



Universidad Autónoma del Estado de México
Facultad de Ciencias

Inserción al azar de 6 aminoácidos en la región aminoterminal del dominio de PDH para la estabilización de la proteína T de *Escherichia coli*

TESIS

Que para obtener el grado de
Licenciada en Biotecnología

Presenta
Mariana Gutiérrez Solano

Dirigida por:
Dr. Humberto Flores Soto
Dra. María Elena Estrada Zúñiga

Febrero de 2019

A mis padres y hermanos

Agradecimientos Académicos

A mis directores de tesis, Dr. Humberto Flores Soto y Dra. María Elena Estrada Zúñiga por su compromiso y dedicación en este proyecto. Por sus conocimientos y el aprendizaje adquirido durante este proyecto.

Al Dr. Xavier Soberón por permitirme realizar este proyecto en su laboratorio.

A mis revisores de tesis, M. en C. Rodolfo Gamaliel Ávila Bonilla y Dra. Carla García Morales por su disposición y compromiso con este proyecto.

Al Instituto de Biotecnología-UNAM por poder permitirme el uso de sus instalaciones para la realización de este proyecto.

El presente trabajo recibió apoyo financiero del PAPIIT-UNAM (proyecto IN208816)

Índice

Índice de Figuras.....	4
Índice de Tablas.....	5
Lista de Abreviaturas.....	6
Resumen.....	7
1. Introducción.....	9
1.1. Proteínas: Definición y estructura.....	9
1.1.1 Dominios de las proteínas.....	11
1.1.2 Tipos de proteínas.....	12
1.1.3 Enzimas.....	13
1.2 Biosíntesis de aminoácidos aromáticos.....	17
1.2.1 La vía de shikimato.....	17
1.2.2 Biosíntesis de fenilalanina.....	19
1.2.3 Biosíntesis de triptófano.....	20
1.2.3. Biosíntesis de tirosina.....	21
1.3 Evolución de las proteínas.....	22
1.3.1 Estrategias para la evolución de proteínas: ingeniería de proteínas y evolución dirigida.....	24
1.3.2 Técnicas para la evolución dirigida.....	25
2. Antecedentes.....	28
3. Planteamiento del problema.....	31
4. Justificación.....	32
5. Hipótesis.....	33

6. Objetivos.....	34
6.1 Objetivo general.....	34
6.2 Objetivos particulares.....	34
7. Materiales y métodos.....	35
7.1. Diseño del oligonucleótido que introduce 6 codones al azar en la región 5' del gen que codifica el dominio PDH.....	35
7.2. Construcción de la librería de variantes del gen que codifica para el dominio PDH de la proteína T de <i>E. coli</i>	37
7.2.1. Clonación.....	37
7.3. Validación de la librería de variantes del gen que codifica para el dominio PDH de la proteína T de <i>E. coli</i>	39
7.4 Sistema de selección de variantes con actividad de PDH.....	40
7.4.1. Validación del fenotipo y del genotipo de las variantes.....	40
7.5. Caracterización de variantes con fenotipo y genotipo validado.....	42
7.5.1. Crecimiento de las variantes seleccionadas en medio líquido.....	42
7.5.2. Validación de la expresión del dominio PDH de las variantes seleccionadas mediante electroforesis en gel de acrilamida.....	43
7.5.3 Ensayo enzimático.....	44
7.6. Determinación del tamaño del banco.....	44
8. Resultados y discusión.....	45
8.1. Validación del oligonucleótido y librería de variantes del gen que codifica para el dominio de PDH de la proteína T de <i>E. coli</i>	45
8.2. Validación del fenotipo y del genotipo.....	47
8.3. Caracterización de las variantes seleccionadas.....	52
8.3.1 Determinación de crecimiento en medio líquido.....	53
8.3.2. Expresión y actividad enzimática específica del dominio PDH de las variantes seleccionadas.....	55
8.5 Determinación del banco de variantes.....	59

9. Conclusiones y perspectivas.....	60
10. Anexos.....	61
11. Bibliografía.....	66

Índice de Figuras

- Figura 1. Representación de la formación del enlace peptídico
- Figura 2. Estructura química general de un aminoácido
- Figura 3. Niveles de organización de la estructura de la proteína
- Figura 4. Pasos en la vía de Shikimato
- Figura 5. Biosíntesis de a) fenilalanina y b) triptofano a partir del corismato
- Figura 6. Biosíntesis de tirosina a partir de corismato
- Figura 7. Representación esquemática de las técnicas enzimáticas DNA shuffling e ITCHY
- Figura 8. Diagrama del plásmido pBadmyc-His A-PDH
- Figura 9. Validación de genotipo de variantes seleccionadas
- Figura 10. Validación del fenotipo de las variantes seleccionadas a través de su crecimiento en diferentes medios de cultivo (LB; MM M9+tyr; MM M9+ara; MM M9)
- Figura 11. Cinéticas de crecimiento celular de las 8 variantes en diferentes medios (MM M9+tyr; MM M9+ara; MM M9)
- Figura 12. Expresión del dominio PDH modificado de las 8 variantes bajo inducción por arabinosa y sin inducción
- Figura 13. Normalización de la cantidad de proteína de PDH modificada a través de su comparación con la proteína T silvestre a diferentes concentraciones
- Figura 14. Actividad enzimática específica del dominio PDH modificado de las 8 variantes, a través de medir la aparición de NADH (acorde a la absorbancia determinada a 340 nm).

Índice de tablas

Tabla 1. Clasificación de enzimas con base en el tipo de reacción que realizan

Tabla 2. Sitios de cortes que reconocen las enzimas HindIII y NcoI

Tabla 3. Secuencias de nucleótidos de colonias seleccionadas

Tabla 4. Secuencias de aminoácidos de clona analizadas

Tabla 5. Secuencias de nucleótidos y aminoácidos de las 8 variantes seleccionadas

Lista de abreviaturas

Abreviatura	Significado
<i>E. coli</i>	<i>Escherichia coli</i>
PEP	Fosfoenolpiruvato
E4P	Eritrosa-4-Fosfato
DAHP	3-deoxi-D arabinoheptuloonato-7-fosfato
SHK-3-P	Shikimato-3-fosfato
ATP	Adenosín trifosfato
EPSP	3-P-5-enolpiruvilshikimato
PRPP	Fosforribosilo pirofosfato
CM	Corismato mutasa
PDH	Prefenato deshidrogenasa
epPCR	Reacción en cadena de la polimerasa propensa a error
DNTP	Desoxirribonucleótidos trifosfato
PDT	Prefenato deshidratasa
EDTA	Ácido etilendiaminotetraacético
PCR	Reacción en cadena de la polimerasa
LB	Luria-Bertani
LB+amp+km	LB adicionado con ampicilina a una concentración 200 µg por ml y kanamicina a 25 µg por ml
Medio SOC	Medio SOB (Super Optimal Broth) suplementado con glucosa
MM M9	Medio mínimo M9
MM M9+ara	MM M9 adicionado con arabinosa a una concentración 0.2%
MM M9+tyr	MM M9 adicionado con tirosina a una concentración de 10 mg por ml
NADPH	ml
NAD+	Nicotinamida adenina dinucleótido fosfato
NADH	Nicotinamida adenina dinucleótido
Taq polimerasa	Nicotinamida adenina dinucleótido en su forma reducida
KDa	ADN polimerasa producida por la bacteria <i>Thermophilus aquaticus</i> Kilodalton

Resumen

La proteína T participa en la biosíntesis del aminoácido aromático tirosina a partir de corismato en la bacteria *Escherichia coli*; dicha proteína es codificada por el gen *TryA*, es homodímera y consta de 373 residuos de aminoácidos por monómero. Cada monómero tiene una masa molecular de 42 kDa. Esta enzima consta de dos dominios, el de corismato mutasa (CM) y el de prefenato deshidrogenasa (PDH). El dominio CM cataliza la conversión de corismato a prefenato por un rearrreglo de Claisen, el segundo dominio, PDH, cataliza la conversión de prefenato a 4-hidroxifenilpiruvato, mediante una descarboxilación oxidativa utilizando NAD^+ como cofactor, posteriormente por medio de una transaminación, catalizada por otro enzima, el 4-hidroxifenilpiruvato se convierte en tirosina, que es un aminoácido aromático esencial para los organismos.

Experimentos anteriores han demostrado que al expresar los dominios por separado, el dominio de CM presenta actividad enzimática y se expresa; no obstante, estos resultados no se observan en el dominio de PDH. Sin embargo, al sustituir el dominio de CM en la región amino terminal del dominio de PDH por dominios capaces de dimerizar se ha observado que hay expresión y actividad enzimática, lo cual indica que la dimerización ayuda a la estabilidad del dominio de PDH.

En el presente proyecto se añadió una etiqueta de 6 aminoácidos al azar en la región amino terminal del dominio PDH, ya que se ha demostrado que la longitud de esta etiqueta es suficiente para observar actividad enzimática específica. Para esto se llevó a cabo el diseño de un oligonucleótido combinando las técnicas de evolución dirigida e ingeniería de proteínas (diseño racional) con el fin de generar diversidad a la secuencia de 6 aminoácidos. Posteriormente se construyó y se validó una librería de variantes. Utilizando esta librería se transformaron células de *E. coli* de la cepa BW25113 carentes del gen *tyrA*, seleccionando aquellas que fueran capaces

de crecer en un medio mínimo, validando su genotipo y su fenotipo, se obtuvieron en total 8 variantes.

Se realizó la caracterización de las 8 variantes las cuales mostraron un nivel de actividad enzimática y expresión de la proteína semejante a la observada por parte de la proteína T silvestre. Con dicho comportamiento podemos decir que una etiqueta de 6 aminoácidos es suficiente para que el dominio de PDH muestre expresión y actividad enzimática, y por tanto, capaz de transformar el prefenato en 4-hidroxifenilpiruvato para la síntesis de tirosina.

1. Introducción

1.1. Proteínas: Definición y estructura

Desde un punto de vista químico, las proteínas son las moléculas estructuralmente más complejas y funcionalmente más sofisticadas que se conocen dentro de las células. Una proteína está formada por una cadena de residuos de aminoácidos, unidos mediante un enlace peptídico; por ello, también se les conoce como polipéptidos (Horton et al., 2008). El enlace peptídico se forma por la unión covalente del carbono del grupo carboxilo con el nitrógeno del grupo amino de otro aminoácido; esta reacción sucede en el ribosoma. La síntesis es direccional, del amino terminal (N-terminal) hacia el carboxilo terminal (C-terminal). Al formarse el enlace peptídico se libera una molécula de agua y por esta razón se llaman residuos de aminoácidos a los bloques que la constituyen, ya que no está completo el aminoácido (Figura 1). Cada tipo de proteína tiene una secuencia única de residuos de aminoácidos, y hay miles de proteínas diferentes en una célula (Alberts et al., 2010), cada una con una función específica. Existen 20 aminoácidos que conforman las secuencias de los polipéptidos, cada uno posee propiedades fisicoquímicas distintas. La estructura química general de los aminoácidos consta de un carbono central llamado α , a este carbono se le une un grupo amino y uno carboxilo. Además de un hidrógeno y una cadena lateral (Figura 2). Esta última es la que define las propiedades fisicoquímicas del aminoácido (Watson et al., 2014). Debido a que sus sustituyentes son distintos, el carbono α es un centro quiral, la forma L es la que generalmente está presente en las proteínas. Podemos dividir a los aminoácidos con base a las propiedades químicas de sus cadenas laterales en las siguientes categorías; apolares, polares sin carga y cargados positiva o negativamente (Alberts et al., 2010).

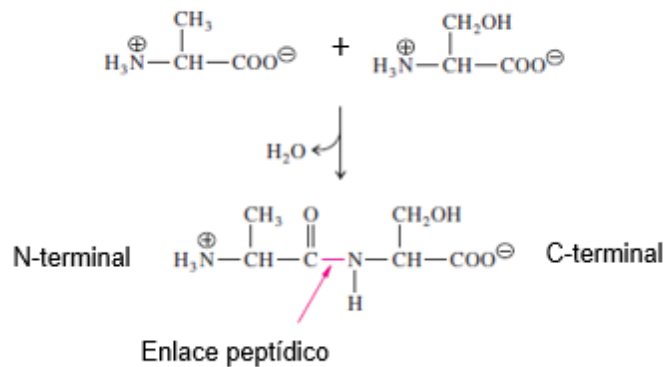


Figura 1. Representación de la formación del enlace peptídico. El enlace se forma por la unión covalente del carbono del grupo carboxilo con el nitrógeno del grupo amino de otro aminoácido, durante este proceso se libera una molécula de agua (Imagen tomada de Horton, 2008).

Las cadenas laterales apolares (hidrófobas) de una proteína tienden a ubicarse en el interior de la molécula, esto evita el contacto con el agua. Esta propiedad ayuda a que las proteínas se plieguen. Por el contrario, los grupos polares tienden a disponerse en la superficie de la molécula donde pueden formar enlaces de hidrógeno con el agua y otras moléculas polares (Alberts et al., 2010).

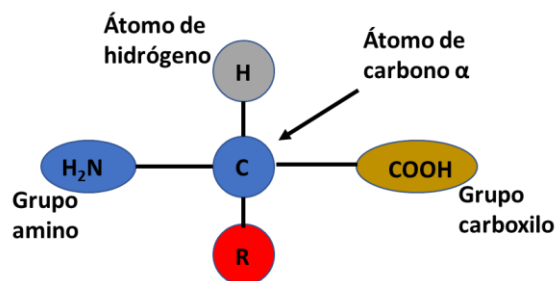


Figura 2. Estructura química general de un aminoácido. Cada aminoácido consta de un carbono central llamado α , a este carbono se le une un grupo amino y uno carboxilo. Además de un hidrógeno y una cadena lateral (Imagen tomada de Horton, 2008).

Se distinguen cuatro niveles de organización en la estructura de una proteína (Figura 3). La secuencia lineal de residuos de aminoácidos en el polipéptido es la estructura primaria y se lee del extremo amino al carboxilo. Los segmentos de una

cadena polipeptídica pueden formar hélices α , láminas β o estructura no periódica constituyen la estructura secundaria. La conformación tridimensional de la cadena polipeptídica, se denomina estructura terciaria, y si la proteína está constituida por más de una cadena polipeptídica, se denomina estructura cuaternaria (Alberts et al., 2010).

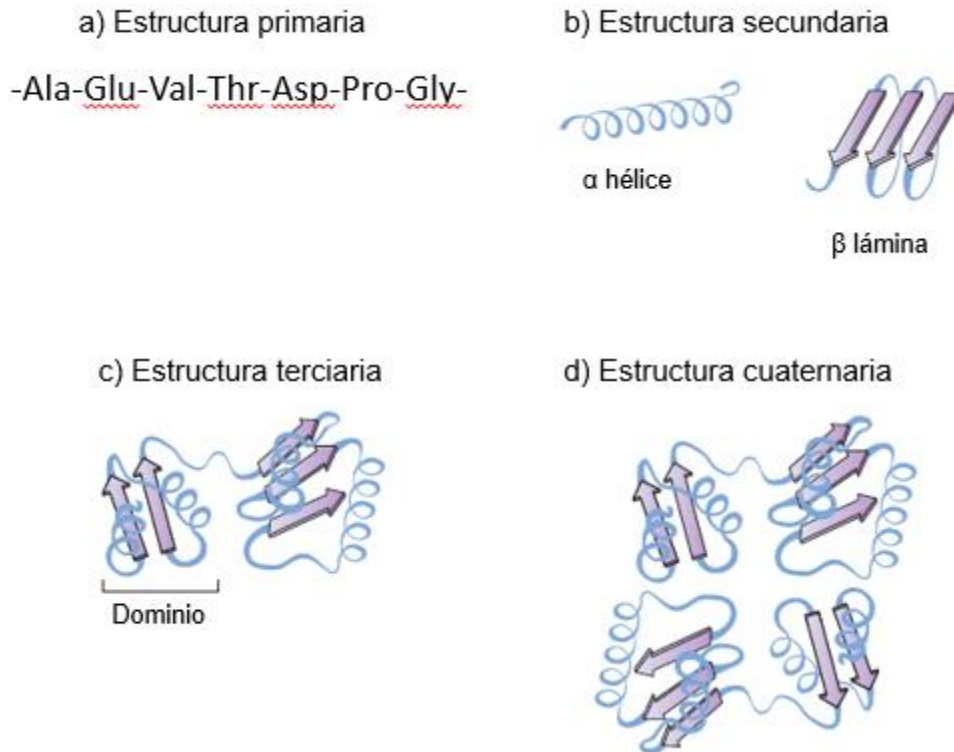


Figura 3. Niveles de organización de la estructura de la proteína. a) Primaria (secuencia lineal de residuos de aminoácidos), b) Secundaria (segmentos de una cadena polipeptídica que forman hélices α , láminas β o estructura no periódica), c) Terciaria (conformación tridimensional de la cadena polipeptídica) y d) Cuaternaria (constituida por más de una cadena polipeptídica) (Imagen tomada de Horton, 2008).

1.1.1. Dominios de las proteínas.

Estudios sobre la conformación, la función y la evolución de las proteínas también han revelado la importancia de un nivel de organización distinta a los cuatro niveles anteriormente mencionados. El análisis de digestión de proteínas con proteasas mostró la existencia de fragmentos que eran resistentes a la degradación y que presentaban una estructura estable. El análisis de la estructura tridimensional de las

proteínas completas y los fragmentos mostraron que estos presentaban una estructura compacta. A estos fragmentos se les llamó dominios y su característica principal es que pueden plegarse independientemente del resto de la proteína, dando lugar a una estructura compacta y estable (Alberts et al., 2010).

Un dominio generalmente contiene de 40 a 350 residuos de aminoácidos y pueden unirse varios dominios para construir proteínas de mayor tamaño (Alberts et al., 2010). Los diferentes dominios de una proteína están asociados con diferentes funciones. Se ha propuesto que esto ha permitido la diversificación de las funciones de las proteínas (Soberón & Montero, 2006).

1.1.2. Tipos de proteínas.

Las proteínas son las moléculas que realizan la mayoría de las funciones en las células, estas son catalíticas, estructurales o regulatorias. El mantenimiento de una célula requiere el trabajo coordinado de diferentes tipos de proteínas (Nelson & Cox, 2008). Entre ellas se encuentran las proteínas estructurales, que proveen rigidez a la célula; proteínas transportadoras, que controlan el flujo de materiales a través de las membranas celulares; proteínas reguladoras, que actúan como sensores e interruptores para controlar la actividad de las proteínas y los genes; proteínas señalizadoras, incluidas las receptoras de la superficie celular y otras proteínas que transmiten señales externas al interior de las células; y las proteínas motoras, que producen el movimiento (Lodish et al., 2005).

Se ha postulado que la gran diversidad de proteínas actuales se ha generado a partir de un número pequeño de moléculas rudimentarias que realizaban reacciones químicas con una baja eficiencia. Por un proceso de divergencia gradual de la secuencia, las proteínas primigenias divergieron aumentando su actividad o generando una amplia gama de nuevas actividades (Soberón & Montero, 2006).

1.1.3. Enzimas.

Existe una gran clase de proteínas muy importantes llamada enzimas (Alberts et al., 2010). Las enzimas son catalizadores biológicos selectivos con una eficiencia extraordinaria. Toda célula viva dispone de cientos de distintas enzimas que catalizan las reacciones de las rutas metabólicas que son necesarias para mantener la vida. El papel principal de las enzimas es aumentar la velocidad de las reacciones. En forma típica, las reacciones catalizadas por las enzimas son de 10^3 a 10^{10} más rápidas que sin ellas. Un catalizador es una molécula que acelera la reacción para alcanzar el equilibrio más rápidamente. Un catalizador no se consume permitiendo que participe cíclicamente en la reacción. Los reactivos se unen a un catalizador y los productos se disocian de él (Horton et al., 2008).

Un catalizador no cambia la posición de equilibrio de la reacción (es decir, no hace que una reacción no favorable sea favorable). Más bien reduce la cantidad de energía necesaria para que se efectúe la reacción, es decir aceleran la reacción. Los catalizadores aceleran las reacciones en ambas direcciones (Horton et al., 2008).

Las enzimas son muy específicas para los sustratos sobre los que actúan, pero pueden mostrar actividad con compuestos relacionados. Esta promiscuidad ha sido propuesta como una pieza fundamental en la evolución de las proteínas ya que esta es una fuente de nuevas funciones (Patrick et al., 2007).

Las enzimas pueden formar complejos proteicos, donde el producto de una enzima es el sustrato de otra enzima del complejo. Esto permite que la segunda enzima obtenga su sustrato más rápidamente y la vía biosintética sea más eficiente. Las reacciones acopladas son una propiedad común de muchas vías metabólicas (Alberts et al., 2010).

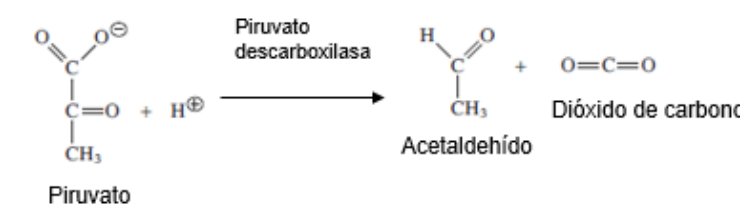


Algunas enzimas funcionan como puntos de control en el metabolismo, la regulación de su actividad permite controlar la producción de metabolitos y sólo producirse cuando son requeridos (Nelson & Cox, 2008).

Las enzimas cuya actividad es regulada, tienen en general, una estructura más compleja que las enzimas no reguladas. Las enzimas reguladas pueden ser moléculas oligoméricas que tienen sitios de enlace separados para sustratos y moduladores, compuestos que actúan como señales de regulación. El hecho de que la actividad enzimática pueda ser regulada es una propiedad importante que distingue a los catalizadores biológicos de los que se realizan en un laboratorio de química. Cuando existe un sitio de regulación que no corresponde al catalítico se habla de enzimas alostéricas (Horton et al., 2008).

Se muestra la clasificación de las enzimas de acuerdo con el tipo de reacción catalítica que realizan en la tabla I

Tabla I. Clasificación de enzimas con base en el tipo de reacción que realizan.

Nombre	Tipo de reacción que llevan a cabo	Ejemplo de enzimas	Ejemplo de reacción
Oxidorreductasas	Catalizan reacciones en las que hay transferencia de electrones	Deshidrogenasas, oxidasas, peroxidasa, oxigenasas o reductasas	$ \begin{array}{ccc} \begin{array}{c} \text{COO}^{\ominus} \\ \\ \text{HO}-\text{C}-\text{H} \\ \\ \text{CH}_3 \end{array} + \text{NAD}^{\oplus} & \xrightleftharpoons{\text{Lactato deshidrogenasa}} & \begin{array}{c} \text{COO}^{\ominus} \\ \\ \text{C}=\text{O} \\ \\ \text{CH}_3 \end{array} + \text{NADH} + \text{H}^{\oplus} \\ \text{L-Lactato} & & \text{Piruvato} \end{array} $
Transferasas	Catalizan las reacciones de transferencia de un grupo y pueden necesitar la presencia de coenzimas.	Transcarboxilasas, transmetilas Transamina	$ \begin{array}{ccc} \begin{array}{c} \text{COO}^{\ominus} \\ \\ \text{H}_3\text{N}^{\oplus}-\text{C}-\text{H} \\ \\ \text{CH}_3 \end{array} + \begin{array}{c} \text{COO}^{\ominus} \\ \\ \text{C}=\text{O} \\ \\ (\text{CH}_2)_2 \\ \\ \text{COO}^{\ominus} \end{array} & \xrightleftharpoons{\text{Alanina transaminasa}} & \begin{array}{c} \text{COO}^{\ominus} \\ \\ \text{C}=\text{O} \\ \\ \text{CH}_3 \end{array} + \begin{array}{c} \text{COO}^{\ominus} \\ \\ \text{H}_3\text{N}^{\oplus}-\text{C}-\text{H} \\ \\ (\text{CH}_2)_2 \\ \\ \text{COO}^{\ominus} \end{array} \\ \text{L-Alanina} & & \text{Piruvato} \quad \quad \quad \text{L-Glutamato} \\ & & \alpha\text{-Cetoglutarato} \end{array} $
Hidrolasas	Catalizan reacciones en las que se produce la ruptura de enlaces como C-O, C-N y O-P por la adición de agua	Esterasas, fosfatasas peptidasas	$ \begin{array}{ccc} \begin{array}{c} \text{O} \quad \quad \text{O} \\ \quad \quad \\ \ominus\text{O}-\text{P}-\text{O}-\text{P}-\text{O}^{\ominus} \\ \quad \quad \\ \ominus\text{O} \quad \quad \text{O}^{\ominus} \end{array} + \text{H}_2\text{O} & \xrightarrow{\text{Pirofosfatasa}} & 2 \begin{array}{c} \text{O} \\ \\ \text{HO}-\text{P}-\text{O}^{\ominus} \\ \\ \text{O}^{\ominus} \end{array} \\ \text{Pirofosfato} & & \text{Fosfato} \end{array} $

Liasas	Catalizan reacciones de adición a dobles enlaces (por lo tanto los transforman en enlaces sencillos), o bien forman dobles enlaces.	Descarboxilasas, hidratasas, deshidratasas, desaminasas y sintetasa.	 <p>Piruvato Acetaldehído Dióxido de carbono</p>
Isomerasas	Catalizan reacciones cuyo resultado es una ordenación intramolecular	Epimerasas, mutasas	 <p>L-Alanina D-Alanina</p>
Ligasas	Catalizan la ligadura o unión de dos sustratos. Estas reacciones necesitan un suministro de energía química de un nucleósido trifosfato, como el ATP.	Carboxilasas, sintetasa	 <p>L-Glutamato L-Glutamina</p>

1.2. Biosíntesis de aminoácidos aromáticos.

Los aminoácidos aromáticos poseen considerable interés industrial, particularmente por su uso como productos farmacéuticos, aditivos para alimentos, suplementos dietéticos y nutracéuticos. Su valor nutritivo se comprende ya que son esenciales en la dieta humana, al ser incapaces de sintetizarlos su ingesta es necesaria para los requerimientos de las células (Bongaerts et al., 2001).

La ruta de biosíntesis de los aminoácidos aromáticos, así como la regulación en su producción en *E. coli* y otras bacterias es bien conocida y ha sido sometida a diversos enfoques de ingeniería metabólica en los últimos años para mejorar la producción de ellos. En contraste con L-fenilalanina y L-triptófano, la producción biotecnológica de L-tirosina ha recibido mucho menos atención, posiblemente debido a que la demanda de este en el mercado es significativamente más baja (Lutke-Eversloh & Stephanopoulos, 2007).

1.2.1. La vía de shikimato.

La vía de Shikimato (Figura 4) es la ruta metabólica por la cual se sintetiza el precursor común para la síntesis de aminoácidos aromáticos y compuestos relacionados. La vía se encuentra presente en plantas, hongos y bacterias pero no en mamíferos (Cortés-Totalpa et al., 2011).

En *E. coli* los aminoácidos aromáticos (fenilalanina, tirosina y triptófano) son sintetizados a partir de fosfoenolpiruvato (PEP) y eritrosa-4-fostato (E4P), los cuales son producidos en la glucólisis y en la vía de las pentosa fosfato respectivamente (Nelson & Cox, 2008). La condensación inicial de PEP y E4P es catalizada por la enzima 3-deoxi-D-arabinoheptulosonato-7-fosfato (DAHP) sintasa para formar DAHP, el primer intermediario de la vía (Kedar et al., 2007). Esta enzima ha sido ampliamente estudiada: posee 3 isoenzimas AroF, AroG y AroH, codificadas respectivamente por los genes *aroF*, *aroG* y *aroH*. La presencia de las tres isoenzimas provee a *E. coli* la capacidad de un control estricto del primer paso de la vía de síntesis de los aminoácidos aromáticos, permitiendo a la vez una actividad enzimática residual en presencia de un exceso de aminoácidos aromáticos que

provee a la célula de precursores para la síntesis de otros compuestos aromáticos. Aunque estas tres isoenzimas catalizan la misma reacción, cada una de ellas es inhibida por los aminoácidos aromáticos tirosina, fenilalanina y triptófano respectivamente (Cortés-Totalpa et al., 2011).

La segunda reacción de la vía es la transformación del DAHP a 3-dehidroquinato (DHQ) mediante una deshidratación, la reacción es catalizada por la enzima dehidroquinato sintasa codificada por *aroB*. Para llevar a cabo la reacción, la enzima utiliza como cofactor a NAD⁺. El DHQ es reducido para formar shikimato (Pittard & Yang, 2013). La reacción es catalizada por la enzima shikimato deshidrogenasa codificada por *aroE* y requiere de NADPH como cofactor. En *E. coli* el shikimato es convertido a shikimato-3-fosfato (SHK-3-P) por acción de las isoenzimas shikimato cinasa I y II, codificadas respectivamente por *aroK* y *aroL*. Esta reacción requiere de ATP. El penúltimo paso consiste en la adición de una molécula de PEP a SHK-3-P para la obtención de 3-P-5-enolpiruvilshikimato (EPSP), reacción catalizada por la enzima EPSP sintasa, codificada por *aroA*. La última reacción de la vía común en la síntesis de los 3 aminoácidos aromáticos es catalizada por la enzima corismato sintasa, codificada por *aroC*, la cual añade un segundo doble enlace al anillo mediante la eliminación del fósforo de EPSP dando corismato como producto final (Cortés-Totalpa et al., 2011).

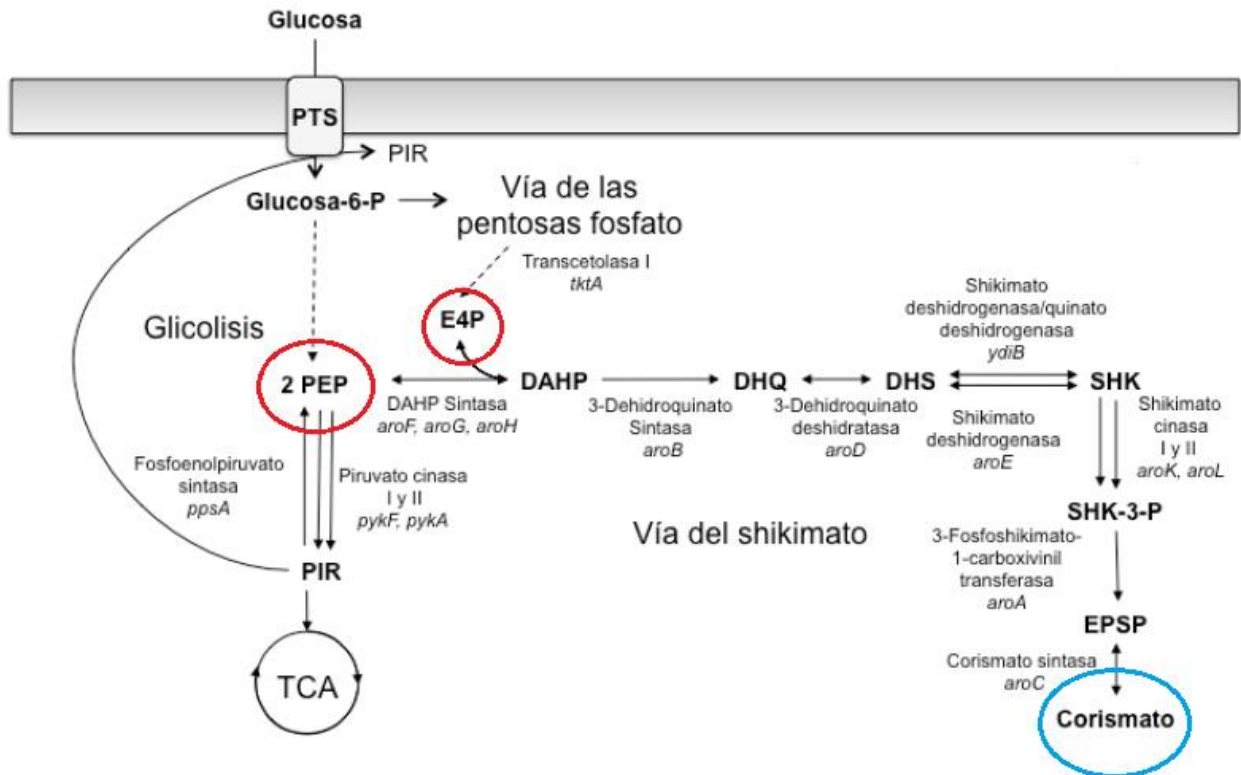


Figura 4. Pasos en la vía del shikimato. El fosfoenilpiruvato (PEP) y eritrosa-4-fostato (E4P), señalados de color rojo, son producidos en la glucólisis y en la vía de las pentosas fosfato respectivamente para formar DAHP, el primer intermediario de la vía. Finalmente se obtiene corismato, señalado de color azul, como precursor común para la síntesis de los aminoácidos aromáticos fenilalanina, triptófano y tirosina (Imagen tomada de Cortés-Totalpa, 2011).

El corismato es un intermediario central de la ruta de síntesis de los aminoácidos aromáticos y metabolitos derivados de ellos, como el shikimato que es el punto de ramificación de cinco rutas metabólicas en microorganismos (Lutke-Eversloh & Stephanopoulos, 2007).

1.2.2. Biosíntesis de fenilalanina.

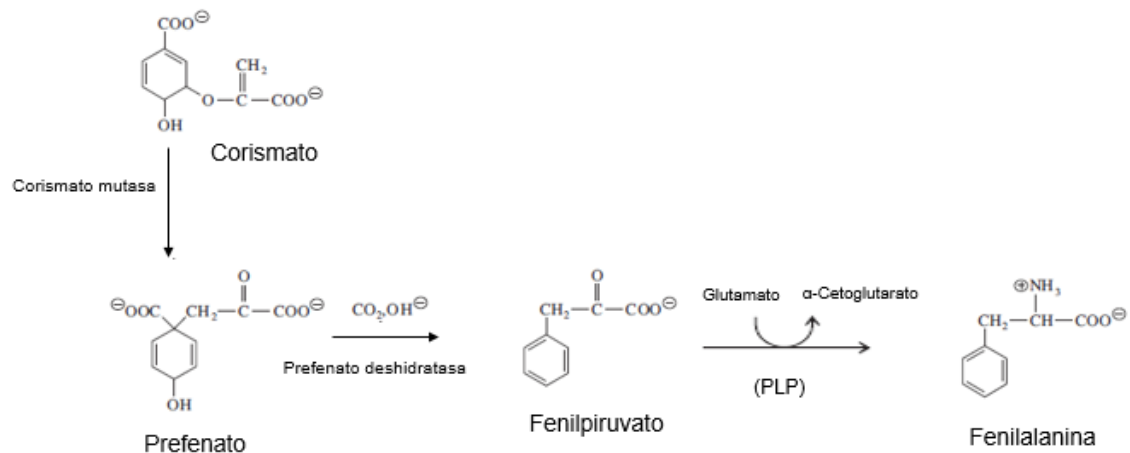
En la síntesis de fenilalanina en *E. coli* (Figura 5a), la proteína P, una enzima bifuncional compuesta de dos dominios el de corimato mutasa y el de preferato deshidratasa, codificada por el gen *pheA*, cataliza el rearrreglo de corismato para producir preferato, compuesto altamente reactivo. A continuación, la enzima cataliza la eliminación de un ion hidróxido y CO₂ del preferato para formar el fenilpiruvato, compuesto aromático, seguido de una transaminación catalizada por

una aminotransferasa (codificada por el gen *tyrB*) para producir la fenilalanina (Horton et al., 2008).

1.2.2. Biosíntesis de triptófano.

En la biosíntesis del triptófano a partir del corismato, el primer paso consiste en que el nitrógeno amida de la glutamina se transfiera al corismato; la eliminación del grupo hidroxilo y la mitad adyacente de piruvato en el corismato, produce antranilato, compuesto aromático, esta reacción es catalizada por la enzima antranilato sintasa. En la segunda reacción, catalizada por la enzima antranilato fosforribosiltransferasa, el antranilato acepta el fosforribosilo del grupo fosforribosilo pirofosfato (PRPP) para formar N-(5'fosforribosilo)-antranilato. Por rearrreglo de la ribosa, descarboxilación y cierre de anillo se genera el fosfato de indol y glicerol (indol glicerol fosfato). Las dos reacciones finales de la biosíntesis del triptófano, la conversión de indol-glicerol fosfato a triptófano, es catalizada por la triptófano sintasa (Horton et al., 2008) (Figura 5b).

a)



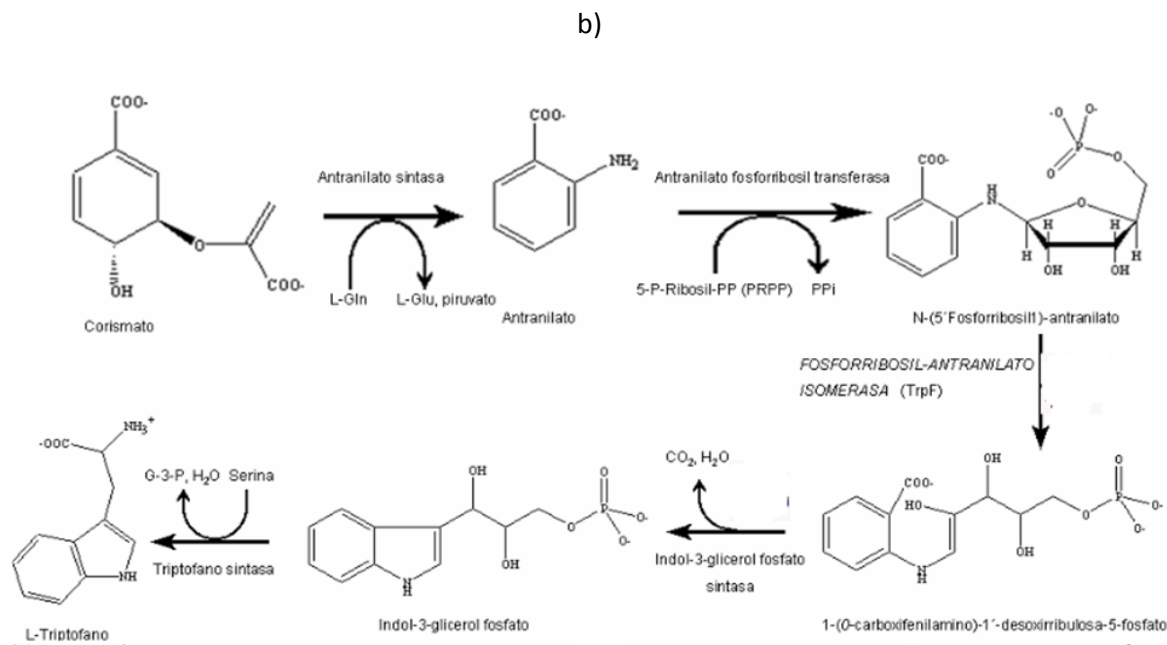


Figura 5. Biosíntesis de a) fenilalanina y b) triptofano a partir del corismato. (imagen tomada de Horton, 2008 y <http://148.206.53.84/tesiuami/UAMI13404.pdf>).

1.2.3. Biosíntesis de tirosina.

La biosíntesis de tirosina es necesaria para la supervivencia y crecimiento de todos los organismos, sin embargo, sólo las bacterias entéricas, levaduras, hongos y plantas la sintetizan. Como muchos otros aminoácidos aromáticos, la tirosina juega un doble papel en la bioquímica del organismo, actuando como producto y como precursor de otros compuestos. En el primer caso, la tirosina es necesaria para la síntesis de proteínas, mientras que en el segundo caso, es sustrato para la producción de otros metabolitos. Los metabolitos aromáticos derivados de la tirosina incluyen quinonas, glucósidos cianogénicos, alcaloides, flavonoides y compuestos fenólicos derivados de la ruta fenilpropanoide. Dado que muchos de estos compuestos están implicados en procesos biológicos primarios, son esenciales para la viabilidad (Sun et al., 2006).

En la bacteria *E. coli* la conversión de corismato a prefenato y de prefenato a 4-hidroxifenilpiruvato en la vía de síntesis de tirosina, es catalizada por la proteína T. Esta enzima es codificada por el gen *TryA*, es homodímera y consta de 373

residuos de aminoácidos por monómero. Cada monómero tiene una masa molecular de 42 kDaa. Esta consta de dos dominios; el de corismato mutasa (CM) en la región amino terminal y que de acuerdo con Chen y colaboradores (Chen et.al., 2003) tiene una longitud de 93 residuos de aminoácidos y el de prefenato deshidrogenasa (PDH) localizado en la región carboxilo terminal que consta de 280 residuos de aminoácidos. En la primera reacción el dominio de corismato mutasa lleva a cabo un rearrreglo de Claisen sobre corismato para formar prefenato, el segundo dominio cataliza la conversión de prefenato a 4-hidroxifenilpiruvato, mediante una descarboxilación oxidativa utilizando NAD^+ como cofactor. Finalmente, por medio de una transaminación el 4-hidroxifenilpiruvato se convierte en tirosina. La regulación de la actividad de ambos dominios se da por inhibición por tirosina (Figura 6).

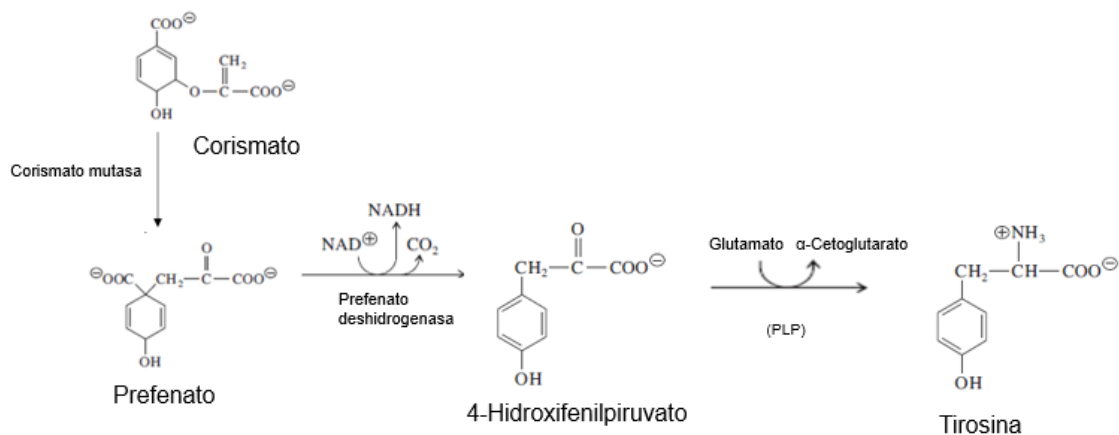


Figura 6. Biosíntesis de tirosina a partir de corismato. En la primera reacción el corismato sufre un rearrreglo de Claisen para formar prefenato, éste se convierte en 4-hidroxifenilpiruvato y finalmente mediante una transaminación obtenemos tirosina (Imagen tomada de Horton, 2008).

1.3. Evolución de las proteínas

Actualmente se acepta que las proteínas se crean por 2 procesos: *de Novo* y por divergencia de proteínas ya existentes. En el primer proceso, una secuencia *de novo* de ADN se expresa y genera una secuencia peptídica estable y con una función. El segundo proceso involucra la diversificación de la secuencia de proteínas ya existentes (Soberón & Segovia, 2007). El primer proceso se dio principalmente

en el origen de las primeras proteínas y después de este evento primigenio se debe de haber dado poco. El segundo evento involucra la generación de cambios en la secuencia de las proteínas que pueden alterar la función de la misma. Esto origina que la proteína pueda tener 2 funciones y que cuando exista una ventaja sobre la nueva actividad, el gen que codifica la proteína se duplique. La duplicación permite que la nueva función se incremente sin afectar la función original, cada copia se especializa en una función. Sabemos que esto ha ocurrido muy a menudo, porque las proteínas actuales pueden agruparse en familias proteicas que presentan secuencias y estructuras tridimensionales parecidas pero con distintas funciones (Soberón & Montero, 2006).

Por otra parte, el análisis de estructuras tridimensionales muestra la existencia de proteínas con secuencias muy distintas, lo que indicaría un origen distinto, pero con la estructura muy parecida lo que indica un origen común y diversificación (Alberts et al., 2010). Se acepta que tienen un origen común, pero que la secuencia ha divergido mucho, esto apoya que la divergencia de secuencia origina proteínas con nuevas funciones.

Puesto que la generación de las mutaciones que originan un cambio de aminoácido es un proceso aleatorio, estos pueden tener distintos efectos de acuerdo con el tipo de cambio y su localización en la estructura de la proteína. Por lo anterior, un cambio de un aminoácido puede o no alterar la función de una proteína. Cuando no produce un efecto en la función se habla que son neutrales. Esto, sin embargo, no significa que alguna propiedad de la proteína no se haya alterado, pueden existir actividades que no están bajo presión de selección y que los investigadores no estudien, que puedan alterarse. Esto es importante ya que pueden en algún momento ser importantes para generar una nueva función. Los cambios que modifican las propiedades de la proteína pueden ser benéficos o deletéreos. Es decir aumentar o disminuir la actividad, estabilidad, solubilidad, etc .

1.3.1. Estrategias para la evolución de proteínas: ingeniería de proteínas y evolución dirigida.

La alta complejidad de las proteínas permite que en ocasiones un solo cambio de residuo en su estructura primaria afecta de manera importante su función o estructura (Nelson & Cox, 2008). El cambio puede originar que la función se altera, se pierda o que la estructura sea menos o más estable. Para poder comprender cómo los cambios afectan a las proteínas se han utilizados 2 estrategias: la ingeniería de proteínas y la evolución dirigida (Soberón & Segovia, 2007).

El primer enfoque, la ingeniería de proteínas, se basa en predecir qué cambio es necesario para poder generar una propiedad deseada. El cambio diseñado se introduce en el gen por mutagénesis sitio dirigida (Bornscheuer & Pohl, 2001). Para este enfoque es necesario tener un conocimiento amplio de la proteína, el cual muchas veces no se tiene. Actualmente no se tiene el suficiente conocimiento y poder computacional para poder predecir con total certeza el efecto que ocasionará un cambio en la proteína o cómo generar una propiedad deseada.

De acuerdo con Soberón (Soberón & Montero, 2006), en esencia, un investigador analiza la información que dispone de una proteína de interés, desde su secuencia de aminoácidos, la secuencia de otras proteínas relacionadas evolutivamente, su función, las condiciones en que funciona, su estructura tridimensional, la identidad y posición en la secuencia de los residuos de aminoácidos que juegan un papel importante en la función de la misma. Con esta información se generan hipótesis específicas sobre el papel que puede jugar un cambio de un residuo en la proteína; se introduce el cambio diseñado en el gen que codifica para la proteína y se expresa (Típicamente en células bacterianas) y se purifica el producto para analizar su comportamiento. Con base en los resultados obtenidos se generan nuevas hipótesis y se repite el ciclo. Un prerrequisito para realizar ingeniería de proteínas es que la enzima ya haya sido clonada, expresada y se conozca su estructura tridimensional. Generalmente es necesario realizar varios ciclos de diseño para lograr el objetivo, se realiza el cambio, se evalúa y se realiza otro hasta lograr el resultado deseado.

El segundo enfoque, evolución dirigida, tiene un concepto sencillo, que implica imitar al proceso evolutivo natural que ha dado origen a las proteínas que existente hoy. La evolución opera sobre dos componentes fundamentales: la variación genética y la selección de los individuos mejor adaptados (Segovia & Peimbert, 2010). La variabilidad genética es el motor de la evolución. La selección natural, por su parte, ejerce un efecto regulador al seleccionar los individuos mejor adaptados a las condiciones del medio y por lo tanto que puedan generar más descendientes que perpetúen su carga genética (Soberón & Segovia, 2007).

Para comprender este enfoque debe tenerse en cuenta que el proceso natural se ha dado en el transcurso de miles de millones de año y operando sobre una cantidad grande de organismos. En el laboratorio, se utilizan organismos que tienen un ciclo de vida corto, esto permite que se reduzca el tiempo y solo se requieran semanas para generar variantes con la propiedad deseada.

1.3.2. Técnicas para la evolución dirigida.

La evolución dirigida de proteínas *in vitro* se inició con el trabajo pionero de Spiegelman en 1965 (Spiegelman et al., 1965), quien demostró que se podría reproducir *in vitro* el proceso darwiniano de la evolución. El experimento consistió en sintetizar *in vitro* una molécula de ARN, seleccionando aquellas que se amplifican más rápidamente. Se purificó la ARN polimerasa del bacteriófago Q β , el genoma de este organismo tiene un tamaño de 4,217 nucleótidos y codifica para 4 proteínas e infecta a la bacteria *E. coli*. Esta enzima junto con un molde de ARN, ribonucleótidos trifosfatados y Mg²⁺ es capaz de sintetizar el genoma de ARN del virus con las mismas características infectivas que el original (Spiegelman et al., 1965). Esta idea, que originalmente nace en un esquema de investigación básica para tratar de entender el origen del proceso evolutivo, fue posteriormente explotada para generar, de manera acelerada, proteínas con propiedades modificadas para su uso en la industria (Soberón & Montero, 2006).

En el proceso natural de la evolución, el material genético sufre mutaciones, aleatorias, a través de diversos mecanismos. En el caso de la evolución *in vivo*, como lo describen Michael S. Packer y David R, revisado en Liu (Packer & Liu, 2015)

es factible emplear métodos tradicionales en los cuales se utilizan agentes físicos y químicos que dañan azarosamente el ADN, como por ejemplo el metanosulfonato de etilo, bases análogas como la 2-aminopurina e irradiaciones ultravioleta. Otros agentes utilizados en la evolución *in vitro* consisten en aumentar el rango de error durante la replicación del ADN, esto mediante métodos enzimáticos que ayudan a introducir cambios durante la síntesis de fragmentos de ADN (Packer & Liu, 2015).

En la actualidad, los métodos enzimáticos son los más utilizados. Esto se debe a la mayor diversidad y eficiencia que permiten las enzimas además de que se pueden introducir los cambios solo en zonas predeterminadas y controlar su tasa de introducción (Soberón & Montero, 2006). Un ejemplo de ello es la técnica de PCR propensa a error (epPCR), la baja fidelidad de la ADN polimerasa en ciertas condiciones genera que introduzca cambios puntales durante la amplificación del gen de interés. Aumento en la concentración de magnesio, introducción de manganeso, el uso de desoxirribonucleótidos trifosfato (DNTP) en proporciones desbalanceadas o análogos que generan cambios, pueden reducir la fidelidad de la replicación del ADN e incrementa el índice de mutación a $10^{-2} \sim 10^{-3}$ por base replicada (Packer & Liu, 2015).

Existen diversas técnicas enzimáticas para introducir los cambios, una de ellas, el DNA shuffling o entremezclado de ADN (Figura 7), recombina los cambios presentes en los genes de los miembros de una familia de proteínas y de esta manera lograr la propiedad deseada más rápidamente. Este método consiste en fragmentar con DNAsa una población de genes relacionados. Los fragmentos de ADN generados de diferentes longitudes se utilizan para reensamblar el gen. La mezcla de fragmentos de distintos genes para armar el gen completo genera que el producto final sea una mezcla de partes de los genes originales, lo cual es la recombinación (Ostermeier et al., 1999). Esta técnica permite tener cambios que se han seleccionado de manera natural en distintos genes y que al estar junto se codifique una proteína con mejores propiedades. Una variante de este método es la estrategia de truncamiento por prueba y error para la creación de enzimas híbridas o ITCHY (Figura 7) para la recombinación de genes con poco o baja identidad. Se

basa en la generación de librerías de fragmentos de ADN que codifican el amino o carboxilo terminal de dos genes por eliminación progresiva de las secuencias con la enzima exonucleasa III seguida por la ligación de los productos para generar librería híbrida (Soberón & Segovia, 2007).

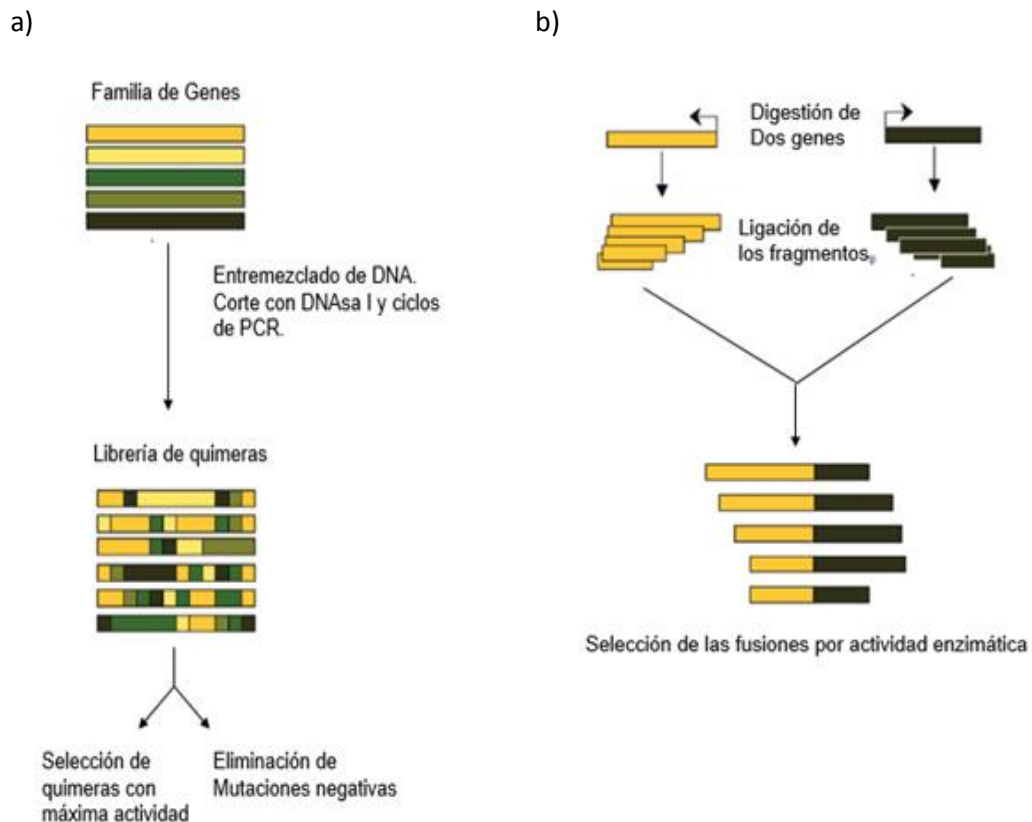


Figura 7. Representación esquemática de las técnicas enzimáticas DNA shuffling e ITCHY. a) Técnica DNA shuffling o entremezclado de genes, utilizada para una familia de genes homólogos. b) Técnica de truncamiento por prueba y error para la creación de enzimas híbridas o ITCHY. (Imagen tomada de Soberón y Montero, 2006).

La principal ventaja de la evolución dirigida es que no es necesario un conocimiento profundo *a priori* de la proteína de interés (Bornscheuer & Pohl, 2001). Sin embargo, cabe resaltar que la principal ventaja de cualquier enfoque evolutivo experimental es que se acorta de manera importante el tiempo necesario para obtener la

propiedad deseada. Sin embargo, la principal diferencia con la evolución natural es que tiene una población de variantes muy pequeña y la selección es planeada.

Por eso los métodos que se utilizan tratan de mejorar la generación de diversidad, sin embargo el principal problema es el limitado número de variantes que se pueden explorar.

2. Antecedentes.

Varias bacterias y algunas plantas utilizan las mismas rutas de síntesis de fenilalanina y tirosina a partir de corismato. Sin embargo, en estas, a diferencia de *E. coli*, las actividades de corismato mutasa y prefenato deshidratasa o deshidrogenasa están en proteínas separadas. Existen otras bacterias que usan rutas diferentes, donde primero se transamina el prefenato y después se descarboxilan (Horton et al., 2008).

La proteína P, es una proteína bifuncional homóloga a la proteína T de *E. coli* presenta el dominio de corismato mutasa (CM) y el dominio prefenato deshidratasa (PDT), el primer dominio cataliza la transformación de corismato a prefenato y a continuación el dominio de PDT convierte el prefenato a fenilpiruvato en la ruta de la biosíntesis de fenilalanina. Ambas proteínas (P y T) presentan numerosas similitudes estructurales y cinéticas. Sin embargo, existe evidencia de que la principal diferencia entre la proteína T y P es la relación espacial y funcional entre las dos actividades catalíticas en cada enzima (Osuna et. al., 2012). Estudios anteriores de mapeos de dominios establecieron que las actividades de CM, PDT y regulatorias residen en dominios discretos y separables que pueden ser clonados y expresados de manera independiente. Sin embargo estudios bioquímicos dirigidos al mapeo de los dominios funcionales en la proteína T, sugieren una relación espacial más compleja de los dominios. Alineamientos de secuencias primarias entre la proteína T y P indican que muestran aproximadamente un 25% de similitud (Chen et al., 2003).

Se ha demostrado que la expresión por separado de los dominios de Corismato Mutasa (CM) y Prefenato Deshidrogenasa (PDH) de la proteína T de *E. coli*, no genera proteínas estables. En la expresión del primer dominio se observa expresión y actividad enzimática. Sin embargo, la expresión del dominio de PDH es baja y no se detecta actividad (Chen et al., 2003). Varios experimentos se han realizado para saber a qué se debe este comportamiento, como los que se describen a continuación. Chen y colaboradores (2003) realizaron varios experimentos con el objetivo de expresar, medir estabilidad y actividad de distintos fragmentos de la proteína T. Esto con el fin de establecer si las actividades se podrían separar y así determinar en qué regiones de la proteína se localizaban. Los resultados arrojaron que todos los fragmentos que presentan los primeros 88 residuos de aminoácidos de la región N-terminal de la proteína T, se expresan bien y muestran actividad de corismato mutasa. Los fragmentos con actividad de CM no mostraron actividad de PDH. También observaron que los fragmentos correspondientes a los residuos 93-373 y 96-3 conservan de un 25% al 50% de la actividad de PDH de la proteína T respectivamente, pero no muestran actividad de CM. Además notaron que el fragmento que va del residuo 101 al 373 no mostró actividad de CM o PDH indicando que los residuos 97-100 de la proteína T son importantes para esta actividad. Asimismo obtuvieron información que muestra que la proteína completa después de ser purificada (esto mediante la adición de una etiqueta de 6 histidinas en la región carboxilo terminal) perdía actividad de PDH al ser almacenada, la actividad de CM era menos afectada (Chen et al., 2003). Estos resultados representaron un primer esfuerzo sistemático para tratar de mapear que zonas de la proteína están involucradas en cada actividad.

Por otro lado, Osuna et al. (2012) reportaron que la sustitución del dominio de corismato mutasa por el dominio $\beta 1$ de la proteína G (una proteína con un tamaño de 56 aminoácidos que se pliega en una lámina β de cuatro hebras empaquetadas contra una hélice α) en la proteína T ayudó a estabilizar al dominio de prefenato deshidrogenasa. Esta construcción mostró niveles de actividad catalítica muy parecidos a los de la proteína silvestre. Este dominio forma un dímero, mutaciones que impiden la dimerización no estabilizaron al dominio de PDH. La estabilidad se

daba sólo cuando el dominio era colocado en la región amino terminal ya que al colocar el dominio $\beta 1$ en la región carboxilo el dominio no se estabiliza. Al mostrar estos niveles de actividad y expresión quedó claro que la dimerización era importante para estabilizar el dominio de PDH. Con este antecedente, Osuna et. al. (2012) llevaron a cabo un experimento en el cual el dominio CM fue remplazado por motivos EF-hand canónico y pseudo que unen Ca^{2+} . Los motivos EF-hand canónicos se diferencian de los motivos pseudo EF-hand, en que los primeros muestran un cambio estructural dependiente de Ca^{2+} . Los motivos de unión a Ca^{2+} consisten en una secuencia de 30 residuos que adopta una estructura hélice-loop-hélice (EF-hand). Estos motivos aislados o insertados tienden a adoptar una estructura homo-dímerica, esta característica fue importante para seleccionar a estos motivos como reemplazo del dominio CM en la proteína T. Esta fusión permitió observar niveles de expresión similar a los de la proteína silvestre. La unión a Ca^{2+} causa que un motivo EF-hand canónico sufra un cambio estructural, es decir, una transición que va de una conformación cerrada a una abierta, que resulta en la exposición de una región de proteína hidrofóbica. Esto permitió que al reemplazar el dominio de CM con este dominio la actividad de PDH se regulará por el calcio. Se observó que la actividad se podría recuperar cuando se agregaba EDTA el cual atrapa el calcio presente en la reacción. Así se concluyó que la naturaleza dimérica del dominio de corismato mutasa es importante para estabilizar el dominio de PDH, y por tanto se expresará y ser capaz de transformar el pefenato en 4-hidroxifenilpiruvato. Dominios diméricos alternativos pueden reemplazar el dominio CM de la proteína T.

3. Planteamiento del problema.

En la síntesis del aminoácido tirosina en la bacteria *E. coli* participa la proteína T. Esta proteína consta de dos dominios; el dominio de Corismato Mutasa (CM) y el dominio de Prefenato Deshidrogenasa (PDH). El dominio CM cataliza la conversión de corismato a prefenato mediante un rearrreglo de Claisen, el segundo dominio, PDH, utiliza el prefenato para formar 4-hidroxifenilpiruvato mediante una descarboxilación oxidativa utilizando NAD^+ como cofactor. Existe evidencia experimental que ha mostrado que al intentar expresar los dominios de manera independiente, el dominio CM es capaz de expresarse y presentar niveles de actividad enzimática altos, no obstante el dominio PDH, no muestra niveles de actividad enzimática o producción de proteína. Varios estudios han demostrado que la dimerización juega un papel importante para la estabilidad del dominio de PDH.

4. Justificación

La tirosina es esencial para el crecimiento de todos los organismos, es necesaria para la síntesis de proteínas y como sustrato para la producción de otros metabolitos (quinonas, glucósidos cianogénicos, alcaloides, flavonoides y compuestos fenólicos derivados de la ruta fenilpropanoide) implicados en procesos biológicos primarios y esenciales para la viabilidad de las células.

A nivel industrial, la tirosina tiene un gran interés ya que se utiliza como producto farmacéutico, aditivo para alimentos, suplemento dietético y nutracéutico.

En el presente proyecto se pretende generar variantes estables del dominio de PDH de la proteína T (que participa en biosíntesis de tirosina en *E. coli*) mediante la adición de una etiqueta al azar en la región amino terminal de dicho dominio. Esta investigación busca ampliar el conocimiento acerca del dominio de PDH, esto con la finalidad de poder obtener información relevante que ayude a comprender mejor su comportamiento cuando se expresa de forma separada.

5. Hipótesis.

La adición de secuencias que codifican para seis aminoácidos al azar en la región 5' del gen que codifica para el dominio de PDH de la proteína T de *Escherichia coli* permitirá la estabilización del dominio, permitiendo la expresión del dominio con actividad catalítica.

6. Objetivos.

6.1. Objetivo general.

- Generar variantes estables del dominio de PDH mediante la inserción al azar de 6 aminoácidos en la región amino terminal para la estabilización de la proteína T de *Escherichia coli*.

6.2 Objetivos particulares.

- Diseñar un oligonucleótido que introduzca 6 codones al azar en la región 5' del gen que codifica el dominio de PDH utilizando las técnicas de evolución dirigida e ingeniería de proteínas.
- Construir y validar una librería de variantes que porten la etiqueta de 6 aminoácidos al azar en la región 5' del gen que codifica para el dominio de PDH.
- Seleccionar variantes que presenten actividad de PDH para analizar sus genotipos y fenotipos.
- Evaluar la actividad catalítica y la expresión de las variantes seleccionadas con actividad de PDH.

7. Materiales y métodos.

Para mayor información acerca de los reactivos, materiales y equipos utilizados en el presente proyecto consultar el ANEXO I.

7.1. Diseño del oligonucleótido que introduce 6 codones al azar en la región 5´ del gen que codifica el dominio PDH.

El diseño del oligonucleótido consistió en combinar la técnica de evolución dirigida e ingeniería de proteínas (diseño racional) ya que se pretendía generar diversidad, pero solo en una región específica del gen que codifica para el dominio de PDH de la proteína T. A este gen se le insertaron 6 codones degenerados (18 nucleótidos) en el extremo 5´. Los codones son degenerados ya que son una mezcla de bases en cada posición (NNG/C), donde N es cualquiera de las 4 bases (A, G, C y T) y la última posición del codón es G o C. Esta mezcla genera 32 posibles codones, donde están representados los 20 aminoácidos y 1 de término.

Los oligonucleótidos que se utilizaron para introducir los codones son los siguientes:

1. Mut6nng-PDH (49 nt)

Secuencia: 5´ AAT-TAA-CCA-TGG-CCN-**NSN-NSN-NSN-NSN-NSN-NSC**-CGT-CAC-TGC-GTC-CGG-T 3´

En negritas se muestran los 6 codones degenerados en donde N=G/A/T/C y S=G/C. En subrayado se muestra el sitio de corte para la enzima de restricción NcoI. Las últimas 17 bases del oligonucleótido aparean con el gen del dominio de PDH y permiten amplificar el gen.

2. PBad Rev (22 nt)

Secuencia: 5´ TCT-GAT-TTA-ATC-TGT-ATC-AGG-C 3´

Este oligonucleótido aparea con la secuencia del plásmido en el cual se encuentra clonado el gen del dominio de PDH.

Los oligonucleótidos utilizados fueron sintetizados en la Unidad de Síntesis y Secuenciación de DNA del Instituto de Biotecnología/UNAM.

Para introducir los codones en el gen se realizó una PCR con los oligonucleótidos mencionados bajo las siguientes condiciones:

Ciclos	Temperatura	Tiempo
1	92 °C	5 min
25	92 °C	30 seg
	55 °C	30 seg
	72 °C	1 min

1	72 °C	5 min
---	-------	-------

Condiciones de reacción:

Reactivos	Cantidad
Buffer F (10X)	5 µl
DNTP´S	5 µl (2.5 mM de cada deoxinucleótido)
Oligonucleotido 1	4 µl (5 picomoles/ µl)
Oligonucleotido 2	4 µl (5 picomoles/ µl)
ADN molde	1 µl (~20 ng del plásmido pBadmyc-His A)
Enzima Taq polimerasa	4 µl (1 U/ µl)
Agua Tetradestilada	27 µl
Volumen Final 50 µl	

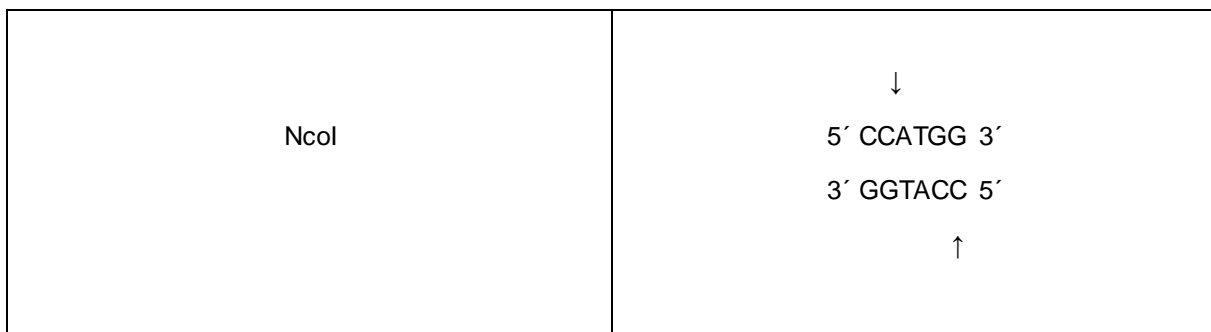
El producto de la reacción de PCR se verificó en un gel de agarosa al 1.5 %. La reacción de PCR se purificó utilizando el Kit High Pure PCR Product Purification (Roche) siguiendo las recomendaciones del fabricante. La elución del producto de la PCR se realizó en 55 µl de buffer de elución.

El producto purificado de la PCR fue digerido en 60 µl de reacción con las enzimas de restricción NcoI/HindIII (10 unidades de cada enzima y 50 µl de producto de PCR) en el buffer de reacción 2.1. La digestión se realizó a 37 °C toda la noche.

Las enzimas de restricción utilizadas, reconocen los siguientes sitios de corte (Tabla 2):

Tabla 2. Sitios de corte que reconocen las enzimas HindIII y NcoI.

Enzima	Sitio de Corte
HindIII	↓ 5'AAGCTT 3' 3'TTCGAA 5' ↑



La flecha indica en donde la enzima realiza el corte del enlace fosfodiéster, en ambos casos se generan extremos cohesivos.

Para desnaturalizar a las enzimas de restricción, se tomó la digestión y se colocó a 65 °C durante 10 minutos. Posteriormente se purificó el ADN de gel de agarosa preparativo al 1.5 %. En cada pozo se colocó 12 µl de buffer de carga para electroforesis DNA 6X (detalles de composición y preparación se describen en el ANEXO II) a los 60 µl de la digestión, se corrió a 90 V durante 40 minutos aproximadamente. El gel se tiñó con bromuro de etidio y se observó en un transiluminador de luz UV de longitud alta (365 nm) (MacroVue UV-25. Hoefer SA). La banda del ADN fue cortada y el ADN se purificó utilizando el Kit High Pure Plasmid Isolation de Roche siguiendo las instrucciones del fabricante. El ADN purificado se verificó en un gel de agarosa al 1.5%.

7.2. Construcción de la librería de variantes del gen que codifica para el dominio PDH de la proteína T de *E. coli*.

7.2.1. Clonación

El producto de la PCR digerido y purificado se ligó al vector pBadmyc-His A-PDH (Figura 8), cortado con las mismas enzimas de restricción (NcoI/HindIII) en una relación molar 1:3 vector/inserto, las condiciones de la ligación fueron las siguientes:

Reactivos	Cantidad
Vector pBadmyc-His A-PDH	4 µl (~100 ng)
Inserto	1 µl (~ 20 ng)
Buffer ligasa (10X)	2 µl
T4 DNA Ligasa	1 µl (1 U/µl)
Agua tetradestilada	12 µl
Volumen Final	20 µl

se dejó la reacción toda la noche a 16 °C en un bloque de temperatura controlada.

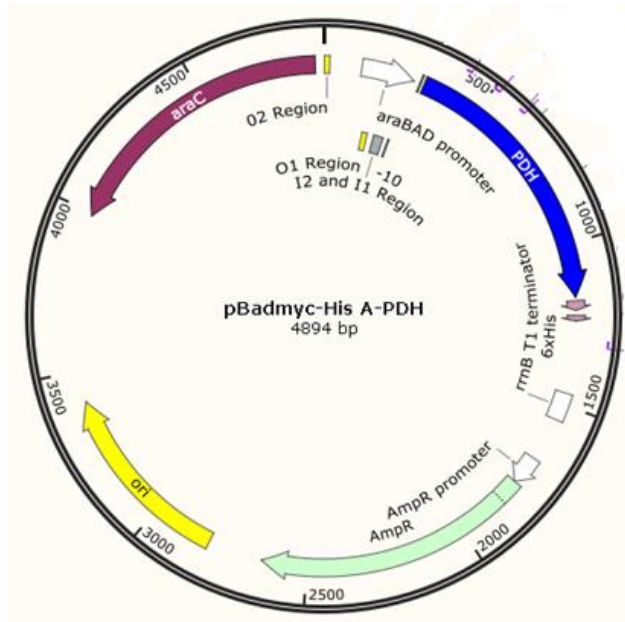


Figura 8. Diagrama del plásmido pBadmyc-His A-PDH. En esta imagen se observa la estructura del plásmido, el cual tiene un sitio de origen de replicación (ori, flecha amarilla), un gen que le confiere resistencia a ampicilina (flecha azul claro), un operador inducible por arabinosa (flecha blanca), y el gen que codifica para el dominio de preñato deshidrogenasa (flecha azul).

El plásmido pBadmyc-His A que utilizamos para su expresión fue usado en proyectos anteriores realizados en el laboratorio asegurando así la calidad del mismo, tiene un promotor inducible por arabinosa, y un gen que le confiere resistencia al antibiótico ampicilina lo que nos permite regular la expresión de la proteína de manera fina en nuestros experimentos.

La reacción de ligación se colocó durante 10 min a 65 °C para inactivar la enzima. Posteriormente fue concentrada por precipitación con butanol como se describe en la sección de ANEXO III.

Células BW25113 ($\Delta tyrA$) de *E. coli* fueron transformadas por electroporación con la reacción de ligación como se describe en el ANEXO IV, y después se platearon 20 μ l y 50 μ l de la transformación en cajas con medio Luria-Bertani (LB) con

ampicilina (amp) (200 µg/ml) (medio de cultivo codificado como LB+amp). La presencia de ampicilina ayuda a evitar el crecimiento de células que no posean el plásmido. Se dejaron incubando a 30 °C durante 24 horas.

7.3. Validación de la librería de variantes del gen que codifica para el dominio PDH de la proteína T de *E. coli*.

Para validar la calidad de la librería, de las colonias que crecieron, se tomaron 4 al azar y se usaron para inocular 5 ml de medio LB líquido con amp y Kanamicina (km) (25 µg/ml) (medio de cultivo codificado como LB+amp+Km). Se dejaron incubando con agitación a 30 °C durante toda la noche. Se purificó plásmido de ellas utilizando el kit High Pure Plasmid Isolation (Roche) siguiendo las recomendaciones del fabricante.

Se realizó una digestión de los plásmidos purificados con las enzimas NcoI/HindIII para verificar si portaban el inserto correcto.

Las condiciones de la digestión fueron las siguientes:

Reactivos	Cantidad
ADN (Plásmido)	8 µl
Buffer 2.1	1 µl
Enzima HindIII	0.25 µl (1 U/µl)
Enzima NcoI	0.25 µl (1 U/µl)
Agua tetradestilada	0.5 µl
Volumen Final	10 µl

Se dejaron a 37 °C durante 2 horas.

Posteriormente se verificó en gel de agarosa al 1.5% la presencia del fragmento clonado. Las cuatro clonas tenían el plásmido con el inserto correcto. Los cuatro plásmidos se mandaron a secuenciar para determinar si presentaban el inserto de 6 codones y la diversidad en sus secuencias. Se usaron 2 µl (5 picomoles/ µl) del oligonucleótido pBadFW el cual apareja con secuencia del vector en la región 5´de gen del dominio de PDH, 12 µl del plásmido purificado y 2 µl de agua grado miliQ.

Oligonucleótido pBadFW (21 nt):

5´ TAT-GCC-ATA-GCA-TTT-TTA-TCC 3´

Las secuencias fueron visualizadas y analizadas con los programas Chromas Lite y BioEdit respectivamente.

7.4 Sistema de selección de variantes con actividad de PDH.

Para la selección de variantes con actividad de PDH, se transformaron células BW25113 ($\Delta tyrA$) por electroporación como se describe en la sección de ANEXO III con la ligación de la librería de variantes del gen de PDH. Después de la recuperación se lavaron las células tres veces con medio mínimo M9 (MM M9; detalles de composición y preparación se describen en el ANEXO V) líquido para quitar la tirosina presente en el medio SOC (detalles de composición y preparación se describen en el ANEXO V), como se describe a continuación; Bajo condiciones estériles se puso un 1 ml de medio SOC que contienen las células transformadas en un tubo eppendorf de 1.5 ml, posteriormente se centrifugó durante 6 minutos a 5 000 rpm. Se desechó el sobrenadante y se agregó 1 ml de MM M9 líquido (ANEXO V), se resuspendió por agitación en un vortex (Velp Scientifica, Europa) y se centrifugó de nuevo. Una vez que las células fueron lavadas se inocularon en dos diferentes medios de cultivo para observar su crecimiento. Se tomó 50 μ l del cultivo lavado y se sembró en MM M9 sólido (ANEXO V). Se tomaron 5 μ l del cultivo de lavado y 50 μ l de MM M9 líquido y se inocularon en medio sólido LB+amp+km ambos medios se dejaron incubando a 37 °C toda la noche. Posteriormente el medio MM M9 se dejó incubando a 37 °C y el medio LB+amp+km a 30 °C durante un día. Al siguiente día se dejaron ambos medios incubando a 37 °C por 24 horas.

7.4.1. Validación del fenotipo y del genotipo de las variantes.

Para validar el fenotipo de complementación de las clonas que crecieron, se inocularon 100 colonias al azar (50 de cada temperatura) en diferentes medios: medio sólido LB+amp+Km, MM M9 con Arabinosa (Ara) (0.2 %) (Medio de cultivo codificado como MM M9+ara) y MM M9 se dejaron incubando a 37 °C.

Para comprobar que las clonas que crecieron en MM M9 presentaran el gen del dominio de PDH clonado de manera correcta se realizó PCR. Para esto se

colocaron 20 μ l de reacción de PCR en un tubo de PCR en condiciones estériles y se le agregó una porción de la colonia a analizar con una punta de 200 μ l. Las condiciones de reacción fueron las mismas utilizadas en la PCR para introducir los 6 codones en el gen (ver sección 6.1). El producto de la PCR se visualizó en un gel de agarosa al 1.5 % para determinar si presentaba el gen de PDH de manera correcta. Se descartaron las clonas que presentaban una banda de barrido en la parte inferior, en experimentos anteriores la presencia de ese barrido es un indicativo de falsos positivos.

En total se obtuvieron 17 variantes que mostraron tener el tamaño del gen correcto y no presentaron el barrido de bajo peso molecular. Se purificó el plásmido de estas clonas utilizando el kit High Pure Plasmid Isolation de Roche siguiendo las recomendaciones del fabricante. Se realizó un patrón de restricción con las enzimas NcoI/HindIII, como se ha indicado previamente (consultar sección 6.1) para verificar que presentarán el tamaño del inserto correcto.

Para corroborar que el plásmido presente en las clonas fuera el responsable del fenotipo de crecimiento en ausencia de tirosina se transformaron células por electroporación como se ha descrito previamente sin el gen cromosomal de *tyrA* con las variantes seleccionadas, se platearon en medio sólido LB+amp+km. Se dejaron incubando a 30°C toda la noche. Se escogió una colonia de cada variante al azar y se colocaron en 5 ml de medio líquido LB +amp+km y se dejaron incubando con agitación a 30° C toda la noche. Se lavaron las células tres veces para desechar rastros de tirosina como se ha descrito anteriormente. Bajo condiciones estériles se realizaron 3 diluciones seriadas en MM M9 líquido. Para esto se tomaron 10 μ l de las células lavadas y se les agregaron 990 μ l de MM M9 líquido (1:100), de esta dilución se agregaron 10 μ l a 990 μ l de MM M9 líquido (1:100), finalmente se agregaron 100 μ l de la última dilución a 900 μ l de MM M9 líquido (1:10) (10^{-5} final). Se prepararon cuatro cajas con distintos medios sólidos: LB, MM M9, MM M9 + Tirosina (*tyr*) (10 mg/ml) (MM M9 +Tyr), MM M9 + ara: Se tomaron 10 μ l de cada dilución y se inocularon mediante la técnica de goteo en cada medio. Como controles se emplearon: proteína T (control positivo) y control negativo (dominio

PDH). Se dejaron incubando a 37 °C hasta observar crecimiento. Se tomaron 8 variantes que mostraban el mejor crecimiento en las cajas con MM M9, nuestro medio selectivo. Se Repitió el experimento para validar el fenotipo de crecimiento observado.

Para la validación del genotipo, las 8 variantes se mandaron a secuenciar para verificar que no hubiera algún cambio en alguna otra parte del gen y poder analizar sus secuencias, para esto se usaron las siguientes condiciones:

Reactivos	Cantidad
Plásmido de cada variante	12 µl
Oligonucleotido pBadFW	2 µl
Agua	2 µl
Volumen final	16 µl

Las muestras se mandaron a la Unidad de Síntesis y Secuenciación de DNA del Instituto de Biotecnología, UNAM. Posteriormente fueron visualizadas y analizadas con los programas Chromas Lite y BioEdit respectivamente.

7.5. Caracterización de variantes con fenotipo y genotipo validado.

7.5.1 Crecimiento de las variantes seleccionadas en medio líquido

Para la preparación de las muestras primero se inoculó 1 colonia de cada variante, el control positivo (proteína T) y el control negativo (dominio PDH) en 3 ml de medio sólido LB+amp+km, se dejaron incubando toda la noche a 37 °C con agitación a 210 rpm. Bajo condiciones estériles el medio de cultivo se pasó a un tubo eppendorf estéril de 1.5 ml, se centrifugaron 6 minutos a 5 000 rpm, se desechó sobrenadante. Las células fueron lavadas cuatro veces para descartar rastros de tirosina como se ha descrito previamente. Después de lavarlas, se resuspendieron en 1 ml de MM M9 utilizando un vortex. Para hacerlas crecer se colocó en una placa de ELISA 200 µl de MM M9 líquido, 200 µl de MM M9+ara líquido, 200 µl de MM M9+tyr líquido y en cada pozo se agregaron 5 µl de muestra de la resuspensión en MM M9 de cada variante. Para determinar el crecimiento se midió la absorbancia a 600 nm. Las mediciones fueron hechas cada 10 minutos utilizando un espectrofotómetro de

placas de 96 pozos controlado por la plataforma Tecan (Tecan Group LTD, Männedorf, Switzerland).

7.5.2. Validación de la expresión del dominio PDH de las variantes seleccionadas mediante electroforesis en gel de acrilamida

Para observar la expresión del dominio PDH de las variantes seleccionadas, se llevó a cabo una electroforesis en gel de poliacrilamida, primero se prepararon las muestras. Para esto se inocularon en 3 ml de medio líquido LB+amp+Km las ocho variantes más el control positivo (Proteína T) y el control negativo (dominio de PDH), se dejaron creciendo en agitación a 37 °C toda la noche. Se usaron 30 µl de estos cultivos para inocular 2 series de tubos con 3 ml de medio líquido LB+amp+ara para cada variante. Se dejaron creciendo 3 horas y después se agregó arabinosa a la segunda serie para inducir la expresión de la proteína. Después se incubaron a 30 °C/ 210 rpm por 4 horas. Se colectaron las células en tubos eppendorf estériles de 1.5 ml. Se centrifugaron 1 min a 12,000 rpm, se descartó el líquido y el botón celular se almacenó a -70 °C. El botón celular se resuspendió en 150 µl de buffer de actividad y se sonó con 2 pulsos de 20 segundos en un sonicador Branson (Danbury, CT, USA) a 6 de potencia y con la micropunta. El sonicado se centrifugó 10 minutos a 12000 rpm. El sobrenadante se transfirió a tubos eppendorf limpios (1.5 ml), se tomaron 10 µl del sobrenadante, y se les agregó 4 µl de buffer de carga (ANEXO II) para electroforesis SDS-PAGE.

Los tubos se calentaron por 2 minutos a 90 °C. Se preparó gel de acrilamida al 10 % y en cada pozo se colocaron 10 µl de muestra, se corrió el gel por 40 minutos aproximadamente a 200 V.

Para teñir el gel se utilizó azul de Coomassie (ANEXO VI). Posteriormente se agregó solución para desteñir (ANEXO VI) y se dejó en agitación a 50 rpm durante toda la noche.

Es importante señalar que se realizó otro gel de poliacrilamida al 10% para poder normalizar la cantidad de proteína de PDH que se utilizó en experimentos posteriores. Para esto se diluyó la proteína silvestre 1:5, 1:10, 1:15 y 1:20, ya que

al no estar purificada nuestra proteína es difícil conocer la concentración si no se realiza esto. Las 4 variantes que utilizamos fueron seleccionadas al azar.

7.5.3 Ensayo enzimático.

Se determinó la concentración de proteínas totales de la fracción soluble mediante una prueba de Bradford, esto se realizó para normalizar el volumen que se debería usar para tener la misma cantidad de proteína en los ensayos de actividad.

La reacción que realiza el dominio de PDH es la conversión del prefenato en 4-hidroxifenilpiruvato, utiliza el NAD⁺ como cofactor y además produce NADH y CO₂ (Nelson & Cox, 2008). Para medir la actividad específica de las variantes seleccionadas se utilizó extracto crudo, siguiendo la aparición de NADH como indicador de la actividad. Esto se hizo mediante la medición de absorbancia a 340 nm. Se utilizó como control positivo la proteína T y como control negativo el dominio de PDH. La reacción se realizó por triplicado en una placa de ELISA de 96 pozos. Se realizaron dos condiciones de cada clona, con y sin tirosina que es un inhibidor de la reacción. Se colocó 200 µl buffer de actividad (ANEXO V), 2 µl prefenato, 2 µl NAD⁺ y 2 µl de extracto de la variante a caracterizar. La reacción se inició con la adición del NAD⁺. En la segunda serie se colocó lo mismo más 3 µl de tirosina esto para determinar si las variantes mantienen la regulación por tirosina. Las mediciones se realizaron durante 10 minutos utilizando un espectrofotómetro de placas de 96 pozos controlado por la plataforma Tecan (Tecan Group LTD, Männedorf, Switzerland).

7.6. Determinación del tamaño del banco

Para calcular el tamaño del banco se contó el número de colonias que crecieron en el medio sólido LB+amp+km de la transformación del banco, 62 en este caso después dividimos entre los 5 µl que se utilizaron para inocular, dando 12.4 colonias por 1 µl de transformación. En un 1 ml que es el volumen total de cada transformación tenemos 12,338 variantes. Durante el experimento se llevaron a cabo 15 transformaciones utilizando 2 µl de ligación en cada una.

8. Resultados y discusión

En los últimos años, el diseño de proteínas y la evolución dirigida han permitido obtener proteínas *de novo* a partir de secuencias completamente nuevas, estas nuevas proteínas pueden realizar reacciones no presentes en la naturaleza o que ya existen en proteínas a las cuales no son homólogas (Hoegler & Hecht, 2016). Así, utilizando la evolución dirigida en combinación con el diseño racional en el presente proyecto se generaron variantes del dominio de PDH de la proteína T de *E. coli* que mostraran actividad catalítica, existían antecedentes que mostraban que la dimerización del dominio de PDH era importante para estabilizarlo y observar actividad catalítica. Resultados previos demostraron que la adición de una etiqueta de 6 aminoácidos (histidinas) en la región amino terminal del dominio de PDH permitía observar actividad, aunque no se veía que la proteína se expresara. Esto planteó desarrollar nuevos experimentos que permitieran demostrar si alguna otra secuencia de aminoácidos funcionaría mejor, y constituyen las bases del presente trabajo.

8.1. Validación del oligonucleótido y librería de variantes del gen que codifica para el dominio de PDH de la proteína T de *E. coli*

La combinación de las técnicas de evolución dirigida e ingeniería de proteínas permitió que el diseño del oligonucleótido generara diversidad en las secuencias sobre una región específica del gen que codifica para el dominio de PDH. El oligonucleótido inserta 6 codones (18 nucleótidos) que codifican para 6 aminoácidos al azar en la región 5' del gen que codifica para el dominio de PDH. Los codones constan de una mezcla de nucleótidos (NNG/C), que da origen a 32 codones, en los cuales están representados los 20 aminoácidos y 1 de término, suficientes para generar la variabilidad en las secuencias. El número de secuencias que se pueden generar por los 6 codones es de 1×10^9 , al ser un número bastante grande, en este proyecto solo se evaluó una pequeña fracción del mismo.

Se tomaron 4 clonas al azar para validar la librería, es decir para comprobar que trajeran los 18 nucleótidos extra en la región 5' del gen que codifica para el dominio PDH, esto nos permitió determinar la diversidad en las secuencias, además permitió

determinar si el diseño del oligonucleótido permite diversidad en las secuencias añadidas.

Al analizar las secuencias se observó que no existe un sesgo en la librería, ya que cada secuencia es única y las clonas no presentaron cambios en otra zona del gen (Tabla 4), por lo tanto al traducirse en aminoácidos se observa nuevamente la diversidad (Tabla 5).

Al analizar el contenido de nucleótidos con base en la relación G/C y A/T en las secuencias de las 4 clonas se observa que hay 42 nucleótidos para la relación G/C y 30 que corresponden a la relación A/T, esto representa el 58.33% y 41.66%. Estos valores son similares a los valores esperados teóricamente (el 60 % debe corresponder a G/C y el 40 % a A/T), esto se debe al diseño del oligonucleótido que se empleó. Vinogradov (2003) reportó que la importancia de que el DNA de doble hélice sea rico en GC radica en que ésta confiere un nivel de organización genómica avanzada, donde la transcripción activa es optimizada, más allá de la termoestabilidad. Cada codón puede tener cualquiera de las 4 bases en la primera y segunda posición y solo G o C en la tercera posición, eso indica que de los 10 posibles nucleótidos 6 son G o C (60%) y 4 A o T (40%). Esto nos indica que las secuencias de las clonas no muestran un sesgo. Todo lo anterior permitió verificar la diversidad de la librería.

Tabla 4. Secuencias de nucleótidos de colonias seleccionadas.

Colonia	Secuencia de nucleótidos
1	TCC-TCG-GTG-GTG-GCC-TAC
2	AAG-GTG-GAC-TCC-GTC-TTG
3	TTC-CTG-GAC-TCG-GTG-TGC
4	GAC-AGG-ATG-ACG-TTG-TAG

Tabla 5. Secuencias de aminoácidos de clonas analizadas.

Colonia	Secuencia de aminoácidos
1	Ser Ser Val Val Ala Tyr

2	Lys	Val	Asp	Ser	Val	Leu
3	Phe	Leu	Asp	Ser	Val	Cys
4	Asp	Arg	Met	Thr	Leu	Cys

8.2. Validación del fenotipo y del genotipo.

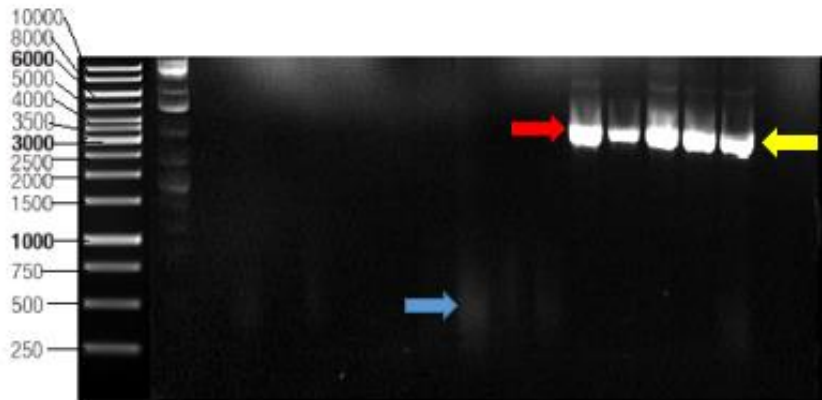
Con la librería generada, se transformaron células nuevas y se sembraron en un medio rico, de las cuales se seleccionaron 100 variantes para realizar PCR y así poder validar su fenotipo. En trabajos preliminares se ha observado que el producto de PCR visualizado en gel de electroforesis al presentar un barrido de bajo peso molecular es un indicativo de que el plásmido sufrió un rearrreglo, lo que genera falsos positivos, y por lo tanto si se llegase a presentar crecimiento de estas variantes, éste no será debido a las modificaciones del gen del dominio de PDH. Así, de las 100 variantes analizadas por PCR, sólo 17 variante mostraron no tener el barrido (Figura 9a), es decir, fueron clonas que posiblemente presentaban actividad de PDH. Además se corroboró que las 17 variantes seleccionadas portaran el plásmido digiriéndolas con enzimas de restricción, posteriormente se corroboró mediante electroforesis en gel de agarosa, en donde todas las variantes mostraron tener el patrón de digestión esperado (Figura 9b).

Las 17 variantes seleccionadas se transformaron nuevamente y se sembraron en diferentes medios para la validación del fenotipo. Fisher et. al. (2011) describen que todas las células que son transformadas y cultivadas en un medio de cultivo rico, presentan crecimiento. Además, dichos autores indican que una forma de validar el fenotipo de células cuyo genoma ha sido modificado es comparar su crecimiento en un medio rico y un medio selectivo; un crecimiento similar indica que el fenotipo observado (crecimiento en medio selectivo) es debido a las modificaciones de su genoma. En el presente trabajo, el sistema de validación consistió en elegir aquellas variantes del dominio de PDH que presentarán un crecimiento en medio selectivo (sin tirosina), ya que las células carentes del gen *tyrA* no pueden crecer en medio carente de tirosina, al ser incapaces de producirla. Sin embargo, aquellas variantes que logren crecer en este medio, es debido a que portan una variante de PDH

funcional, que le permite la producción de 4-hidroxifenilpiruvato y así la producción de tirosina. Para llevar a cabo la producción de tirosina, el prefenato que utiliza el dominio de PDH (para transformarlo en 4-hidroxifenilpiruvato y así generar tirosina) lo produce el dominio Corismato Mutasa de la proteína P, una proteína homóloga en *E. coli* encargada de la síntesis de fenilalanina (Chen et al., 2003). Cuando las variantes se encuentran en un medio rico (es decir hay tirosina en el medio) presentan un buen crecimiento al tener todo lo necesario para crecer, este comportamiento se observa también cuando el medio selectivo es suplementado con tirosina. Por tanto, las variantes fueron cultivadas en los medios LB (medio rico, contiene tirosina) y MM M9 (medio selectivo, sin tirosina), MM M9+tyr (medio selectivo+tirosina), MM M9+ara (medio selectivo+arabinosa (el cual es un inductor del plásmido), y así se seleccionaron aquellas que crecieron en todos los medios lo que indica que el dominio PDH es estable. Por tanto, las variantes que se seleccionaron crecieron en un medio selectivo además de ser capaces de crecer en medios suplementados con tirosina y arabinosa.

En los medios donde hay presencia de tirosina, las variantes mostraron buen crecimiento aunque en diversos grados de crecimiento de acuerdo a las diluciones realizadas (Figura 10a-10c). Sin embargo, en el medio selectivo (MM M9) no todas las variantes mostraron crecimiento (Figura 10d). Por lo tanto se seleccionaron 8 variantes que mostraron el mejor crecimiento, debido a que posiblemente presentan un dominio de PDH estable. Estas variantes se seleccionaron para caracterizarlas en mayor detalle, el círculo amarillo en la Figura 10d las muestra; estas son: 1, 4, 5, 6, 7, 9, 12 y 13, que en lo sucesivo tendrán la siguiente nomenclatura: 1 = 1, 4 = 2, 5 = 3, 6 = 4, 7 = 5, 9 = 6, 12 = 7, 13 = 8.

a)



b)

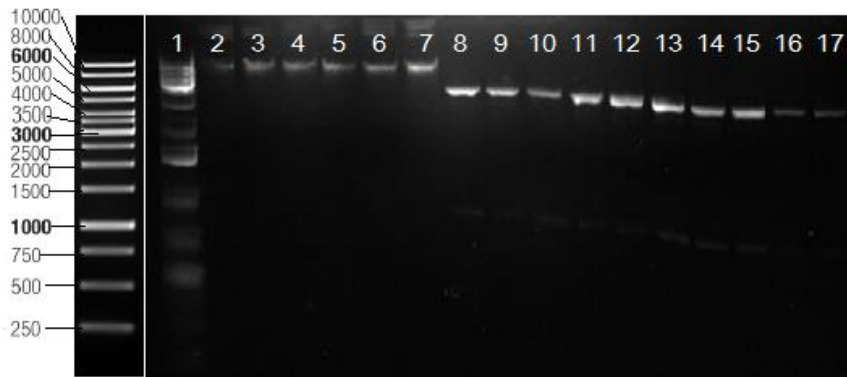
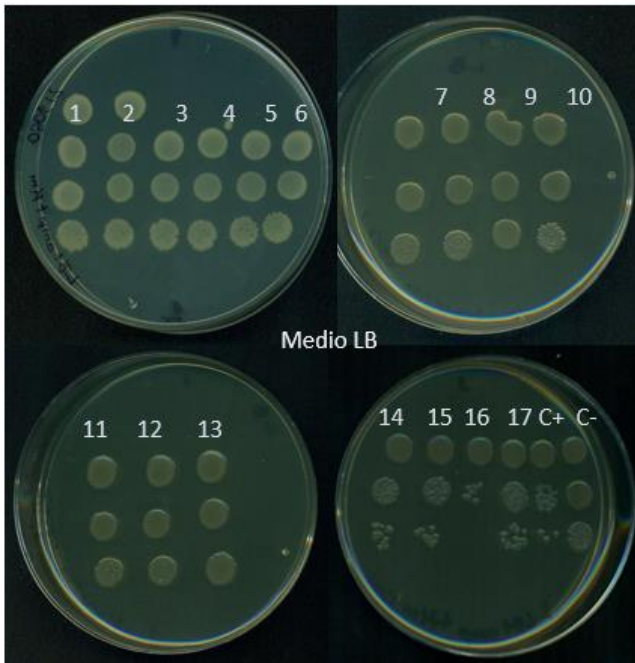
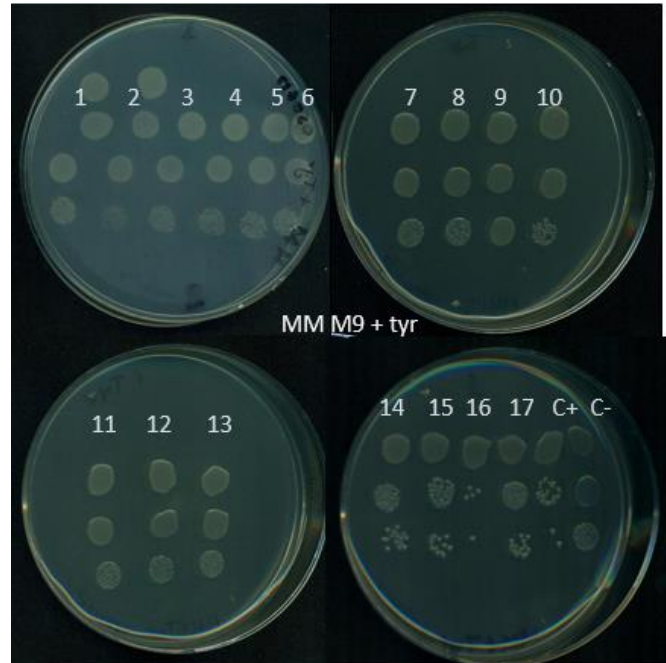


Figura 9. Validación de genotipo de variantes seleccionadas. a) Algunos productos de PCR visualizadas en gel de electroforesis. La flecha azul muestra el barrido de bajo peso molecular, la flecha roja indica a la variante con el plásmido y sin la presencia del barrido, la flecha amarilla indica a la variante con el plásmido; b) Productos de digestión enzimática de las 17 variantes seleccionadas. 1: marcador de peso molecular, 2-7: primeros plásmidos sin digerir, 8-17: primeras 10 variantes digeridas, las cuales presentan el patrón de digestión correcto, mostrando dos bandas: 4046 pb y 848 pb. La imagen es una muestra representativa de una unidad experimental.

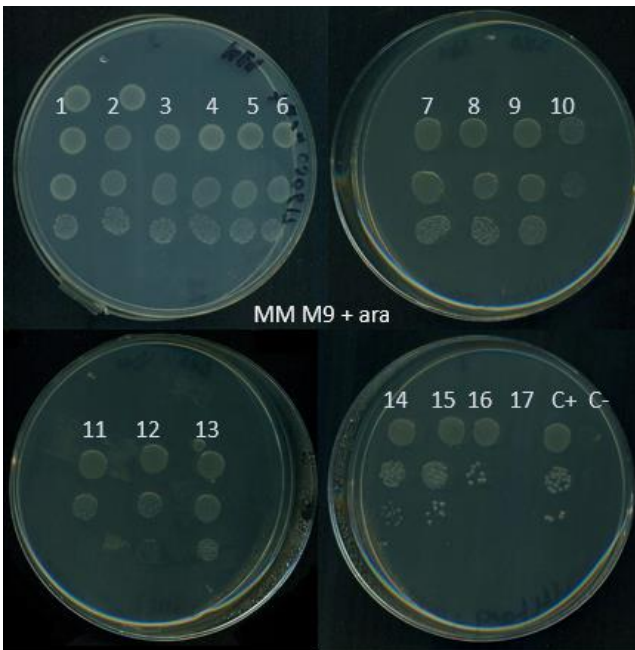
a)



b)



c)



d)

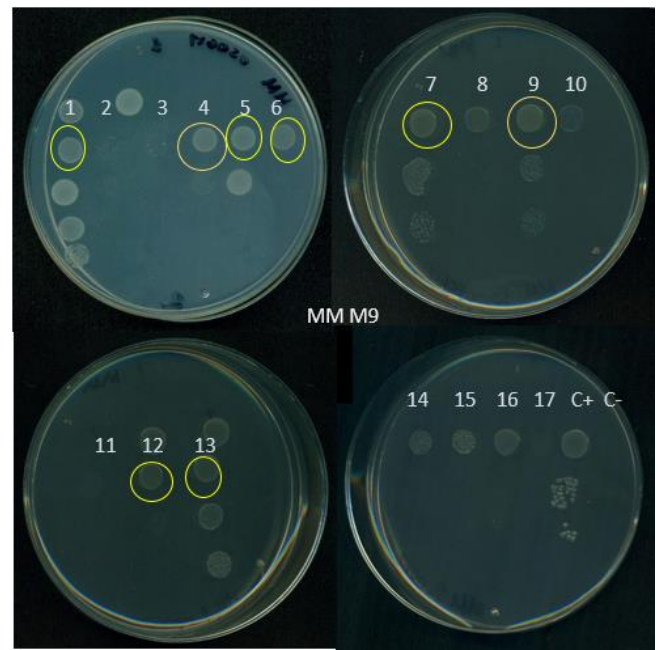


Figura 10. Validación del fenotipo de las variantes seleccionadas a través de su crecimiento en diferentes medios de cultivo (LB; MM M9+tyr; MM M9+ara; MM M9). Los números del 1 al 17 corresponden a las variantes seleccionadas y para cada número se muestra el grado de crecimiento de las diluciones empleadas por el método de goteo, C+ corresponde al control positivo (proteína T silvestre), C- se refiere al control negativo que (dominio PDH). a) Crecimiento en medio LB. b)

Crecimiento en medio MM M9 + tyr. c) Crecimiento en MM M9 + ara. d) Crecimiento en MM M9. La imagen es una muestra representativa de una unidad experimental.

La validación del genotipo de las 8 variantes seleccionadas en los medios de cultivo, consistió en confirmar que el crecimiento observado en medio selectivo se debe a la adición de la etiqueta de 6 aminoácidos y no a mutaciones presentes en el gen, por lo cual se mandaron a secuenciar. Las secuencias de las 8 variantes no presentaron cambios adicionales en el gen del dominio de PDH, sólo mostraron los 18 nucleótidos adicionales en su extremo 5' (Tabla 6), lo que indica que el fenotipo observado se debe a la etiqueta que portan. El análisis de la composición de nucleótidos de las 8 variantes muestra que existe un sesgo en su composición, esto quiere decir que con base en la relación G/C y A/T en las secuencias de las 8 variantes se observó que hay 70 nucleótidos para la relación G/C y 74 que corresponden a la relación A/T, esto representa el 48.61% y 51.39%. La relación G/C aumentó y la relación A/T se redujo en comparación con los valores teóricos (60% y 40% respectivamente) ya que al haber 144 bases relación G/C se reduce (y el par A/T aumenta) con respecto al teórico (60% y 40% respectivamente), esto nos indica que no todas las secuencias son capaces de estabilizar al domino de PDH y así permitir el crecimiento de la bacteria que lo porta. Al traducir en aminoácidos las secuencias y analizarlas (Tabla 6) no se observó la existencia de un patrón que indique alguna preferencia hacia ciertos aminoácidos. Así como una relación entre las secuencias y la actividad observada.

Tabla 6. Secuencias de nucleótidos y aminoácidos de las 8 variantes seleccionadas.

Secuencia de clonas seleccionadas	
Secuencia de Nucleótidos	
Variante 1	TGC-TTC-GAC-AAG-ATG-TGC
Variante 2	TTC-AAG-GGC-TTG-ATC-CAC
Variante 3	GAG-CTC-ATG-AGG-ACG-AAC

Variante 4	TCG-TTG-AAC-TAC-AAG-TCG					
Variante 5	CTG-CAC-AAC-ACG-ACC-AAC					
Variante 6	GTG-AAG-TTC-AAC-CCG-TTC					
Variante 7	TTC-ATC-AAG-AAG-TCG-AGC					
Variante 8	AAG-TAC-TCG-TGC-CTC-GGC					
Secuencia de Aminoácidos						
Variante 1	Cys	Phe	Asp	Lys	Met	Cys
Variante 2	Phe	lys	Gly	Leu	Ile	His
Variante 3	Glu	Leu	Met	Arg	Thr	Asn
Variante 4	Ser	Leu	Asn	Tyr	Lys	Ser
Variante 5	Leu	His	Asn	Thr	Thr	Asn
Variante 6	Val	Lys	Phe	Asn	Pro	Phe
Variante 7	Phe	Ile	Lys	Lys	Ser	Ser
Variante 8	Lys	Tyr	Ser	Cys	Leu	Gly

8.3. Caracterización de las variantes seleccionadas

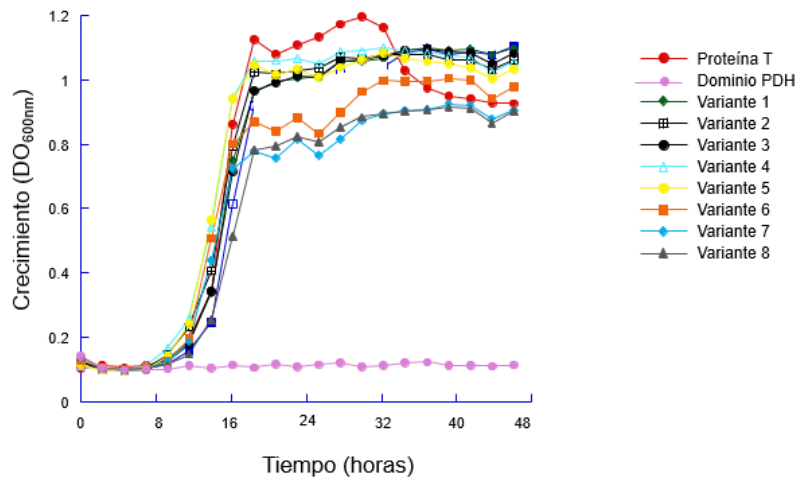
Las 8 variantes se caracterizaron en mayor detalle, es decir se analizó su crecimiento en medio líquido, así como la expresión y actividad enzimática específica del dominio PDH que contiene la etiqueta de aminoácidos (en adelante a este dominio PDH se le llamará “dominio PDH modificado”).

8.3.1 Determinación de crecimiento en medio líquido.

La determinación del crecimiento en medio líquido consistió en transformar nuevamente células de las 8 variantes y cultivarlas en 3 medios de cultivo: MM M9+tyr, MM M9+ara y medio MM M9; se tomó como control positivo a la proteína T y como control negativo al dominio de PDH silvestre. Con lo anterior se determinó que las variantes capaces de crecer en un medio selectivo (MM M9) presentan un dominio de PDH estable funcional. En medios suplementados con tirosina (MM M9+tyr) o con inductor de la expresión (MM M9+ara) se observó el crecimiento de las 8 variantes (Figura 11a,b). En medio selectivo se observó crecimiento de las 8 variantes similar al de la proteína T (Figura 11c). No obstante las variantes 2,5 ,6 y 7 mostraron una fase lag más larga en comparación a la fase lag de la proteína T (Figura 11c). Con estos resultados se sugiere que el dominio de PDH que portan es estable, ya que al ser estable, transforman el prefenato en 4 hidroxifenilpiruvato y así se lleva a cabo la producción de tirosina, permitiendo el crecimiento de las variantes en un medio selectivo. El crecimiento de estas células, bajo condiciones selectivas nos indica que las etiquetas seleccionadas posiblemente permiten la dimerización del dominio de PDH.

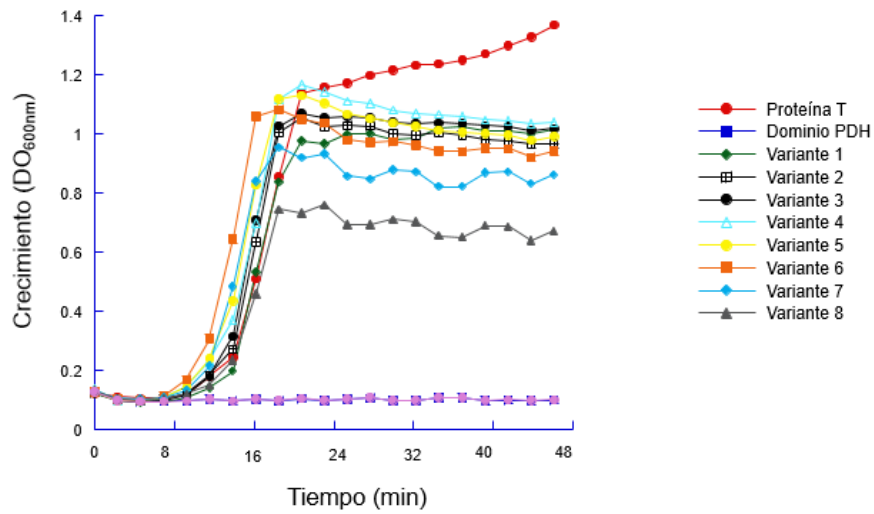
a)

Crecimiento en MM M9 con tirosina



b)

Crecimiento en MM M9 + ara



c)

Crecimiento en MM M9

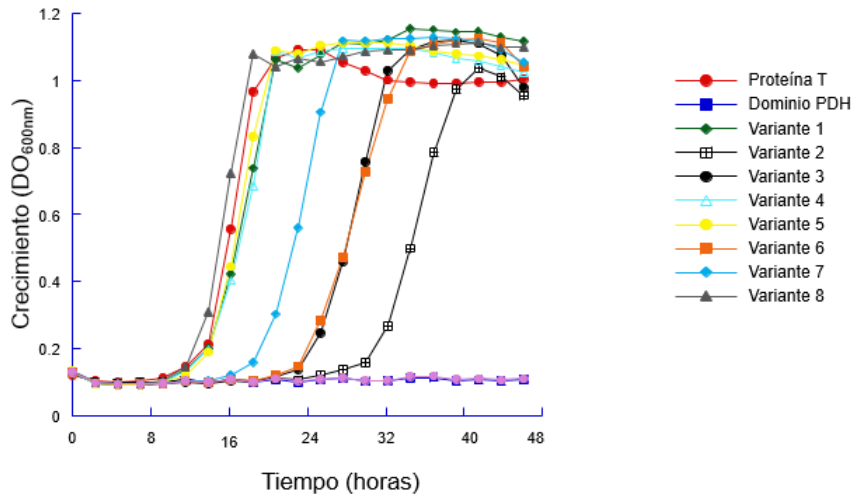


Figura 11. Cinéticas de crecimiento celular de las 8 variantes en diferentes medios de cultivo (MM M9+tyr; MM M9+ara; MM M9). a) Crecimiento en MM M9 + tyr. b) Crecimiento en MM M9 + ara. c) Crecimiento en MM M9. En las gráficas la leyenda “Proteína T silvestre” se refiere al control positivo mientras que la leyenda “Dominio PDH silvestre” se refiere al control negativo.

8.3.2. Expresión y actividad enzimática específica del dominio PDH de las variantes seleccionadas.

Al observar crecimiento en medio MM M9 de las 8 variantes se infiere la presencia de un dominio de PDH estable y funcional, por lo que el siguiente paso fue determinar si había expresión del dominio PDH modificado. Para esto se realizaron dos tipos de cultivos; con inducción (medio de cultivo con arabinosa) y sin inducción (medio de cultivo sin arabinosa) así como un control positivo (proteína T) y un control negativo (dominio PDH silvestre).

Al inducir la expresión del dominio PDH modificado en las células, estas mostraron niveles mayores de expresión del dominio en comparación con las células sin inducción (Figura 12a y 12b). El nivel de expresión de proteína T fue mayor que el nivel de expresión observable de las variantes bajo inducción (Figura 12b). La expresión del dominio PDH modificado de las 8 variantes con inducción no fue

uniforme ya que las bandas presentaron distintos grosores (Figura 12b). Las variantes se expresan mejor que el dominio de PDH silvestre. Esto se podría deber a que son más estables y por lo tanto se degradan menos que el dominio silvestre.

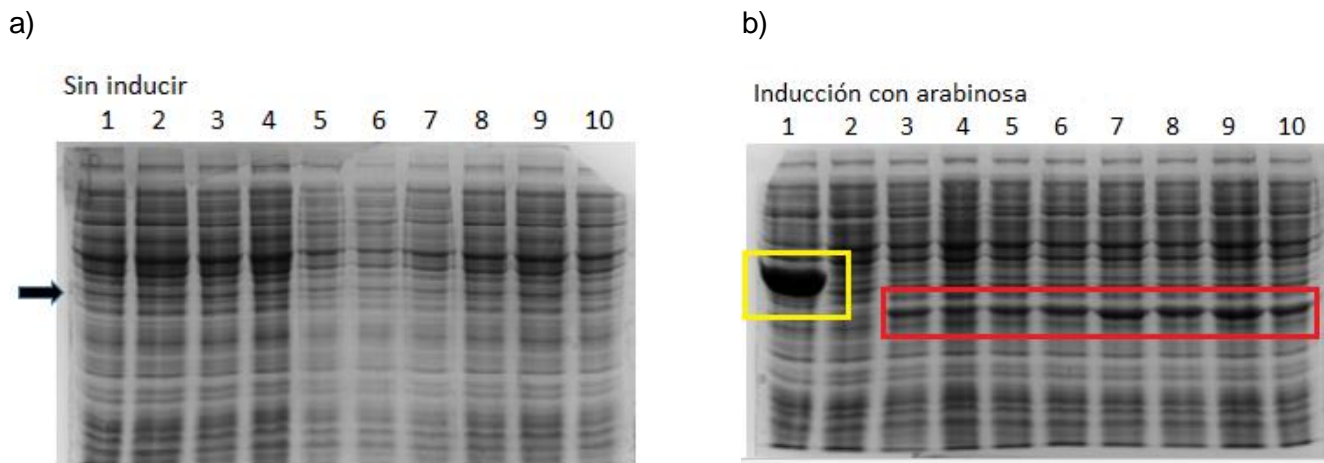


Figura 12. Expresión del dominio PDH modificado de las 8 variantes bajo inducción por arabinosa y sin inducción. a) Sin inducción, b) Con inducción. 1 corresponde a la proteína T, 2 corresponde al dominio PDH silvestre, 3-10 corresponden a las 8 variantes seleccionadas; la flecha negra indica el lugar donde se localiza la banda correspondiente a la proteína T; en el cuadrado amarillo se observa un fuerte marcaje de la banda correspondiente a la proteína T (lo que indica una mayor expresión de la proteína) comparado con las demás variantes (en rojo).

Para determinar la actividad enzimática específica del dominio PDH de las 8 variantes es importante señalar que se realizó otro gel de poliacrilamida al 10% para poder normalizar la cantidad de proteína de PDH a usar.

Con base en lo anterior se observó que el grosor de las bandas de las 8 variantes era similar al grosor de la banda correspondiente a la proteína T diluida 1:10 (Figura 13) así se determinó que la concentración de las 8 variantes era 10 veces menos que la proteína T. Esto permitió usar la misma cantidad de proteína de PDH modificada en la cinética de actividad enzimática específica, ya que al no estar purificada es difícil conocer su concentración si no se realiza de esta manera.

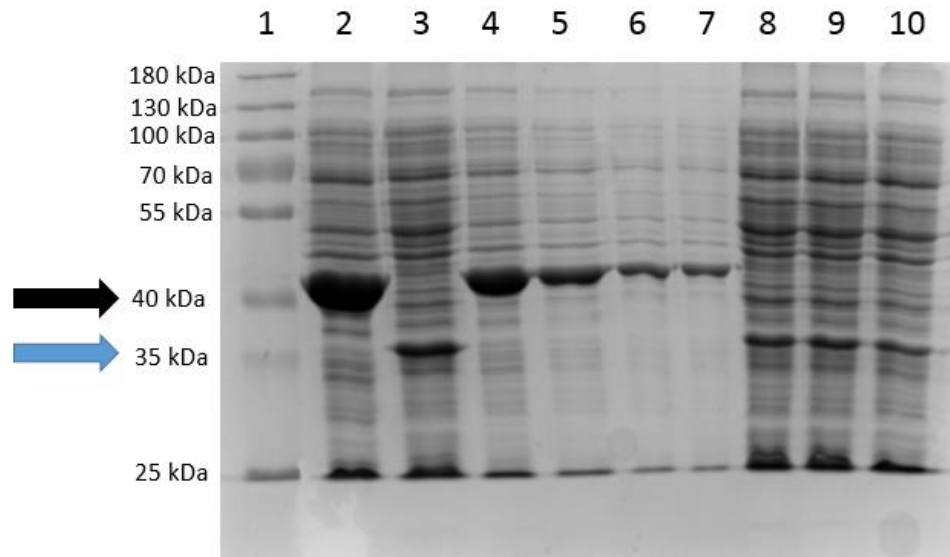


Figura 13. Normalización de la cantidad de proteína de PDH modificada a través de su comparación con la proteína T a diferentes concentraciones. 1 corresponde al marcador de peso molecular, 2 - proteína T concentrada, 3 - variante 7, 4- proteína T diluida (1:5), 5 - proteína T diluida (1:10), 6 - proteína T diluida (1:15), 7- proteína T diluida (1:20), 8 - variante 4, 9 - variante 5, 10 - variante 6. La flecha negra indica el lugar de la banda donde se muestra el peso molecular de la proteína T. La flecha azul indica el lugar de la banda donde se muestra el peso molecular del dominio de PDH modificado.

La actividad de PDH se determinó monitoreando la aparición de NADH acorde a la absorbancia medida a 340nm además se evaluó si las variantes presentaban inhibición por tirosina. Se utilizó a la proteína T (diluida 1:10) como control positivo y al dominio PDH silvestre como control negativo.

Al analizar la actividad enzimática específica del dominio PDH modificado de las 8 variantes se observó que estos presentaron actividad enzimática mostrando un comportamiento similar al de la proteína T. También se observó que todas las variantes, así como la proteína T, presentaron inhibición por tirosina (Figura 14). Se observó que las variantes seleccionadas mostraron distintos niveles de actividad (Figura 14) y de expresión del dominio PDH (Figura 12). La variante 7 fue la que mostró un mayor nivel de expresión y la más alta actividad. Esto se podría atribuir a que al existir más cantidad de proteína se incrementa la actividad enzimática. Sin

embargo, debe purificarse la proteína para tener datos más claros sobre la existencia de diferencias entre la actividad catalítica respecto al dominio expresado. Experimentos posteriores a este trabajo (datos no mostrados; información personal de los autores) han confirmado que algunas variantes son más activas que la proteína T completa. Una posible explicación de la relación que hay entre el nivel de actividad y la cantidad de proteína es que la etiqueta de 6 aminoácidos en la región amino terminal del dominio PDH promueva la estabilización del dominio PDH.

El uso de técnicas como la evolución dirigida permite la síntesis de proteínas en el laboratorio con secuencias completamente de novo, es decir que no se encuentran en la naturaleza y que ayudan a la supervivencia de organismos en condiciones en las que usualmente resultarían letales. Un ejemplo de ello es el experimento realizado por Hoegler et al., (2016), en el cual obtuvieron una proteína de novo la cual aislaron y denominaron ConK, la cual permitió el crecimiento de E.coli bajo niveles altos de cobre, mismos que normalmente serían considerados como letales, mostrando así una resistencia al cobre. Sin embargo, no mostraron resistencia a otros metales divalentes. Fisher et. al (2011) obtuvieron 18 secuencias de novo que compensaron la delección de 4 diferentes genes necesarios para la viabilidad de E. coli. Estas secuencias permitieron su crecimiento. No obstante, las proteínas de novo mostraron niveles de actividad menores que las proteínas silvestres que reemplazaron. Con lo anterior podemos decir que si bien las proteínas de novo permiten la supervivencia del organismo, se espera que sean menos activas que las proteínas que se encuentran en la naturaleza y que son el resultado de millones de años de evolución.

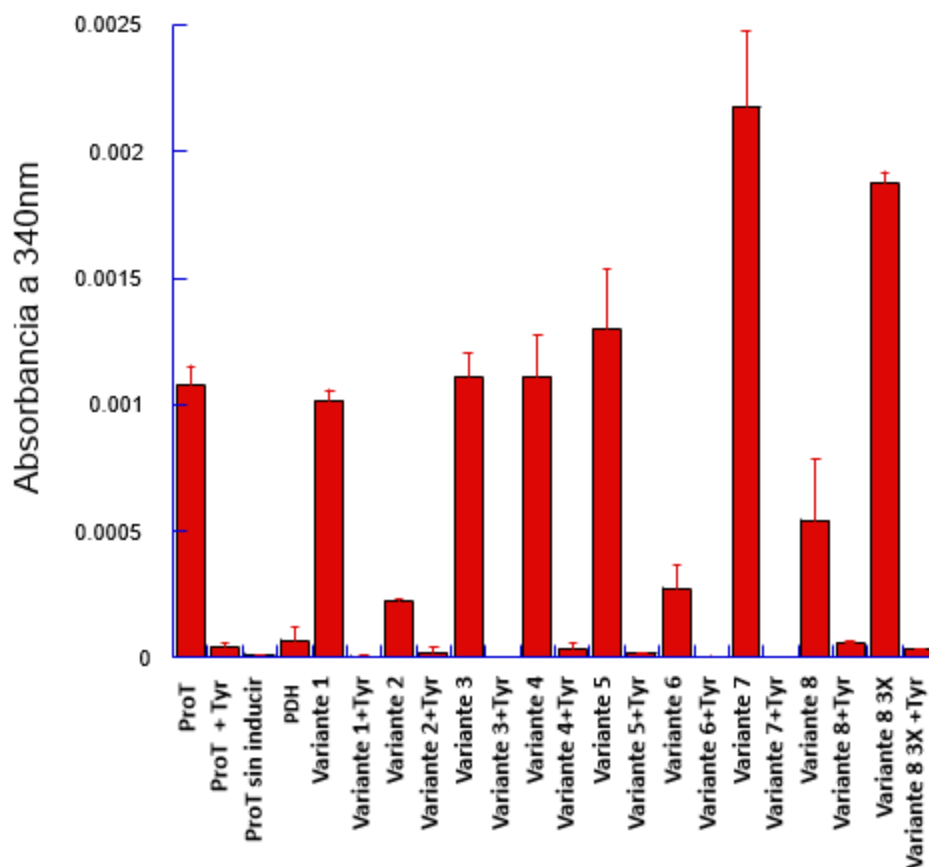


Figura 14. Actividad enzimática específica del dominio PDH modificado de las 8 variantes, a través de medir la aparición de NADH (acorde a la absorbancia determinada a 340 nm). Muestras empleadas: ProT corresponde a la proteína T (control positivo), ProT+Tyr corresponde a la proteína T con tirosina, PDH corresponde al dominio PDH silvestre (control negativo), +Tyr indica empleo de tirosina como posible inhibidor, Variante 8 3X corresponde a la muestra de la variante 8 con una cantidad de proteína 3 veces mayor que el resto de las variantes utilizadas. Los datos mostrados corresponden a la media y desviación estándar de dos unidades experimentales (n=2).

8.5 Determinación del banco de variantes

El tamaño del banco fue de 186,000 variantes. A pesar de ser un banco pequeño, esto en comparación con la cantidad de posibles variantes que se pueden obtener teóricamente (1×10^{13}) (Packer & Liu, 2015), permitió obtener variantes estables y funcionales.

9. Conclusiones y perspectivas.

Se demostró que la introducción de etiquetas de 6 aminoácidos al azar en la región amino terminal del dominio PDH de la proteína T es capaz de estabilizarlo, permitiendo su expresión y actividad catalítica, ésta última igual que la proteína T. El método de selección a pesar de no ser preciso ya que se seleccionaron falsos positivos, si permitió seleccionar variantes con la propiedad deseada.

Del banco de 186,000 variantes, solo 8 presentaron un PDH estable. Estas mostraron crecimiento, expresión (10 veces menor que la proteína T silvestre) y actividad enzimática específica similar a la proteína T silvestre. De lo anterior podemos decir que solo 1 de cada 25 000 variantes muestra actividad de PDH.

El presente trabajo, permitió lograr la estabilización del dominio PDH, así como observar su expresión y actividad catalítica, algo que no se había logrado en experimentos previos. Sin embargo, para entender mejor el funcionamiento del dominio es necesario ampliar la perspectiva de futuras hipótesis, si bien la etiqueta de 6 aminoácidos al azar pudo permitir la estabilización del dominio, al analizar las secuencias no encontramos ningún patrón que indique la preferencia hacia ciertos aminoácidos. Asimismo, experimentos previos han demostrado el rol importantísimo que tiene la dimerización para estabilizar al dominio PDH, al reemplazar el dominio de corismato mutasa por dominios que dimerizan (dominio $\beta 1$ de la proteína G, un motivo EF-hand y un dominio Rop). El presente trabajo deja la duda si las secuencias de 6 aminoácidos seleccionadas están estabilizando al dominio de la misma manera y de ser así, si es posible que secuencias más cortas tengan el mismo efecto.

10. Anexos

ANEXO I

Reactivos, materiales y equipo.

Reactivo/materiales/ equipo	Marca, País
Enzima HindIII	New England Biolabs (Ipswich, MA, USA)
Enzima NcoI	New England Biolabs (Ipswich, MA, USA)
Enzima Taq polimerasa	Purificada en el laboratorio
Buffer F (10X)	Thermo scientific, USA
Buffer ligasa	Thermo scientific, uSA
Buffer 2.1	New England Biolabs (Ipswich, MA, USA)
Buffer de carga (5x) para electroforesis SDS-PAGE	Los reactivos empleados para preparar el buffer de carga fueron adquiridos con Sigma USA, Bio-Rad USA, y J.T. Baker, Mex. Información sobre los reactivos se describe en el ANEXO V
Buffer de elución	10 mM Tris, pH 8.5, Sigma USA
Buffer de carga para electroforesis DNA 6X	Los reactivos empleados para preparar el buffer de carga para electroforesis DNA 6X fueron adquiridos con Sigma USA y J.T.Baker, Mex. Información sobre los reactivos se describe en el ANEXO V
Buffer de actividad	Los reactivos empleados para preparar el buffer de actividad fueron adquiridos con Sigma USA. Información sobre los reactivos se describe en el ANEXO V
Medio de cultivo mínimo M9 (MM M9)	Todos los reactivos empleados para formular este medio de cultivo fueron adquiridos con J.T. Baker, Mex. Información sobre los reactivos y preparación se describe en el ANEXO VI.
Medio de cultivo SOC	Todos los reactivos empleados para formular este medio de cultivo fueron adquiridos con J.T. Baker, Mex. y BD, USA. Información sobre los reactivos y preparación se describe en el ANEXO VI.

Medio de cultivo LB	Todos los reactivos empleados para formular este medio de cultivo fueron adquiridos con J.T. Baker, Mex. y BD, USA. Información sobre los reactivos y preparación se describe en el ANEXO VI.
Kit High Pure PCR Product Purification	Roche Applied Science, Indianapolis, IN, USA
Kit High Pure Plasmid Isolation	Roche Applied Science, Indianapolis, IN, USA
Vector pBadmyc-His A	Invitrogen, USA
Prefenato	Sigma, USA
NAD ⁺	Sigma, USA
Marcador de peso molecular de proteínas	Thermo Fisher Scientific, USA
Marcador de peso molecular de ADN	Thermo Fisher Scientific, USA
Trans-iluminador de luz UV de longitud alta (365 nm).	MacroVue UV-25. Hoefer SA
Vortex	Velp Scientifica, Europa
Centrifuga	Legend Micro 17, Thermo Scientific, Alemania

ANEXO II

Composición de buffers

Buffer	Composición
Buffer de carga (5x) para electroforesis SDS-PAGE	2.5 ml de 1M Tris pH 6.8 (Sigma, USA) 2.5 ml de 10% SDS (Bio-Rad, USA) 5 ml de glicerol al 100% (J.T. Baker, Mex) 100 µl de β-MEOH (Bio-Rad, USA)
Buffer de carga para electroforesis DNA 6X	0.25 % azul de bromofenol sigma USA 0.25 % xilen Cianol sigma USA 20 % Glicero J.T.Baker, Mex
Buffer de actividad	50 mM Tris-Cl pH 7.5 Sigma USA 0.1mg/mL BSA, Sigma USA

ANEXO III

Reacción de ligación concentrada por precipitación con butanol

Se le agrega 500 µl de butanol a la reacción de ligación y se mezcla en vortex, esto con el fin de quitar el agua de la reacción. La mezcla se centrifuga a 12,000 rpm durante 10 min. Se descarta la fase líquida con una micropipeta y en el fondo queda el ADN. El botón se pone a secar para eliminar restos de butanol en un secador al vacío a 60 °C durante 20 min. Finalmente, se resuspende en 10 µl de agua grado milliQ.

ANEXO IV

Transformación por electroporación de Células BW25113 (Δ tyrA) de *E. coli*

En condiciones estériles se toma una alícuota de 60 células electrocompetentes, se les agrega 2 µl de la reacción de ligación y se colocan en una celda para electroporar de 0.1 cm. El pulso del electroporador (MicroPulser) es de 1.8 kV, inmediatamente después se agregó 1 ml de medio SOC, posteriormente son transferidas a un tubo de cultivo estéril se dejan en una incubadora con agitación para que las células se recuperen durante 1 hrs a 30 °C/ 250 rpm.

ANEXO V

Composición y preparación de medios de cultivo.

A) Medio mínimo M9 (MM M9)

Composición	Preparación
-Sales M9 10X: Na ₂ HPO ₄ 30g/L KH ₂ PO ₂ 15g/L NH ₄ Cl 5g/L NaCl 2.5g/L -MgSO ₄ 1 M. -Glucosa 20%	Agregar los siguientes reactivos siguiendo el orden que se describe a continuación -Medio Sólido Para 1 L. 100 ml de Sales M9 10X 10 ml de Glucosa 20% 1 ml de MgSO ₄

<p>-Ampicilina (200 mg/ml)</p> <p>-Kanamicina (25 mg/ml)</p>	<p>1 ml de amp</p> <p>1 ml de km</p> <p>pH 7.0</p> <p>Agregar agar hasta llegar a un volumen de 1 litro</p> <p>-Medio líquido</p> <p>Para 1 L.</p> <p>Mismo procedimiento sólo cambiar el agar por agua estéril.</p> <p>Nota: se agrega 1 ml de tirosina (10 mg/ml), o 0.5 ml de arabinosa (0.2%) cuando el experimento lo requiera.</p>
--	--

B) Medio LB

Composición	Preparación
<p>Triptona 10 g/L BD, USA</p> <p>Extracto de levadura 5g/L BD, USA</p> <p>NaCl 5 g/L, J.T. Baker, Mex.</p>	<p>-Medio Sólido</p> <p>Para 1 L</p> <p>15 g de Bacto agar</p> <p>10 g de triptona</p> <p>5 g de extracto de levadura</p> <p>Ajustar el volumen de 1 L con agua desionizada.</p> <p>pH 7.0</p> <p>-Medio Líquido</p> <p>10 g de Bacto-triptona</p> <p>5 g de extracto de levadura</p> <p>10 g de NaCl</p> <p>950 ml de agua desionizada</p> <p>pH 7.0</p>

C) Medio SOC

Composición	Preparación
Tryptona 2 % BD, USA	-Medio líquido
Extracto de levadura 0,5 % BD, USA	Para 1 L
NaCl 10 mM J.T. Baker, Mex	20 g de Tryptona
KCl 2,5 mM J.T. Baker, Mex	5 g de Extracto de levadura
MgCl ₂ 10 mM J.T. Baker, Mex	10 ml de NaCl 1M
MgSO ₄ 10 mM J.T. Baker, Mex	2.5 ml de KCl 1M
Glucosa 20 mM J.T. Baker, Mex	5 ml de MgCl ₂
	20 ml de Glucosa
	Llevar a un litro con agua desionizada

ANEXO VI

Composición de disoluciones empleadas para electroforesis

A) Solución para teñir (Azul de Coomassie)

- 5% ácido acético
- 10% etanol
- 0.1% azul de Coomassie.
- Agua

B) Solución para desteñir

- 40 % etanol
- 7 % de ácido acético

11. Bibliografía.

- Alberts B., Johnson A., Lewis J., Raff M., Roberts K., & Walter P. (2010). Proteínas (M. L. i. S. Mercé Durfort i Coll, Trans.). *Biología molecular de la célula*. Vol. 5, pp. 125-152.
- Bongaerts J., Krämer M., Müller U., Raeven L., & Wubbolts M. (2001). Metabolic engineering for microbial production of aromatic amino acids and derived compounds. *Metabolic Engineering*. Vol. 3, pp. 289-300.
- Bornscheuer U., & Pohl M. (2001). Improved biocatalysts by direct evolution and rational protein design. *Current Opinion in Chemical Biology*. Vol 5 (2), pp. 137-143.
- Cortés-Totalpa L., Carmona B., Cervantes-Salinas A., Garcia M., Gosset G., Escalante A., & Bolivar F. (2011). Ingeniería de vías metabólicas en *Escherichia coli* para la producción de shikimate como precursor para la síntesis de compuestos antivirales contra influenza. *Revista de la Sociedad Mexicana de Biotecnología y Bioingeniería A.C.* Vol. 15, pp. 30-47.
- Chen S., Vincent S., Wilson D., & Ganem B. (2003). Mapping of chorismate mutase and prephenate dehydrogenase domains in the *Escherichia coli* T-protein. *The FEBS Journal*. Vol. 270(4), pp. 757-763.
- Fisher M., McKinley K., Bradley L., Viola S., & Hecht M. (2011). *De novo* designed proteins from a library of artificial sequences function in *Escherichia coli* and enable cell growth. *Plos one*, Vol. 6(1), pp. 1-9.
- Hoegler K., & Hecht M. (2016). A *de novo* protein confers copper resistance in *Escherichia coli*. *The Protein Society*. Vol. 2, pp. 1249-1259.
- Horton R., Moran L., Scrimger G., Perry M., & Rawn D. (2008). Proteínas: Estructura tridimensional y función (G. y. P. Virgilio, Trans.). In F. R. Rubén (Ed.), *Principios de Bioquímica* (4 ed.). México
- Kedar P., Colah Roshan., & Shimizu K. (2007). Proteomic investigation on the pyk-F gene knockout *Escherichia coli* for aromatic amino acid production. *Enzyme and Microbial Technology*, Vol. 41(4), pp. 455-465.

- Lodish H., Berk A., Matsudaira P., Kaiser A., Krieger M., Scott M., Darnell J. (2005). *Biología Celular y Molecular* (S. N., Trans. 5 ed.). México, Médica Panamericana.
- Lutke-Eversloh T. & Stephanopoulos G. (2007). L-tyrosine production by deregulated strains of *Escherichia coli*. *Applied Microbiology and Biotechnology*. Vol. 75(1), pp. 103-110.
- Nelson L. D. & Cox M. M. (2008). *Principles of Biochemistry* (K. Ahr Ed. 5 ed.). United States of America.
- Ostermeier M., Hoon J. & Benkovic J. (1999). A combinatorial approach to hybrid enzymes independent of DNA homology. *Nature Biotechnology*. Vol. 17, pp. 1205-1209.
- Osuna J., Flores H. & Gaytan P. (2012). A reporter system that discriminates EF-hand-sensor motifs from signal-modulators at the single-motif level. *FEBS Letters*, Vol. 586(19), pp. 3398-3403.
- Osuna J., Flores, H. & Saab-Rincon G. (2012). The beta1 domain of protein G can replace the chorismate mutase domain of the T-protein. *FEBS Letters* Vol. 586(4), pp. 466-471.
- Packer M. & Liu D. (2015). Methods for the directed evolution of proteins. *Nature Reviews Genetics*. Vol. 16, pp. 379-393.
- Patrick W., Quandt E., Swartzlander D. & Matsumura I. (2007). Multicopy suppression underpins metabolic evolvability. *Molecular Biology and Evolution*. Vol. 24, pp. 2716-2722.
- Pittard J. & Yang J. (2013). Biosynthesis of the aromatic amino acids. *EcoSal Plus*. Vol 3(1) pp. 1-40.
- Segovia L. & Peimbert. (2010). *Ingeniería de proteínas y evolución dirigida*. Disponible en:
http://bq.unam.mx/wikidep/uploads/MensajeBioquimico/Mensaje_Bioq10v34p135-142_LorenzoSegovia2010.pdf
- Soberón X. & Montero G.M. (2006). *Ingeniería de proteínas y evolución dirigida*. Disponible en: http://fenix.cichcu.unam.mx/libroe_2006/1038967/11_c07.pdf

- Soberón X. & Segovia L. (2007). Evolución experimental de proteínas. Disponible en http://www.ibt.unam.mx/computo/pdfs/libro_25_aniv/capitulo_26.pdf
- Spiegelman S., Haruna I., Hollands B., Beaudreau G. & Mills D. (1965). The synthesis of a self-propagation and infectious nucleic acid with purified enzyme. *Proceedings of the National Academy of Sciences*, Vol. 54, pp. 919-927.
- Sun W., Singh S., Zhang R., Turnbull J. L., & Christendat D. (2006). Crystal structure of prephenate dehydrogenase from *Aquifex aeolicus*. Insights into the catalytic mechanism. *Journal of Biological Chemistry*. Vol. 281(18), pp. 12919-12928.
- Vinogradov E. A. (2003). DNA helix: the importance of being GC-rich. *Nucleic Acids Research*. 37(7), pp.1838–1844.
- Watson D. J., Baker A. T., Bell P. S., Gann A., Levine M. & Losick R. (2014). The Structure of Proteins In W. Beth (Ed.), *Molecular Biology of the Gene*. 7 ed., pp. 121-145.