 2012)

>gi|49175990:c3428860-3428042 Escherichia coli str. K-12 substr. MG1655 chromosome, complete genome

```
METYAVFGNP IAHSKSPFIH QQFAQQLNIE HPYGRVLAPI NDFINTLNAF FSAGGKGANV  
TVPFKEEFAF RADELTERAA LAGAVNTLMR LEDGRLLGDN TDGVLLSDL ERLSFIRPGL  
RILLIGAGGA SRGVLLPLLS LDCAVTITNR TVSRAEELAK LFAHTGSIQA LSMDELEGHE  
FDLIINATSS GISGDIPAIP SSLIHPGIYC YDMFYQKGKT PFLAWCEQRG SKRNADGLGM  
LVAQAAHAFL LWHGVLPDVE PVIKQLQEEL S*
```

A continuación se muestra la secuencia de cada dominio de AroE.

- Secuencia correspondiente al dominio catalítico (aminoácidos 1-101 y 245-271).

Aminoácidos 1-101

```
METYAVFGNP IAHSKSPFIH QQFAQQLNIE HPYGRVLAPI NDFINTLNAF FSAGGKGANV  
TVPFKEEFAF RADELTERAA LAGAVNTLMR LEDGRLLGDN T
```

Aminoácidos 245-271

```
AAHAFL LWHGVLPDVE PVIKQLQEEL S
```

- Secuencia correspondiente al dominio Rossmann (aminoácidos 102-244).

Aminoácidos 102-244

```
DGVLLSDLE·RLSFIRPGLR·ILLIGAGGAS·RGVLLPLLSL·DCAVTITNRT·  
VSRAEELAKL FAHTGSIQAL·SMDELEGHEF·DLIINATSSG·ISGDIPAIPS·  
SLIHPGIYCY·DMFYQKGKTP FLAWCEQRGS·KRNADGLGML·VAQ
```

ANEXO 2: DOMINIOS ROSSMANN CONSENSOS DETERMINADOS POR EL PROGRAMA SEAVIEW

>Consenso1

DGIGLLSDLE·RLXFIKPGDR·ILLIGAGGAA·RGVILPLLSY·GCXITITNRT·
FSRAQXLAXX·FXHXGXIXAX·XXXLXXXXF·DLIINATSSG·IXGEIPXIPX·
SLIXXTXCY·DMFYQXGLTP·FLXWCXQXGA·XXXADGLGML·VGQ

>Consenso2

DGIGLVRDLX·RLGXXLXGKR·ILLIGAGGAA·RGVLLPLLXA·GPALVIANRT·
XSKAEXLAXX·FAXXGXIXAX·XXDXLXGXXF·DLIINATSAS·LXGELPPLPX·
XLFXXTLAY·DMMYGXXPTP·FLXWAXXGA·XXXADGLGML·VEQ

>Consenso3

DGXGFXXLX·XXXXXXXXXR·VLILGAGGAA·RAVAXALLXL·GXXXXIXNRT·
XXRAXXLAXX·FXXGXIXAX·XXDXLXXXXF·DLLINATSXG·MXGDXPXXPX·
XLLXGXVVY·DXVYXPXXTP·FLXXAXXGX·XRTXDGLGML·VXQ

ANEXO 3: DOMINIOS ROSSMANN CONSENSO DETERMINADOS POR EL PROGRAMA SEAVIEW CON LOS AMINOÁCIDOS INDETERMINADOS SUSTITUIDOS

>Consenso1

DGIGLLSDLE·RLSFIKPGDR·ILLIGAGGAA·RGVILPLLSY·GCAITITNRT·
FSRAQELAKL·FAHTGSIQAL·SMDELEGHEF·DLIINATSSG·ISGEIPAIPS·
SLIHPGTICY·DMFYQKGLTP·FLAWCEQRGA·KRNADGLGML·VGQ

>Consenso2

DGIGLVRDLE·RLGFILPGKR·ILLIGAGGAA·RGVLLPLLSA·GPALVIANRT·
VSKAEELAKL·FAHTGSIQAL·SMDELEGHEF·DLIINATSAS·LSGELPPLPS·
SLFHPGTLAY·DMMYGKGPPTP·FLAWAEQRGA·KRNADGLGML·VEQ

>Consenso3

DGVGFVSDLE·RLGFIRPGKR·VLVLGAGGAA·RAVLLALLSL·GCAITITNRT·
VSRAEELAKL·FAHTGSIQAL·SMDELEGHEF·DLIINATSSG·MSGDIPAIPS·
SLLHPGIYVY·DMVYQPGETP·LLAWAEQRGA·KRTADGLGML·VAQ

ANEXO 4: SECUENCIAS CONSENSO (AMINOÁCIDOS) SINTETIZADAS

>GenConsenso1

METYAVFGNP·IAHSKSPFIH·QQFAQQLNIE·HPYGRVLAPI·NDFINTLNAF·
FSAGGKGANV TVPFKEEFA·RADELTERAA·LAGAVNTLMR·LEDGRLLGDN·
TDGIGLLSDL·ERLSFIKPGD RILLIGAGGA·ARGVILPLLS·YGCAITITNR·
TFSRAQELAK·LFAHTGSIQA·LSMDELEGHE FDLIINATSS·GISGEIPAIP·
SSLIHPGTYC·YDMFYQKGLT·PFLAWCEQRG·AKRNADGLGM LVGQAAHAFL·
LWHGVLPDVE·PVIKQLQEEL·SHHHH*

>GenConsenso2

METYAVFGNP·IAHSKSPFIH·QQFAQQLNIE·HPYGRVLAPI·NDFINTLNAF·
FSAGGKGANV TVPFKEEFA·RADELTERAA·LAGAVNTLMR·LEDGRLLGDN·
TDGIGLVRDL·ERLGFILPGK RILLGAGGA·ARGVLLPLLS·AGPALVIANR·
TVSKAEELAK·LFAHTGSIQA·LSMDELEGHE FDLIINATSA·SLSGELPLP·
SSLFHPGTLA·YDMMYGKGP·PFLAWAEQRG·AKRNADGLGM LVEQAAHAFL·
LWHGVLPDVE·PVIKQLQEEL·SHHHHHH*

>GenConsenso3

METYAVFGNP·IAHSKSPFIH·QQFAQQLNIE·HPYGRVLAPI·NDFINTLNAF·
FSAGGKGANV TVPFKEEFA·RADELTERAA·LAGAVNTLMR·LEDGRLLGDN·
TDGVGFVSDL·ERLGFIRPGK RVLVLGAGGA·ARAVLLALLS·LGCAITITNR·
TVSRAEELAK·LFAHTGSIQA·LSMDELEGHE·FDLIINATSS·GMSGDIPAIP·
SSLLHPGIYV·YDMVYQPGET·PLLAWAEQRG·AKRTADGLGM LVAQAAHAFL·
LWHGVLPDVE·PVIKQLQEEL·SHHHHHH*





THÉSIKA

DISEÑO DE TESIS

- www.thesika.mx -



Diseño y Caracterización

de un Dominio

Rossmann Consenso



CLAUDIA IRIS BRAVO BONILLA