

# UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO



Facultad de Medicina

Instituto de Investigaciones Biomédicas

Instituto de Fisiología Celular

## LICENCIATURA EN INVESTIGACIÓN BIOMÉDICA BÁSICA

Biología Sintética: Modelo de estudio de mecanismos de recombinación genética codificante para proteínas con potencial terapéutico, producidas en *E. coli*.

Tesis que para obtener el grado de Licenciado en Investigación Biomédica Básica presenta Daniel Isui Aguilar Salvador.

Directora de tesis:

Dra. Norma Adriana Valdez Cruz

Ciudad Universitaria, México D.F. 2012

*El presente trabajo fue realizado bajo la asesoría de la Dra. Norma Adriana Valdez Cruz en su laboratorio, perteneciente al Departamento de Biología Molecular y Biotecnología del Instituto de Investigaciones Biomédicas de la Universidad Nacional Autónoma de México, Ciudad Universitaria, México, Distrito Federal. Durante la realización de este proyecto se contó con los apoyos económicos de CONACyT Ciencia Básica 104951-Z, 103393, 82533, al proyecto CONACYT INNOVAPYME-137854, a los proyectos PAPIIT-UNAM IN28509-3 y 210013-2 otorgados a la Dra. Valdez.*

## **Agradecimientos**

A la Dra. Adriana Valdez Cruz por aceptar dirigir este proyecto.

Al Dr. Mauricio Trujillo Roldán por su apoyo financiero y académico.

A la Dra. María de los Ángeles Cancino por su asistencia técnica en la ejecución de algunos experimentos.

Al Q. Jorge Yáñez de la Unidad de Secuenciación del Instituto de Biotecnología-UNAM.

Al Taller de Biología Sintética de la Facultad de Ciencias-UNAM por la donación de los BioBricks necesarios para el desarrollo del proyecto.

Al Dr. Guillermo Gosset Lagrada por la donación del plásmido pJW168.

A Isabel Ángeles Santander, Gilberto Gómez Correa y Román Alfonso Castillo Diaz por sus aportaciones teóricas durante el planteamiento inicial del proyecto.

A los proyectos CONACYT 103393, 82533, 104951-Z; al proyecto CONACYT INNOVAPYME 137854; al proyecto PAPIIT-UNAM IN28509-3 y al proyecto DGAPA 210013-2.

A mi familia, a mis amigos, a Isabel

Claro que, observa el que te dije, a pesar de ese obstruccionismo subjetivo el tema subyacente es muy simple: 1) La realidad existe o no existe, en todo caso es incomprendible en su esencia, así como las esencias son incomprendibles en la realidad, y la comprensión es otro espejo para alondras, y la alondra es un pajarito, y un pajarito es el diminutivo de pájaro, y la palabra pájaro tiene tres sílabas, y cada sílaba tiene dos letras, y así es como se ve que la realidad existe (puesto que alondras y sílabas) pero que es incomprendible, porque además qué significa significar, o sea entre otras cosas *decir* que la realidad existe; 2) La realidad será incomprendible pero existe, o por lo menos es algo que nos ocurre o que cada uno hace ocurrir, de manera que una alegría, una necesidad elemental lleva a olvidar todo lo dicho (en 1) y pasar a 3) Acabamos de aceptar la realidad (en 2), sea lo que sea o como sea, y por consiguiente aceptamos estar instalados en ella, pero ahí mismo sabemos que, absurda o falsa o trucada, la realidad es un fracaso del hombre aunque no lo sea del pajarito que vuela sin hacerse preguntas y se muere sin saberlo. Así, fatalmente, si acabamos de aceptar lo dicho en 3), hay que pasar a 4) Esta realidad, a nivel de 3), es una estafa y hay que cambiarla.

*Libro de Manuel, Julio Cortázar*

## Resumen

Los anticuerpos son las moléculas de reconocimiento específico más usadas en investigación, diagnóstico e incluso con fines terapéuticos. Sin embargo, los costos asociados tanto al desarrollo como a la producción de nuevos anticuerpos aún son altos, debido principalmente a su complejidad estructural. Además de los anticuerpos se han planteado otro tipo de moléculas de reconocimiento específico. Por un lado se encuentran proteínas derivadas de dominios de inmunoglobulinas de diversas especies y por otro se encuentran las “plataformas proteicas diseñadas”. Estas últimas se derivan de dominios proteicos sin relación con las inmunoglobulinas, por lo que ofrecen propiedades estructurales diferentes a estas últimas; además de que pueden producirse de manera recombinante con mayor facilidad debido a su menor tamaño y a que no requieren de modificaciones post-traduccionales.

En esta tesis se presenta el diseño de un sistema de recombinación somática sintética en *Escherichia coli* desde el enfoque de la biología sintética. El sistema fue diseñado para imitar el proceso celular de generación de diversidad de los anticuerpos -la recombinación VDJ-, sobre la base estructural de las DARPinas. El sistema se compone de un circuito genético diseñado a partir del mecanismo de recombinación Cre/*loxP* y se diseñó para generar hasta 18 secuencias diferentes.

Como parte del diseño se obtuvo un modelo de la estructura tridimensional de una de las proteínas resultantes que sugiere que la estructura terciaria de la DARPina no se verá afectada por la inserción de sitios *loxP*. En el ensamblaje de las secuencias se utilizaron biopartes (*Biobricks*) y se utilizó el ensamblaje estándar. Aunque, en principio, puede aumentarse el tamaño del Módulo Recombinante para producir una mayor diversidad. Aunque inicialmente el ensamblaje de las secuencias diseñadas fue exitoso, el ensamblaje del módulo recombinante propuesto no se logró. Tampoco se logró comprobar la recombinación *in vitro* de las secuencias diseñadas. Lo anterior posiblemente debido a la formación de estructuras secundarias en ADN, producidas por su alta repetitividad, que no permitieron el ensamble y la recombinación.

## Índice

Introducción .....	15
Biología sintética .....	15
Desacoplamiento, abstracción y estandarización .....	17
Parte, dispositivo, sistema y chasis.....	18
Ensamblaje Estándar .....	20
Anticuerpos .....	22
Estructura de los anticuerpos.....	23
Generación de variabilidad en los anticuerpos .....	25
Recombinación V(D)J .....	25
Antecedentes .....	28
Plataformas proteicas diseñadas .....	29
DARPinas.....	33
Sistemas de recombinación sitio-especificos .....	35
Serina-recombinasas.....	36
Tirosina-recombinasas .....	37
El sistema Cre/loxP.....	39
Modificaciones al sistema Cre/loxP.....	40
Alternativas celulares de la biología sintética para la generación y selección de proteínas con reconocimiento específico .....	44
Hipótesis.....	46
Objetivo general .....	46
Objetivos específicos.....	46
Metodología .....	47
Obtención y alineamiento de secuencias de DARPinas .....	47
Transcripción in silico, retrotranscripción y optimización de secuencias .....	47
Predicción de estructura terciaria y visualización de estructuras tridimensionales .....	47

Preparación de E.coli TOP 10 quimiocompetentes.....	48
Transformación de E. coli TOP 10 quimiocompetentes.....	48
Síntesis de ADN .....	48
Secuenciación de ADN.....	49
Purificación de ADN plasmídico .....	49
Digestión de ADN .....	49
Ligación de ADN .....	49
Ensayo de recombinación <i>in vitro</i> .....	49
Resultados .....	52
Diseño de la estructura general del sistema de recombinación.....	52
Identificación de módulos constantes y variables de las DARPinas .....	54
Selección de secuencias loxP heterólogas.....	58
Predicción de estructura terciaria de los productos de recombinación .....	59
Selección de promotor, señal de localización a membrana y sitio de unión a ribosoma .....	62
Secuencias finales de las regiones C1, V1 y V2+C2 .....	63
Síntesis de secuencias .....	65
Ensamblaje de secuencias .....	67
1) Ensamblaje de las tres regiones en un vector mediante ligación y amplificación <i>in vitro</i> .....	67
2) Ensamblaje de las regiones C1 y V2+C2 en un vector .....	72
3) Ensayo de recombinación <i>in vitro</i> .....	74
Discusión .....	76
Conclusiones.....	78
Perspectivas.....	79
Bibliografía.....	80
Anexo 1: Secuencia de aminoácidos de las DARPinas usadas para el diseño de los módulos constantes y variables.....	89
Anexo 2: Secuencias nucleotídicas utilizadas del Registry of Standard	



Biological Parts .....	90
Anexo 3: Modelos estructurales para las proteínas finales con diferentes combinaciones de secuencias loxP heterólogas generados por I-TASSER.....	91
Anexo 4: Alineamiento de la secuencia obtenida con la secuencia V1 .....	94
Anexo 5: Artículo de revisión publicado en la revista BioTecnología.....	95

## Abreviaturas

AAs	Aminoácidos
AID	<i>Activation-Induced Deaminase</i>
C1	Región constante 1
C2	Región constante 2
CDR	<i>Complementarity Determining Region</i>
DNA-PK	DNA-activated protein kinase
<i>E. coli</i>	<i>Escherichia coli</i>
Fab	<i>Fragment antigen binding</i>
Flp	Flippase recombination enzyme
Fv	<i>Fragment variable</i>
GFP	<i>Green Fluorescent Protein</i>
iGEM	competencia internacional de Máquinas Genéticamente Modificadas ( <i>international Genetically Engineered Machine competition</i> )
KDa	Kilo Daltones
NHEJ	<i>Non-homologous end joining</i>
OmpA	Proteína de membrana externa A
PDB	<i>Protein Data Bank</i>
PDB ID	Identificador del <i>Protein Data Bank</i>
PCR	Reacción en cadena de la polimerasa ( <i>Polymerase Chain Reaction</i> )
PLacI	Promotor regulado por lactosa
Rag	Recombination activating gene
RBS	Sitio de unión a ribosoma
RFP	<i>Red Fluorescent Protein</i>
RMCE	Intercambio de casete mediado por recombinasa ( <i>Recombinase-Mediated Cassette Exchange</i> )
RSE	Recombinasas sitio-específicas

Ter	Terminador de la transcripción
TdT	Terminal deoxynucleotidyl transferase
V1	Región variable 1
V2+C2	Región variable 2 + constante 2
XRCC4	<i>X-ray repair cross-complementing protein 4</i>

## Lista de Figuras

- Figura 1.** “Represilador de Elowitz”.
- Figura 2.** Secuencias del prefijo y del sufijo.
- Figura 3.** Esquema del ensamblaje estándar para dos *Biobricks*.
- Figura 4.** Estructura tridimensional de un anticuerpo IgG1.
- Figura 5.** Esquema del proceso de Recombinación V(D)J en el locus de la cadena pesada.
- Figura 6.** Esquema de polipéptidos derivados de anticuerpos.
- Figura 7.** Estructura tridimensional de distintas plataformas proteicas diseñadas.
- Figura 8.** Módulos que conforman una DARPina.
- Figura 9.** Esquema del mecanismo de reacción de RSE.
- Figura 10.** Estructura de la secuencia loxP WT
- Figura 11.** Secuencia *lox P* tipo (WT) y secuencias *lox P* mutadas (511, 511-I, FAS, 2272 y 5171).
- Figura 12.** Esquema del RMCE.
- Figura 13.** Esquema del sistema LE/RE.
- Figura 14.** Esquema del sistema de selección de proteínas capaces de interactuar con otra proteína blanco del equipo de la Universidad de Varsovia (iGEM 2008).
- Figura 15.** Estructura general del Sistema de Recombinación.
- Figura 16.** Esquema del sistema de recombinación con la distribución de las secuencias señales en tres diferentes regiones de ADN.
- Figura 17.** Alineamiento de las DARPinas 2QYJ, 2V5Q:D, 2P2C:R y 1SVX:A
- Figura 18.** Ubicación de los residuos variables en las DARPinas
- Figura 19.** Estructura de las secuencias de proteínas finales con diferentes combinaciones de secuencias *loxP* heterólogas.
- Figura 20.** Esquema de la región C1.
- Figura 21.** Comprobación mediante PCR de transformación en *E. coli* de secuencias

sintetizadas comercialmente.

**Figura 22.** Ejemplo del proceso de ensamblaje *in vitro* para la unión de la región C1 y la RFP.

**Figura 23.** Ensamblaje de las tres regiones en un vector mediante ligación y amplificación *in vitro*.

**Figura 24.** Amplificación de secuencias ensambladas *in vitro*.

**Figura 25.** Esquema del Ensamblaje de las regiones C1 y V2+C2 en un vector.

**Figura 26.** Amplificación de la ligación V1-V2+C2.

**Figura 27.** Estrategia de recombinación *in vitro*.

## Lista de Tablas

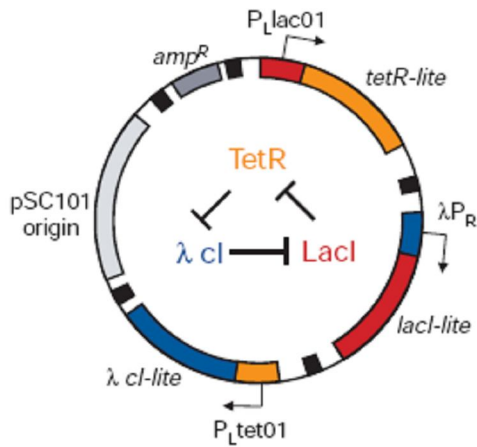
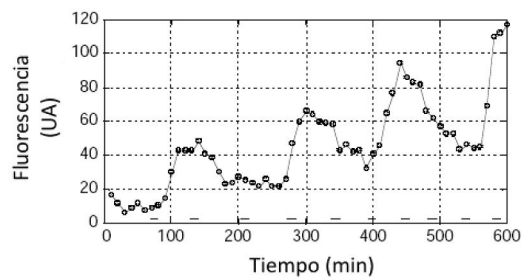
- Tabla 1.** Eficiencia de recombinación entre distintas secuencias *loxP*.
- Tabla 2.** Módulos constantes y variables de las DARPinas.
- Tabla 3.** Traducción *in silico* de secuencias *loxP* heterólogas.
- Tabla 4.** Selección de secuencias *loxP* heterólogas de acuerdo a su eficiencia de recombinación.
- Tabla 5.** Estructura de las secuencias de proteínas finales con diferentes combinaciones de secuencias *loxP* heterólogas.
- Tabla 6.** Comparación estructural entre las diferentes secuencias de proteínas finales y la DARPina consenso 2QYJ.
- Tabla 7.** Detalle de los plásmidos en que se recibieron las secuencias sintetizadas.

# I Introducción

## I.1 Biología Sintética

El desarrollo de técnicas que facilitan la manipulación del material genético a gran escala, desde el ADN recombinante hasta la síntesis de genomas enteros, ha permitido el desarrollo de un nuevo enfoque en las ciencias biológicas, pasando de la modificación de genes o vías metabólicas de manera aislada, al diseño de sistemas biológicos artificiales con un mayor número de elementos funcionales y de regulación (Andrianantoandro *et al.* 2006; Gibson *et al.* 2010).

La biología sintética se entiende ahora como el diseño y construcción de nuevas partes, dispositivos y sistemas biológicos; así como el rediseño de sistemas biológicos naturales ya existentes con un fin determinado ([www.syntheticbiology.org](http://www.syntheticbiology.org)). Bajo este enfoque basado en el diseño y construcción sistemática de sistemas biológicos se han creado ya un sinnúmero de este tipo de sistemas. Uno de los trabajos fundacionales de esta disciplina es el “repretilador de Elowitz” (Elowitz y Leibler 2000), que consiste de una red genética oscilatoria completamente artificial diseñada a partir de tres sistemas de represión transcripcional, obteniendo un ciclo ultradiano en la expresión de la proteína reportera (GFP), con una duración mayor al tiempo de duplicación celular (Figura 1). Con este trabajo, Elowitz y Leibler mostraron la posibilidad de diseñar y construir redes genéticas artificiales con propiedades funcionales nuevas a partir de componentes naturales.

**A****B**

**Figura 1.** “Represilador de Elowitz”. **A.** Red genética del represilador, consta de tres represores y sus promotores correspondientes. En este arreglo, el primer represor reprime la expresión del segundo, el cuál reprime la expresión del tercero, que a su vez reprime al primero. **B.** Oscilaciones observadas *in vivo* mediante fluorescencia de la proteína GFP, que se utilizó como reportero del promotor  $P_{Ltet01}$ . (Modificado de Elowitz y Leibler 2000)

El principal objetivo de la biología sintética es sistematizar y facilitar la ingeniería genética, apoyándose en la modularidad de los sistemas biológicos (Wolf y Arkin 2003). Para lograr esto se han desarrollado nuevas herramientas computacionales y moleculares que permiten el modelado y disección de sistemas biológicos; así como el ensamblaje y caracterización sistemática de nuevos sistemas biológicos. Ejemplos de herramientas computacionales son el *Standard Virtual Biological Parts* (Cooling *et al.*, 2010), BioJADE (<http://web.mit.edu/jagoler/www/biojade/>, 2012), SynBioSS (<http://synbioss.sourceforge.net/>, 2012), GenoCAD (<http://www.genocad.org>, 2012) y ClothoCAD (<http://www.clothocad.org/>, 2012); estas herramientas buscan conjuntar herramientas de modelado de sistemas biológicos con el acceso a bases de datos con información biológica, para facilitar el diseño de nuevos sistemas biológicos.

De acuerdo a Drew Endy (2005), el interés por la biología sintética lo conducen principalmente cuatro grupos de investigadores: biólogos, químicos, “imitadores” e



ingenieros. Los biólogos buscan métodos directos y eficaces para poner a prueba el conocimiento actual de los sistemas biológicos. Los químicos aprovechan la habilidad de los seres vivos de sintetizar nuevas moléculas, exploran nuevos sistemas genéticos o utilizan un enfoque sintético para estudiar el origen de la vida. A los “imitadores” les interesa producir sistemas biológicos alternativos a los existentes, optimizados para algún fin en particular. Para los ingenieros, la vida es una tecnología más, en la que aún se debe trabajar para facilitar el diseño y construcción de nuevos sistemas biológicos.

### **I.1.1 Desacoplamiento, abstracción y estandarización**

Desde el enfoque ingenieril de la biología sintética, se hace hincapié en tres conceptos usados por otras ingenierías: desacoplamiento (*decoupling*), abstracción (*abstraction*) y estandarización (*standardization*). El desacoplamiento es la idea de que un problema de diseño complicado puede separarse en problemas más sencillos que se pueden resolver de manera independiente. La abstracción es una manera de manejar la complejidad, por ejemplo, en los sistemas biológicos se pueden observar comportamientos diferentes a escalas diferentes. Estas escalas se pueden separar conceptualmente unas de otras: se puede comprender el comportamiento de una población sin necesidad de saber todos los detalles del sistema a nivel molecular. Por último, la estandarización consiste en establecer procesos y herramientas que permitan la descripción, el diseño y construcción sistemática de sistemas (Hartwell, *et al.* 1999, Endy 2005, Wolf y Arkin 2003).

Uno de los primeros pasos que ha tomado la biología hacia un enfoque ingenieril es la estandarización de datos biológicos, que se ha producido como resultado de las grandes cantidades de información que es posible adquirir actualmente. Ejemplo de ello son los estándares para información de secuencias de ADN (<http://insdc.org/>), datos de microarreglos (Brazma *et al.* 2001), datos estructurales (<http://www.rcsb.org/pdb/home/home.do>) y modelos en biología de sistemas ([http://sbml.org/Main\\_Page](http://sbml.org/Main_Page)). Sin embargo, aun faltan por desarrollarse y refinarse estándares para medir funciones biológicas; como actividad de promotores, eficiencia de

secuencias terminadoras de la transcripción y de la traducción, fondo genético, condiciones ambientales, condiciones de crecimiento, entre otros (Hartwell, *et al.* 1999, Endy 2005).

La estandarización de la información sobre los procesos y partes biológicas, así como la manera en que se adquiere esta información, resultaría en la acumulación de información útil para todo aquel que esté interesado en diseñar nuevos sistemas biológicos. Además de que podrían identificarse y abstraerse ciertos aspectos del comportamiento de los sistemas biológicos, resultando en la generación de bibliotecas de partes biológicas que pudieran utilizarse fácilmente (Canton *et al.* 2008).

### **I.1.2 Parte, dispositivo, sistema y chasis**

Un resultado de la aplicación de estos conceptos ingenieriles a la biología es el desarrollo del *Registry of Standard Biological Parts*, una colección abierta de partes genéticas funcionales, con más de 18 mil *Biobricks* (nombre que reciben las partes biológicas). El objetivo de esta colección es generar una base de datos de partes biológicas estandarizadas, donde se describe a detalle el comportamiento de cada secuencia de ADN. Además de esto, las distintas biopartes, o *Biobricks*, se pueden ensamblar sistemáticamente mediante técnicas bien establecidas de biología molecular ([http://partsregistry.org/Main\\_Page](http://partsregistry.org/Main_Page)).

En este registro se manejan ciertas jerarquías que resultan de la abstracción del funcionamiento de sistemas biológicos. Una "parte" (*part*) es un objeto genéticamente codificado que realiza una función biológica y que ha sido diseñada para cumplir con ciertas especificaciones o desempeñarse de cierta manera bajo condiciones dadas (Canton *et al.* 2008). Una parte puede ser un promotor, una secuencia codificante para proteína o para algún ARN funcional, un terminador de la transcripción, un sitio de unión al ribosoma, entre muchos otros.

Es posible ensamblar dos o más partes siguiendo algún estándar de ensamblaje para codificar una función específica. Al resultado de este ensamblaje se le llama “dispositivo” (*device*) (Canton *et al.* 2008). Estos dispositivos pueden catalogarse de acuerdo a su función, muchas de las cuales son funcionalmente equivalentes a dispositivos diseñados en otras ingenierías, como la mecánica y la electrónica. Ejemplos de esto son dispositivos equivalentes a sensores (French *et al.* 2007), puertas lógicas (Anderson *et al.* 2007) o actuadores (Shaner *et al.* 2005).

Es deseable que tanto partes y dispositivos se caractericen de manera cuantitativa para conocer distintas propiedades, como desempeño, respuesta ante los estímulos de entrada, tiempo de respuesta, compatibilidad con otras partes y dispositivos, confiabilidad, demanda transcripcional y requerimientos de fondo genético (Canton *et al.* 2008). Esta caracterización facilita su uso entre distintos grupos de trabajo y su utilización para construir elementos de la siguiente jerarquía: sistemas.

Un (*system*) sistema es un conjunto de dispositivos y partes que, una vez ensamblados, cumplen con un propósito específico (Canton *et al.* 2008). Si los dispositivos se encuentran bien caracterizados y se tiene un conocimiento razonable de la influencia de parámetros externos sobre el sistema diseñado (como fondo genético, condiciones de cultivo, entre otros) los sistemas pueden comportarse de manera predecible, como sucede al construir sistemas electrónicos y mecánicos.

Parte central de la caracterización de partes, dispositivos y sistemas es el fondo biológico sobre el que se construye, a esto se le conoce como “chasis” (*chassis*). Un chasis es un organismo, o bien ciertas condiciones bioquímicas para el caso del sistema libre de células, que puede usarse para sustentar y suministrar poder a un dispositivo biológico diseñado en particular (Canton *et al.* 2008). Los chasis que se encuentran en el *Registry of Standard Biological Parts* son tres: *Escherichia coli*, *Bacillus subtilis*, y el sistema libre de células (<http://partsregistry.org/Catalog>). Aunque la mayoría de las biopartes de este

catálogo se han creado y probado sólo en *E. coli*.

Un chasis ideal sería un organismo que fuera capaz de replicarse de manera eficiente bajo ciertas condiciones de cultivo y cuyo fondo genético tuviera la menor complejidad posible, con el fin de reducir las posibles interacciones de este con los sistemas que se le introducen (Heinemann y Panke, 2006). La bacteria *E. coli* es el chasis biológico más utilizado en biología sintética. Esto se debe a que es uno de los organismos modelo más utilizados en biología molecular y celular, y por lo tanto se tiene un amplio conocimiento de su genética, fisiología y metabolismo. Tiene además un tiempo de duplicación corto comparado con otros procariontes y líneas celulares eucariontes, presenta bajo riesgo de trabajo, y puede ser manipulado fácilmente, pues existen muchas herramientas moleculares para lograrlo ([http://partsregistry.org/Escherichia\\_coli\\_chassis](http://partsregistry.org/Escherichia_coli_chassis)).

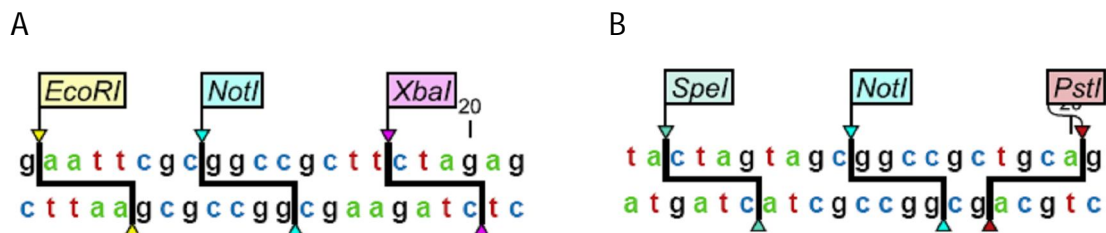
### **I.1.3 Ensamblaje Estándar**

La estandarización de una parte se refiere a las características que una secuencia funcional de ADN debe tener para poder ser ensamblada con otra secuencia de manera sistemática y a ciertos detalles sobre su comportamiento (tasa de transcripción, señales de entrada, señales de salida, requerimientos de fondo genético, entre otros). El “ensamblaje estándar” (*standard assembly*) es la serie de pasos que se siguen para ensamblar a dos o más partes que ya han sido estandarizadas.

Este tipo de ensamblaje se realiza usando técnicas normales de biología molecular (corte con enzimas de restricción, purificación de ADN, ligación y transformación). Las partes estandarizadas contenidas en el *Registry of Standard Biological Parts* están diseñadas para ser “componibles” (*composable*), es decir que, al menos en principio, pueden ensamblarse cualquier número de partes y en cualquier orden.

Para poder utilizar el ensamblaje estándar, cada secuencia funcional de ADN debe estar flanqueada por dos secuencias nucleotídicas denominadas “prefijo” (*prefix*) y “sufijo”

(*suffix*) ([http://partsregistry.org/Help:BioBrick\\_Prefix\\_and\\_Suffix](http://partsregistry.org/Help:BioBrick_Prefix_and_Suffix)); ambos son secuencias de ADN que contienen sitios de restricción de algunas enzimas comerciales (Figura 2). El Prefijo que se usa en los *Biobricks* incluye los sitios de restricción para las enzimas EcoRI, NotI y XbaI (Figura 2); y se ubica río arriba, en el extremo 5' de la secuencia funcional. El Sufijo incluye sitios de restricción para las enzimas SpeI, NotI y PstI, y se encuentra río abajo de la parte. Es muy importante que el resto de la cadena de ADN no contenga sitios de restricción para ninguna de estas enzimas. La parte, junto con su prefijo y sufijo, puede estar contenida en un plásmido con algún marcador de selección.

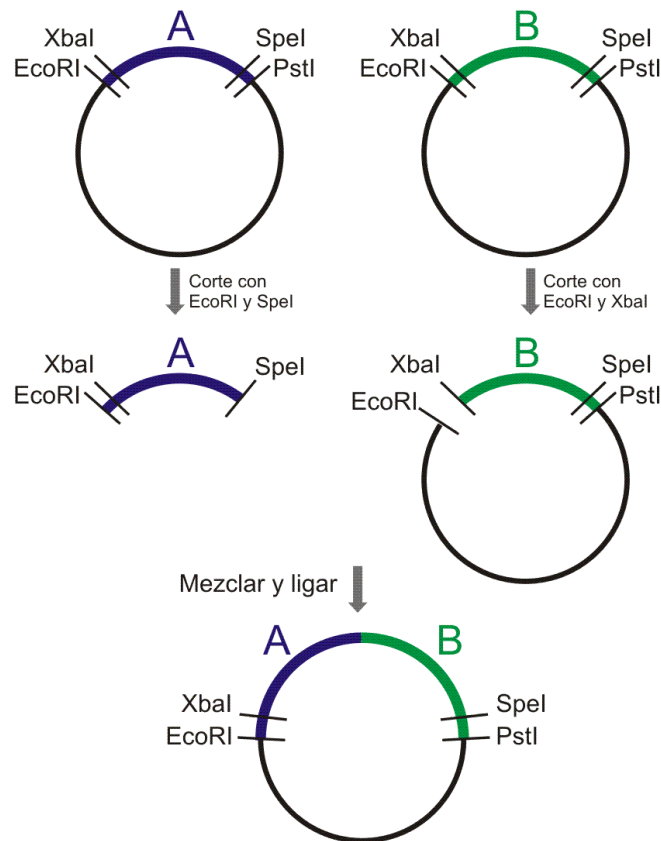


**Figura 2.** Secuencias del prefijo y del sufijo. **A.** Secuencia del prefijo, se indican los sitios de corte de las enzimas de restricción. **B.** Secuencia del sufijo, se indican los sitios de corte de las enzimas de restricción.

Gracias a este arreglo de sitios de restricción se puede ensamblar una parte estándar “A” con otra parte “B”, de manera que B quede río abajo de A (Figura 3). Para lograrlo, se debe digerir la parte B con las enzimas EcoRI y XbaI, de manera que quede abierto el vector con los extremos cohesivos que dejan estas enzimas. Por otro lado, la parte A se debe digerir con EcoRI y SpeI, de manera que se separe la parte A del plásmido. Posteriormente, se purifica el vector abierto con la parte B, y la parte A libre del plásmido.

El extremo cohesivo que deja XbaI es complementario al extremo cohesivo que deja SpeI, de manera que al ligar el vector con la parte B, y la parte A, esta se inserta en el vector que contiene a B. Durante este proceso se eliminan los sitios de restricción y como resultado se obtiene una nueva parte “AB”, flanqueada por un prefijo y un sufijo. Esta nueva parte puede a su vez ensamblarse con otras partes de la misma manera para

ensamblar un dispositivo.



**Figura 3.** Esquema del ensamblaje estándar para dos *BioBricks* (Modificado de <http://partsregistry.org/Help:Assembly>). Para ensamblar dos *BioBricks*, se extrae el *BioBrick* A de su vector con las enzimas EcoRI y SpeI, por otro lado se abre el vector que contiene al *BioBrick* B con las enzimas EcoRI y XbaI. Finalmente, se mezcla el *BioBrick* A y el vector abierto con el *BioBrick* B. Los extremos cohesivos que dejan las enzimas SpeI y XbaI son complementarios entre sí, y al ligarlos se eliminan los sitios de restricción de estas enzimas.

## I.2 Anticuerpos

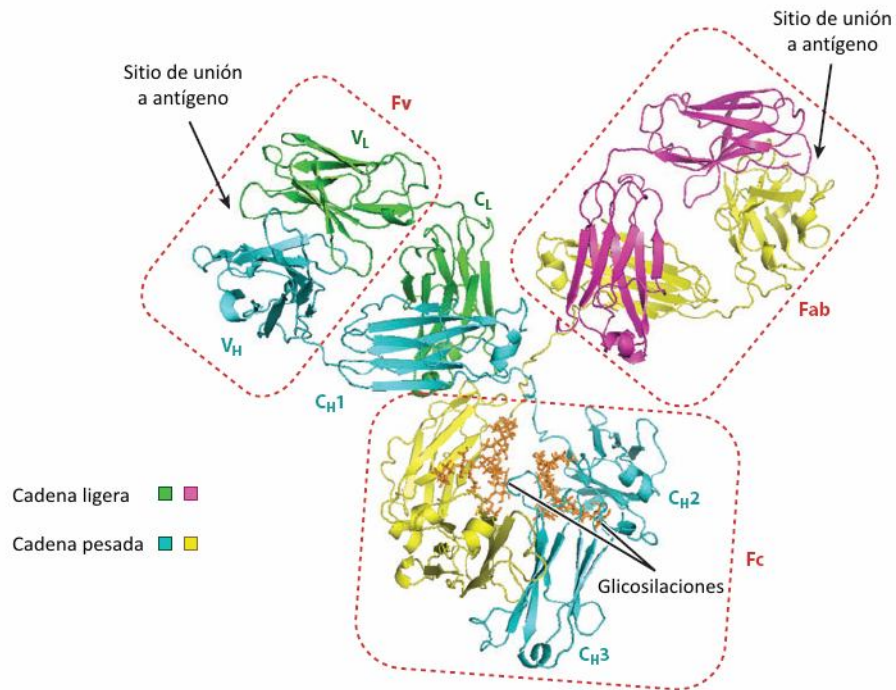
Los anticuerpos han tenido éxito como moléculas de reconocimiento debido a la alta afinidad que pueden presentar por un antígeno o un hapteno, alcanzando afinidades del orden nanomolar o incluso picomolar. Otra razón de su éxito, es la relativa facilidad con la que hoy en día se puede generar un anticuerpo específico; ya sea mediante inmunización

de animales y la posterior generación de hibridomas, o bien, mediante selección *in vitro* de librerías genéticas mediante despliegue en fagos (Hoogenboom y Chames, 2000).

Hoy en día, está bien establecida la eficacia de los anticuerpos como moléculas de reconocimiento y la seguridad de su uso como moléculas terapéuticas. También se encuentran bien descritas sus propiedades de especificidad, inmunogenicidad, farmacocinética y funciones efectoras relacionadas con el sistema inmune (Lonberg 2008, Boder y Jiang 2011, Steinmeyer y McCormick 2008, Holliger y Hudson 2005). Más de 20 anticuerpos diferentes han sido aprobados para su uso terapéutico en Estados Unidos y Europa, y un sinnúmero son usados diariamente en diagnóstico e investigación básica (Boder y Jiang 2011, Holliger y Hudson 2005).

### **I.2.1 Estructura de los anticuerpos**

Un anticuerpo tiene una estructura base formada por dos cadenas ligeras y dos cadenas pesadas, ambos pares idénticos entre sí (Figura 4). Tanto las cadenas pesadas como las ligeras constan de una región amino-terminal variable y una carboxilo-terminal constante. Las regiones variables difieren en su secuencia de aminoácidos entre distintas clonas de linfocitos B, y son estas secuencias las que participan en el reconocimiento específico de los antígenos.



**Figura 4.** Estructura tridimensional de un anticuerpo IgG1. Se muestran los dominios variables (V) y constantes (C) de las cadenas pesadas (H) y ligeras (L). Fab, fragmento de unión a antígeno. Fc, fragmento cristalizable. Fv, fragmento variable (Modificado de Boder y Jiang 2011).

La estructura primaria de diferentes anticuerpos es similar, mas no idéntica, por lo que su estructura tridimensional también es similar. La mayoría de estas diferencias se encuentran en la parte más externa de las cadenas ligera y pesada, principalmente en seis asas hipervariables que forman las “regiones determinantes de complementariedad” (*Complementary Determining Regions*, CDRs), donde el antígeno se une al anticuerpo (Figura 4). Tres de las asas hipervariables se encuentran en la cadena pesada y tres en la cadena ligera. En ambos casos las asas se encuentran en la parte externa de la hoja-β que forma el dominio de inmunoglobulina. Las propiedades fisicoquímicas de la superficie formada por estas seis asas es lo que determina en mayor proporción la afinidad de un anticuerpo por su antígeno. La variabilidad de esta interfaz resulta de la alta diversidad de secuencias codificadas en el genoma para estas asas hipervariables. A esto se suman las sustituciones de aminoácidos que resultan del proceso de hipermutación somática durante el curso de una respuesta inmune (Abul *et al.* 2010).



### **I.2.2 Generación de variabilidad en los anticuerpos**

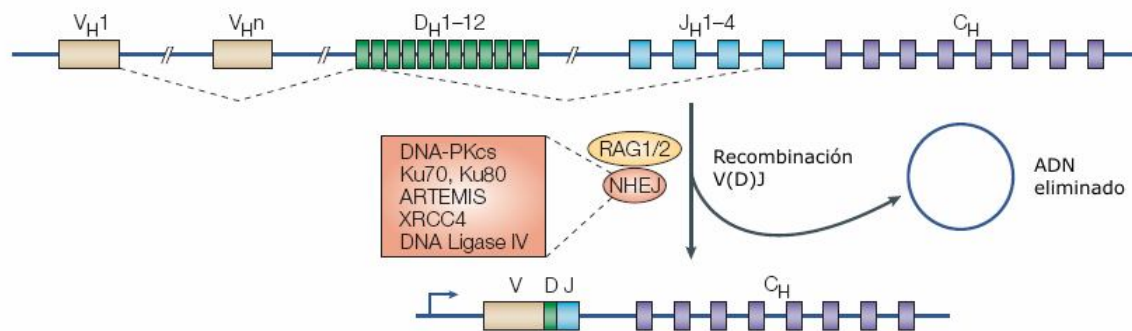
En la médula ósea, las células madre hematopoyéticas maduran en progenitores linfoides que darán origen a células B, células T, células NK y a algunas células dendríticas. Una vez que los progenitores linfoides están comprometidos a convertirse en células B, tiene lugar un proceso de re-arreglo genético denominado recombinación V(D)J, en el cuál una de muchas regiones variables se selecciona al azar y se une a un segmento río abajo. Este re-arreglo no depende de la presencia de antígenos (Abul *et al.* 2010).

### **I.2.3 Recombinación V(D)J**

En el extremo 5' del locus del gen que codifica para la cadena ligera, hay un conjunto de segmentos V (de "Variability", variabilidad). Río arriba de cada segmento, hay un promotor desde el que se puede iniciar la transcripción. Río abajo de este segmento se encuentran varios segmentos J (*Joining*, unión), que están ligados a exones que codifican para las regiones constantes de la cadena ligera. El locus del gen que codifica para la cadena pesada tiene una estructura similar, pero entre los segmentos V y los segmentos J hay segmentos adicionales denominados D (*Diversity*, diversidad) (Figura 5).

En el extremo 3' de cada segmento V se encuentran secuencias de ADN llamadas "secuencias señal de recombinación" (*Recombination Signal Sequences*, RSS), las cuales también se encuentran en el extremo 5' de los segmentos J y a ambos lados de los segmentos D (Abul *et al.* 2010). Estas secuencias señales son reconocidas por la Recombinasa V(D)J, formada por las proteínas Rag-1 y Rag-2 (Abul *et al.*, 2010). La primera se encarga de reconocer las secuencias señales de recombinación, pero es sólo hasta que forma un complejo con Rag-2 que se activa y realiza un corte en la cadena de ADN (Figura 5). Posteriormente, otras enzimas, como Artemisa, DNA-PK, exonucleasas, TdT, Ku70, Ku80, XRCC4 y DNA-Ligasa IV, se encargan de reparar los cortes en el ADN mediante un proceso denominado unión de extremos no homólogos (*Non-Homologous End Joining*, NHEJ), dejando unidos los segmentos V y J; o V, D y J en el caso de las cadenas pesadas

(Abul *et al.* 2010).



**Figura 5.** Esquema del proceso de Recombinación V(D)J en el locus de la cadena pesada. Uno de los segmentos de la región variable (V<sub>H</sub>) se une con uno de los segmentos de la región de diversidad (D<sub>H</sub>), que a su vez se une con uno de los segmentos de unión (J<sub>H</sub>). Este proceso, mediado por las proteínas Rag-1 y Rag-2, y el complejo de proteínas que realiza la unión de extremos no homólogos, resulta en la generación de una secuencia codificante para la cadena pesada del anticuerpo (Modificado de Chaudhuri y Alt 2004).

En el locus de la cadena ligera de humanos hay 150 secciones V y cinco secciones J, lo que resulta en 750 combinaciones posibles. Para el caso de la cadena pesada existen 80 secciones V, 50 secciones D y 6 secciones J; lo que da lugar a 24 mil combinaciones posibles (Abul *et al.* 2010). Los anticuerpos están constituidos por cadenas ligeras y pesadas, por lo que si se toman en cuenta las 750 posibles cadenas ligeras y las 24 mil cadenas pesadas posibles, se pueden generar 18 millones de anticuerpos diferentes sólo mediante recombinación somática. La diversidad de anticuerpos aumenta en varios órdenes de magnitud debido a otros procesos, como la adición de nucleótidos durante el proceso de recombinación y la hipermutación somática durante la maduración de afinidad de los linfocitos (Abul *et al.*, 2010).

La actividad de las proteínas Rag-1 y Rag-2 es necesaria y suficiente para iniciar el proceso de recombinación, como se ha demostrado *in vitro* (Hiom *et al.*, 1998, Agrawal *et al.*, 1998) y mediante complementación funcional (Schatz *et al.*, 1989, Oettinger *et al.*, 1990).

Aunque aún se desconocen muchos detalles sobre el origen evolutivo de estas proteínas, se acepta que se originaron por la inserción de un transposón en el genoma de un ancestro común a todos los vertebrados con mandíbula (Litman *et al.*, 2010). Esto se concluyó por la manera en que se realiza el corte al ADN (reacciones de hidrólisis y transesterificación), que es la misma que llevan a cabo otras transposasas; y también por la semejanza que tienen algunos de sus dominios con transposasas encontradas en invertebrados (Litman *et al.*, 2010).

Los locus de las cadenas pesada y ligera se encuentran en diferentes cromosomas. La manera en que se organizan los segmentos que sufren re-arreglo, es esencialmente la misma en todos los mamíferos, aunque el número y la secuencia de los diferentes segmentos puede variar (Abul *et al.*, 2010).

## II Antecedentes

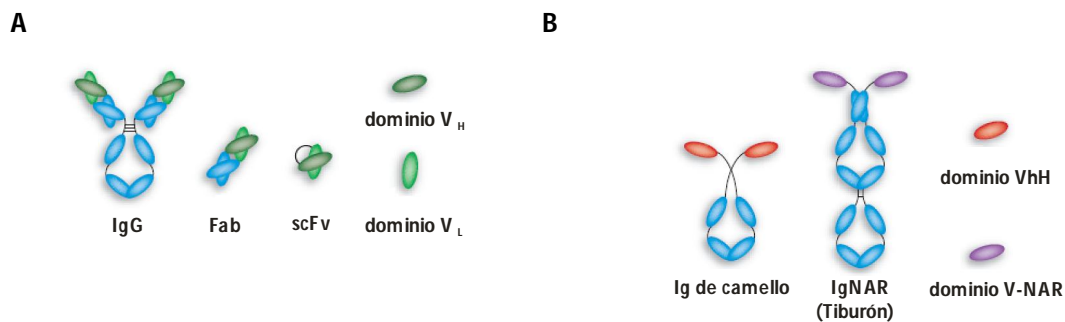
Uno de los mayores avances en la producción de anticuerpos fue el desarrollo de la tecnología de hibridomas (Kohler y Milstein 1975), la cual permitió que los anticuerpos se produjeran a partir de cultivos de linfocitos inmortalizados. A este desarrollo le siguió la producción en *E. coli* de fragmentos de anticuerpos funcionales (Skerra y Pluckthun 1988). Actualmente es posible seleccionar anticuerpos específicos a partir de librerías de anticuerpos fusionados a proteínas de una cápside viral mediante despliegue en fagos (Hoongboom y Chames, 2000), y es posible producir tanto anticuerpos enteros como fragmentos de estos en diferentes sistemas recombinantes como levaduras, bacterias y algas (Jeong *et al.* 2011, Tran *et al.* 2009, Mayfield *et al.* 2003). Estos anticuerpos conservan la capacidad de reconocimiento específico (Figura 6A), como los Fab de dos cadenas polipeptídicas, o los scFv de una sola cadena (Ruigrok *et al.* 2011).

Otra alternativa que se ha explorado es la producción aislada de dominios variables de inmunoglobulina de las cadenas pesada ( $V_H$ ) o ligera ( $V_L$ ) de anticuerpos de mamífero (Ward *et al.* 1989). Debido a que cada dominio variable presenta sólo tres de los seis CDRs que presenta un anticuerpo completo, el área de interacción con el antígeno se reduce a la mitad, y por esta razón es común que los dominios aislados pierdan la capacidad de reconocimiento específico del anticuerpo del que se derivaron (Holliger y Hudson 2005).

Además de los dominios de inmunoglobulina aislados de mamíferos, también se ha trabajado con inmunoglobulinas de otras especies (Holliger y Hudson 2005, Saerens *et al.* 2008, Harmsen y De Haard, 2007) (Figura 6B). La ventaja de las inmunoglobulinas de camélidos y tiburones es que, de manera natural, la superficie responsable del reconocimiento específico se encuentra en un solo dominio de inmunoglobulina (De Genst *et al.* 2004, Streltsov *et al.* 2004). Este dominio puede producirse de manera aislada del resto del anticuerpo y mantener su especificidad y solubilidad, a diferencia de los  $V_H$  y  $V_L$  (Stanfield *et al.* 2004; Conrath *et al.* 2004). El menor tamaño de estas proteínas les

permite tener una mayor penetración en los tejidos, lo que las hace candidatas para aplicarse en inmunoterapia (Huhlov y Chester, 2004); además, su menor tamaño permite que sean excretadas rápidamente, con lo que se disminuye su concentración en sangre, aumentando el contraste en aplicaciones de imagenología *in vivo* (Sundaresan *et al.* 2003).

Sin embargo, las inmunoglobulinas de otras especies requieren ser “humanizadas” para administrarse en humanos con el fin de evitar respuestas inmunes adversas (Holliger y Hudson 2005), aunque se ha reportado que los dominios VhH de llama son poco inmunogénicos (Cortez-Retamozo *et al.* 2004), por lo que podrían no requerir ser humanizados.



**Figura 6.** Esquema de polipéptidos derivados de anticuerpos. **A.** Fragmentos de anticuerpos derivados de corte enzimático de inmunoglobulinas, o producidos mediante ingeniería genética (Fab: *Fragment antigen-binding*), scFv: *Single chain variable fragment*, V<sub>H</sub>: *Heavy chain variable segment*, V<sub>L</sub>: *Light chain variable segment*). **B.** Inmunoglobulinas de otras especies y polipéptidos derivados de las regiones variables de estos (Modificado de Holliger y Hudson, 2005).

## II.1 Plataformas proteicas diseñadas

De manera paralela al desarrollo de plataformas basadas en dominios de inmunoglobulina se han desarrollado otras moléculas de reconocimiento basadas en estructuras sin relación alguna con estos dominios. Sobre estas estructuras pueden seleccionarse sitios de unión mediante ingeniería combinatoria, mutagénesis aleatoria dirigida, despliegue en fagos, despliegue en ribosomas, entre otras técnicas (Rothe, *et al.* 2006, Skerra 2003). Esto ha ampliado la gama de proteínas con potencial de reconocimiento para su uso en

biotecnología y medicina (Sheridan, 2007).

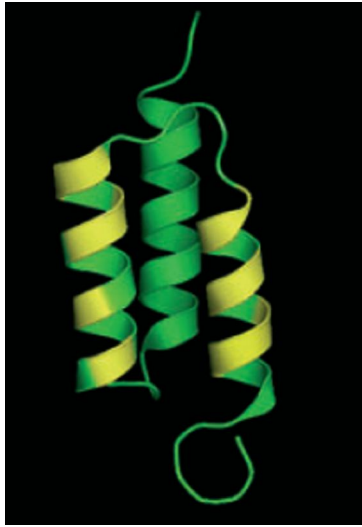
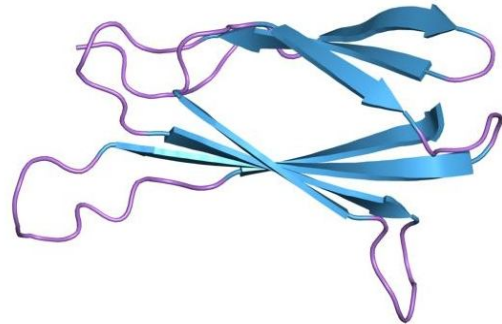
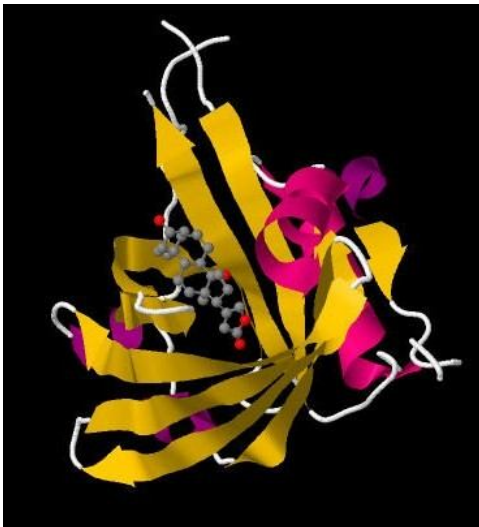
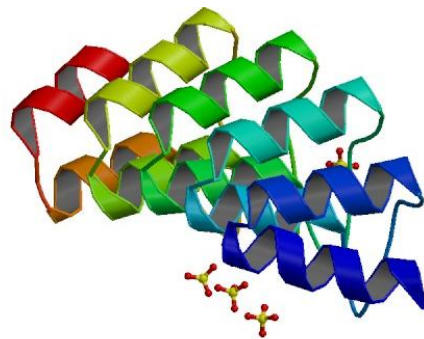
Al conjunto de proteínas utilizadas con funciones de reconocimiento análogas a los anticuerpos se les llama "Plataformas proteicas diseñadas" (*Engineered protein scaffolds*). Estas plataformas se derivan de la estructura terciaria que adopta algún dominio de una familia de proteínas. Es deseable que esta estructura sea rígida, estable y de fácil producción y escalamiento (Skerra 2000). Con base en estas plataformas estructurales se pueden generar proteínas con ciertas propiedades deseables, como solubilidad, termoestabilidad, resistencia ante proteasas, que además pueden producirse de manera recombinante (Gebauer y Skerra 2009).

Se han propuesto más de 50 plataformas diferentes, algunas de las más desarrolladas son las siguientes:

- **Affibodies.** Se basan en el dominio-Z de la proteína A del género *Staphylococcus*, y su interfaz de reconocimiento se encuentra en dos de sus tres hélices- $\alpha$  (Figura 7A). Tiene 58 aminoácidos y puede producirse de manera recombinante, o bien, mediante síntesis de péptidos en fase sólida (Nygren 2008). Se han producido affibodies con afinidades micromolares por la *Taq* polimerasa y la apolipoproteína A1 (Nord *et al.* 1997, 2000). También se ha demostrado su funcionalidad para reconocer e inhibir la función de CD28 (Sandstrom 2003).
- **Monobodies o Adnectinas.** Se basan en un dominio extracelular de la fibronectina humana III, que adopta una estructura de sándwich- $\beta$  de 94 residuos con alta termoestabilidad (Figura 7B). La compañía Adnexus ha desarrollado un antagonista del receptor VEGF para el tratamiento de glioblastoma que se encuentra en pruebas clínicas fase II (Bloom y Calabro, 2009).
- **Anticalinas.** Se derivaron de las lipocalinas humanas, una familia de proteínas con

estructura de barril- $\beta$ , que unen ligandos hidrofóbicos pequeños como esteroides, bilinas y retinoides (Figura 7C). Tienen un tamaño de alrededor de 180 residuos. Por su estructura es posible fusionarla con otros dominios o proteínas tanto en el extremo amino-terminal como en el carboxilo-terminal, tales como GFP u otras proteínas reporteras o efectoras (Skerra 2008).

- **DARPinas** (acrónimo de *Designed Ankyrin Repeat Proteins*). Se construyeron con un enfoque de diseño consenso a partir de las ankirinas, proteínas de unión entre proteínas transmembranales y el esqueleto de actina; las cuales son muy abundantes en mamíferos. Tienen una estructura basada en módulos con estructura hélice-giro-hélice (Figura 7D). Los módulos amino y carboxilo-terminal son constantes y entre ellos se encuentran módulos variables. Son termoestables (su temperatura de desnaturalización promedio es de 66°C) y tienen un peso aproximado de 130 aminoácidos (Pancer *et al.*, 2004; Stumpp *et al.*, 2008).

**A****B****C****D**

**Figura 7.** Estructura tridimensional de distintas plataformas proteicas diseñadas. **A.** Estructura de un Affibody (Tomado de Nygren, 2008). **B.** Estructura de un Monobody (Tomado de PDB, PDBID: 1TTG). **C.** Estructura de una Anticalina en complejo con digitoxenina (Tomado de PDB, PDBID: 1LNM). **D.** Estructura de una DARPin en complejo con iones sulfato (Tomado de PDB, PDBID: 2QYJ).

Todas estas plataformas combinan una alta estabilidad estructural con los beneficios de poder producirse en sistemas de expresión microbianos. Además de que no interaccionan



de manera inespecífica con receptores  $F_c$  y presentan una excreción más rápida a través del riñón, lo que mejora el contraste entre la concentración de la molécula en el tejido blanco y la concentración en sangre, lo que las hace ideales para aplicaciones biomédicas como imagenología *in vivo* (Gebauer y Skerra 2009).

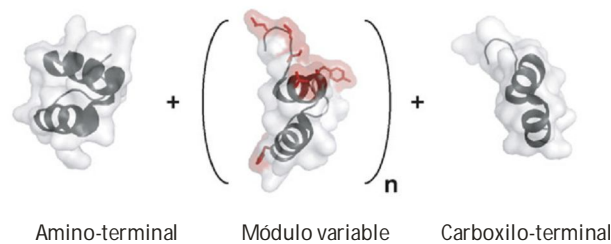
A pesar de los avances en el diseño de plataformas proteicas diseñadas como análogos de los anticuerpos, la inversión necesaria para generar una proteína funcional de este tipo es más alta que la preparación de un anticuerpo convencional, por lo que la mayoría de los esfuerzos de los grupos y empresas propietarias de estas tecnologías se enfocan a producir proteínas con fines terapéuticos, con el fin de obtener, posiblemente, un retorno más rápido de la inversión (Gebauer y Skerra, 2009).

## **II.2 DARPinas**

Las DARPinas son proteínas pequeñas, con un solo dominio que fueron diseñadas a partir de los dominios con repeticiones de ankirina (Stumpp *et al.*, 2008, Patente Internacional WO 02/20565). Estos dominios forman parte de una gran diversidad de proteínas en mamíferos y son responsables de interacciones entre proteínas muy diversas. Su estructura consiste en repeticiones de un motivo estructural de 33 residuos que consiste en un giro- $\beta$  seguido de dos hélices antiparalelas y un loop que se enlaza con el giro- $\beta$  de la siguiente repetición. Estas repeticiones forman dominios estables de entre 4 y 6 repeticiones, aunque se han reportado hasta 12 repeticiones en un dominio (Binz *et al.* 2003).

La estructura de las DARPinas se construyó con un enfoque de diseño consenso, tomando 229 repeticiones de 33 residuos de longitud del motivo estructural base de las ankirinas antes mencionado. Estas repeticiones se alinearon y se escogieron los residuos que presentaran una frecuencia  $\geq 30\%$  para formar parte de la secuencia consenso (Binz, *et al* 2003). Esta secuencia consenso se refinó utilizando información estructural de proteínas cristalizadas que presentan dominios con repeticiones de ankirina (Binz, *et al* 2003). Con

esta secuencia consenso es posible obtener proteínas que constan de módulos repetidos, dentro de cada uno de los cuáles seis residuos pueden sustituirse por otros aminoácidos. Estos módulos repetidos variables se encuentran flanqueados por motivos amino y carboxilo terminales que protegen el núcleo hidrofóbico formado por los demás módulos (Pancer *et al.*, 2004; Stumpp *et al.*, 2008) (Figura 8). El número de residuos variables por módulo resulta en un espacio de secuencias posibles de  $4 \times 10^{15}$  para una DARPin constituida sólo por dos módulos variables. Esta diversidad puede aumentarse añadiendo más módulos.



**Figura 8.** Módulos que conforman una DARPin; amino-terminal, el módulo de la librería (en rojo se muestran los aminoácidos variables) y el carboxilo-terminal (Modificado de Stumpp *et al.*, 2008).

El grupo de Stumpp mostró que, partiendo de un repertorio de DARPins construidas con motivos de la librería que ellos construyeron, es posible generar DARPins con una alta afinidad y especificidad a distintas moléculas (Amstutz, *et al.* 2001), mediante despliegue en ribosomas y despliegue en fagos (Steiner *et al.* 2008). La estructura basada en repeticiones de ankirinas es termoestable, alcanzando temperaturas de desnaturalización de entre 66 °C y 95 °C , dependiendo de la secuencia de aminoácidos particular de cada DARPin, siendo menos estables las que tienen ciertos residuos sustituidos en el contacto entre las hélices antiparalelas que forman el núcleo hidrofóbico de las DARPins (Binz *et al.* 2003).

Se han desarrollado DARPins mediante despliegue en fagos, que reconocen

específicamente a HER-2, proteína asociada a ciertos tipos de cáncer de mama, que pueden usarse con fines terapéuticos o de diagnóstico. Algunas DARPinas ya se encuentran en pruebas clínicas fase I y II (DARPinas inhibidoras de VEGF, para el tratamiento de la degeneración macular y del edema macular) (Steiner *et al.*, 2008; Boersma y Plückthun, 2011).

De las plataformas proteicas diseñadas, las DARPinas son las que presentan una mayor modularidad estructural, debido a que su superficie de interacción puede modificarse dependiendo del número de módulos variables, además de que son proteínas monoméricas estables en condiciones fisiológicas, libres de residuos de cisteína y pueden producirse en grandes cantidades (Binz *et al.*, 2003).

### **II.3 Sistemas de recombinación sitio-específicos**

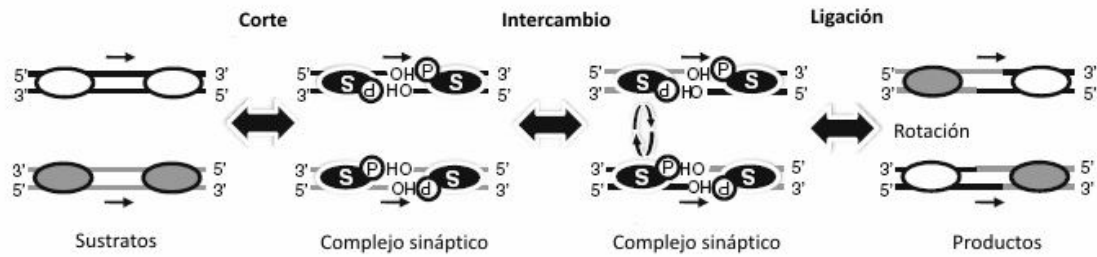
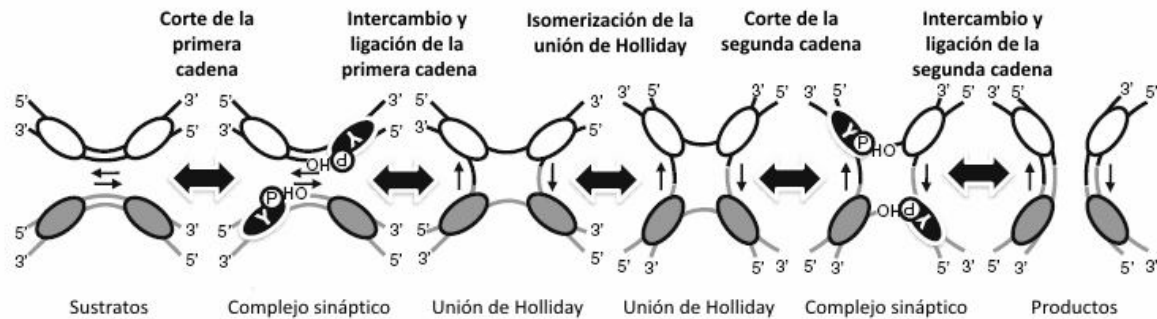
Los sistemas de recombinación sitio-específicos catalizan la recombinación entre dos regiones de ADN que contienen secuencias que son reconocidas por las recombinasas. Estas proteínas se encuentran presentes en todos los dominios de los seres vivos, jugando papeles importantes en la replicación de fagos, plásmidos y transposones (Austin *et al.* 1981, Broach *et al.* 1982, Rice *et al.* 2010); regulando fases de crecimiento bacteriano y la virulencia de fagos, mediando el re-arreglo genómico durante la esporulación de hongos y, pueden ser el origen evolutivo de las enzimas Rag-1 y Rag-2 (Agrawal *et al.* 1998). La especificidad de estas enzimas las ha convertido en herramientas ideales en la manipulación genética de organismos, ya que permiten insertar secuencias en los genomas de los organismos, o bien, eliminarlas (Hirano *et al.* 2011).

Aunque todas las recombinasas catalizan la recombinación entre dos secuencias de ADN, se clasifican en dos familias de acuerdo al aminoácido que se encuentra en su sitio activo: las serina-recombinasas y las tirosina-recombinasas.

### II.3.1 Serina-recombinasas

A esta familia pertenecen las resolvasas  $\gamma\delta$  y Tn3, las integrasas  $\phi$ C31 y R4, y las invertasas Hin y Gin. Estas proteínas catalizan la recombinación sitio-específica entre secuencias que pueden ser repeticiones directas o invertidas, lo que determina si el fragmento de ADN que se encuentra entre ellas se elimina, se integra o se invierte (Hirano *et al.*, 2011).

Las recombinasas de esta familia forman un “complejo sináptico” de cuatro enzimas (Figura 9A), dos dímeros que reconocen, cada uno, a dos sustratos de ADN. Simultáneamente, todas las recombinasas se activan y los grupos hidroxilo de los residuos de serina del centro catalítico realizan un ataque nucleofílico al enlace fosfodiéster de una de las cadenas. Esto resulta en la formación de un complejo sináptico donde los fosfatos 5' de las cadenas de ADN se encuentran unidas covalentemente a las recombinasas. El intercambio de cadenas se da por una rotación de 180° del complejo sináptico, y posteriormente se unen los extremos hidroxilo 3' a los fosfatos 5' (Rice *et al.* 2010).

**A****B**

**Figura 9.** Esquema del mecanismo de reacción de RSE. **A.** Esquema del mecanismo de recombinación de las serina-recombinasas. **B.** Esquema del mecanismo de recombinación de las tirosina-recombinasas (Modificado de Hirano *et al.* 2011).

Cada enzima tiene diferentes requerimientos para llevar a cabo la recombinación, por ejemplo, las invertasas Hin y Gin necesitan de factores adicionales del huésped, como las proteínas Fis (*Factor for inversion stimulation*) y HU (*Histone-like protein*) y de que la cadena de ADN adopte una topología especial (Hirano *et al.* 2011). Esto ha limitado su uso en ingeniería genética, aunque se han generado mutantes llamadas “resolvasas/invertasas activadas” que no requieren la formación de un intermediario altamente estructurado (Arnold *et al.* 1999).

### II.3.2 Tirosina-recombinasas

El complejo sináptico de estas enzimas consiste en un par de dímeros. Para catalizar la reacción, los residuos de tirosina de uno de los dímeros actúan como nucleófilos y cortan

cada cadena de ADN en el enlace fosfodiéster 5', mientras que el extremo 3' se une covalentemente al residuo de tirosina. Los grupos hidroxilo 5' atacan los enlaces fosfotirosina del complejo nucleoproteico adyacente, formando una unión de Holliday. Esta unión de Holliday se isomeriza, mientras que el otro dímero se activa y cataliza la segunda ronda de corte e intercambio de cadenas, lo que resulta en la resolución de la unión de Holliday (Figura 9B) (Hirano *et al.* 2011).

Las tirosina-recombinasas se clasifican a su vez en integrasas de fagos y en recombinasas simples. Las integrasas de fagos catalizan de manera unidireccional la recombinación de dos secuencias específicas (Hirano *et al.*, 2011) y requieren de factores adicionales que complementan su función, por ejemplo, la direccionalidad de la recombinación de la integrasa del fago  $\lambda$  depende de la proteína IHF (*Integration Host Factor*) de *E. coli* (Bushman *et al.*, 1985).

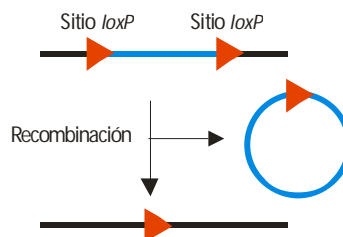
En cambio, las recombinasas simples no requieren de proteínas accesorias o ADN estructurado (Gosh y Van Duyne, 2002, Turan *et al.* 2011). Ejemplos de estas recombinasas son Cre y Flp, que catalizan la recombinación entre dos secuencias idénticas. La recombinasa Flp participa en la replicación del círculo de 2  $\mu$ m de levaduras, un plásmido de aproximadamente 6 Kb, catalizando reversiblemente la recombinación entre dos secuencias *FRT* de 48 pb (Broach y Guarascio 1982). La recombinasa Cre es parte del mecanismo de replicación del bacteriófago P1, cuyo genoma lineal se cicliza al recombinar dos secuencias *loxP* de 34 pares de bases que se encuentran en sentido directo. Esto le permite replicarse y segregarse a las células hijas. De los sistemas de recombinación, el sistema Cre/*loxP* es el más usado en la manipulación genética para eliminar de manera específica secuencias de ADN en varios organismos, incluyendo eucariontes, esto se debe a que no necesita de proteínas accesorias y al menor tamaño de su secuencia *loxP* (Hirano *et al.*, 2011).

## II.4 El sistema Cre/*loxP*

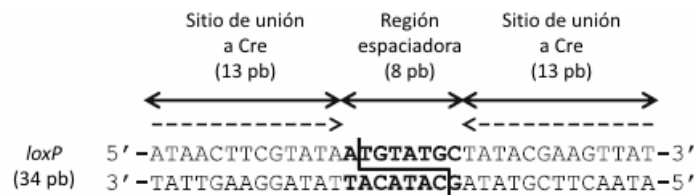
La recombinasa Cre es una enzima de 38 kDa que cataliza la recombinación entre dos secuencias idénticas de 34 pb, llamadas *loxP* (Figura 10). Esta proteína forma parte del mecanismo de replicación del bacteriófago P1. El genoma de este fago no se integra en el genoma del huésped, sino que se replica de manera independiente como un plásmido. Después de la infección, la recombinación mediada por Cre cicliza el genoma lineal del fago, permitiéndole replicarse y segregarse a las células hijas (Austin *et al.* 1981).

El sitio *loxP* se compone de dos secuencias repetidas invertidas de 13 pares de bases y un espaciador asimétrico de 8 pares de bases (Strachan y Read 1999). El espaciador asimétrico hace que el resultado de la recombinación dependa de la orientación de los sitios *loxP*. Cuando estos están alineados, el ADN que se encuentra entre ellos es escindido, mientras que cuando la secuencia se encuentra invertida, la recombinación resulta en la inversión del ADN.

**A**



**B**



**Figura 10.** Recombianción mediada por la recombinasa Cre. **A.** Esquema de la recombinación mediada por Cre. **B.** Estructura de la secuencia *loxP* WT. Las regiones espaciadoras se muestran en negritas. Las repetidas invertidas se indican con las flechas punteadas (Modificado de Hirano *et al.* 2011).

#### II.4.1 Modificaciones al sistema Cre/*loxP*

Debido a que de manera natural la recombinación entre dos sitios idénticos resulta en la eliminación de la secuencia de ADN entre ellos, se han desarrollado arreglos de secuencias heterólogas *loxP* que permiten la inserción dirigida, orientada y permanente de material genético.

Además de la secuencia *loxP*, se han identificado y desarrollado secuencias funcionales con mutaciones tanto en las repetidas invertidas como en el espaciador (Figura 11). Estas secuencias recombinan eficientemente con otras idénticas a ellas, y tienen una baja tasa de recombinación con otras secuencias. La eficiencia de recombinación entre dos secuencias heterólogas depende de la similitud entre las secuencias usadas y de la orientación de las secuencias, de acuerdo a un ensayo desarrollado por Siegel *et al.*, (2001).

		12345678
<b>WT <i>LoxP</i></b>	<u>ATAACTTCGTATAGC</u> <u>CAT</u> <u>TATACGAAGTTAT</u>	
<b>511 <i>LoxP</i></b>	<u>ATAACTTCGTATAGT</u> <u>ATA</u> <u>CATTATACGAAGTTAT</u>	
<b>511-I <i>LoxP</i></b>	<u>ATAACTTCGTATAA</u> <u>TGT</u> <u>ATACTATACGAAGTTAT</u>	
<b>FAS <i>LoxP</i></b>	<u>ACA</u> <u>ACTTCGTATATA</u> <u>CCTTTCT</u> <u>TATACGAAGTTGT</u>	
<b>2272 <i>LoxP</i></b>	<u>ATAACTTCGTATAGG</u> <u>ATA</u> <u>CCTTATACGAAGTTAT</u>	
<b>5171 <i>LoxP</i></b>	<u>ATAACTTCGTATAGT</u> <u>AC</u> <u>CATTATACGAAGTTAT</u>	

**Figura 11.** Secuencia *loxP* tipo (WT) y secuencias *loxP* mutadas (511, 511-I, FAS, 2272 y 5171). Las secuencias repetidas invertidas se encuentran subrayadas. La secuencia espaciadora se indica con números del 1 al 8 (Tomado de Siegel *et al.*, 2001)

El ensayo consistió en insertar un plásmido portador de una construcción con la estructura Promotor-*loxP* 3'-GFP-*loxP* 5' en una cepa de *E. coli* que expresaba la recombinasa Cre, de manera que al darse la recombinación el segmento codificante para la GFP dejaba de



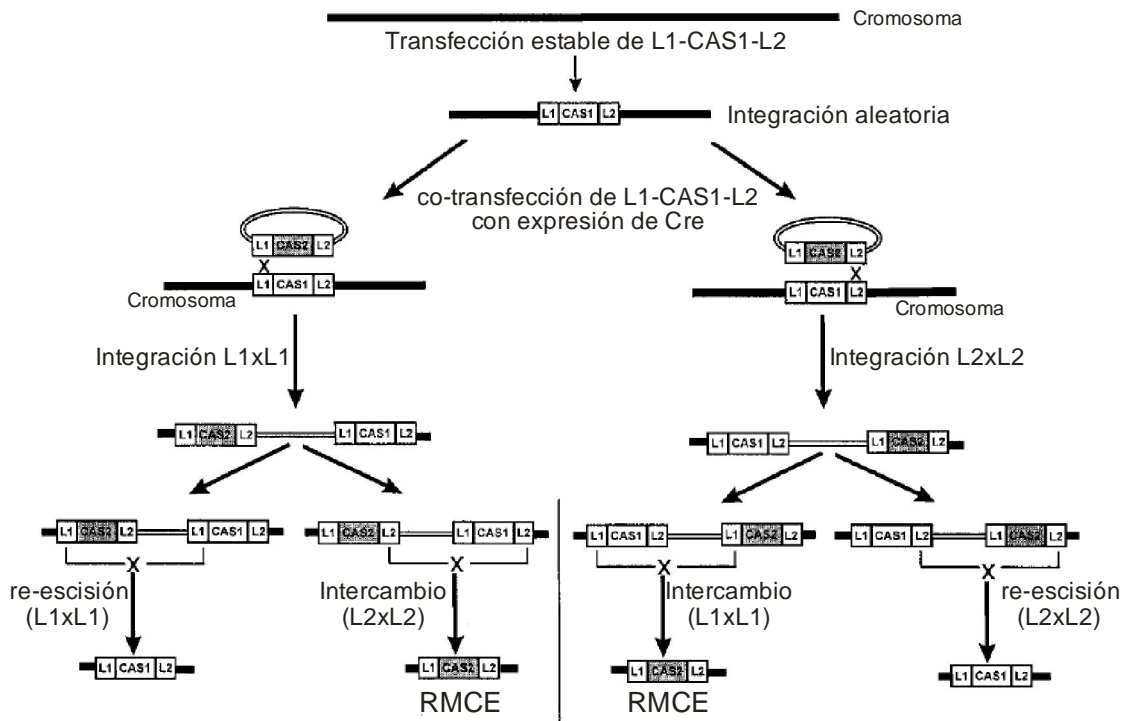
expresarse. En la construcción se probaron distintas combinaciones de sitios *loxP* heterólogos, y la eficiencia de recombinación entre los distintos sitios *loxP* se analizó mediante conteo de colonias individuales que no expresaran GFP (Siegel *et al.*, 2001) (Tabla 1).

**Tabla 1.** Porcentaje de recombinación entre distintas secuencias heterólogas *loxP* (WT, 511-I, 511, FAS, 2272, 5171) orientadas en la misma dirección (Modificado de Siegel *et al.*, 2001).

Sitio 3' <i>loxP</i>	Sitio 5' <i>loxP</i>					
	WT	511-I	511	FAS	2272	5171
WT	99.6					
511-I	1.4	99.2				
511	10.3	75.3	99.8			
FAS	0.2	5.7	0	99.4		
2272	0.5	0.3	1.6	1.7	99.7	

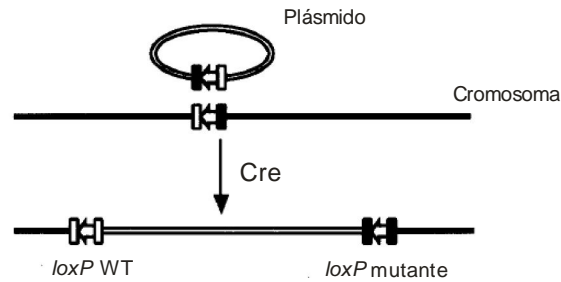
Usando arreglos de estos y otros sitios *loxP* heterólogos se han desarrollado metodologías que permiten la inserción de material genético en genomas de manera específica, como el RMCE (*Recombination-mediated cassette exchange*) (Bouhassira *et al.* 1997) y el sistema mutante LE/RE (*Left element/Right element*) (Albert *et al.* 1995).

El sistema RMCE (Figura 12) se diseñó para insertar una secuencia de manera controlada en el genoma de células MEL. Para lograrlo primero se integra de manera aleatoria un casete de expresión flanqueado con dos secuencias *loxP* heterólogas en el genoma de la célula. Posteriormente se seleccionan aquellas células que presentan la inserción en la región del genoma deseada y, finalmente, este primer casete se reemplaza por un segundo casete con la secuencia deseada. Para lograrlo, este segundo casete se flanquea con las mismas dos secuencias *loxP* heterólogas del primero. Esta secuencia se inserta en un plásmido que se cotransfecta con otro plásmido que expresa constitutivamente Cre. El producto de la recombinación depende del orden en que se dé la recombinación de las dos secuencias *loxP* en la sustitución del primer casete por el segundo (Bouhassira *et al.*, 1997).



**Figura 12.** Esquema del RMCE. Una vez que se transfecta e integra en el genoma, el casete 1 (CAS1) flanqueado por dos sitios *loxP* heterólogos (L1 y L2), se cotransfecta un plásmido que expresa Cre. De acuerdo al orden de la recombinación entre los sitios L1 y L2 se pueden obtener cuatro resultados, dos de ellos resultan en el intercambio del casete 1 por el casete 2 (CAS2). (Modificado de Bouhassira *et al.* 1997)

Por otro lado, el sistema LE/RE se basa en la modularidad de la secuencia *loxP*. En este sistema se utilizaron mutantes de los elementos que se encuentran a la izquierda y a la derecha de la región espaciadora de la secuencia *loxP*. Este sistema, probado en plantas (Albert *et al.* 1995), integra el contenido de un plásmido, que incluye un sitio *loxP* mutado en uno de sus elementos, en otro sitio que tiene mutado el otro elemento (Figura 13). La recombinación entre estos dos sitios resulta en la inserción del contenido del plásmido flanqueado por dos sitios *loxP*, uno WT y otro mutado, incapaces de recombinar entre sí.



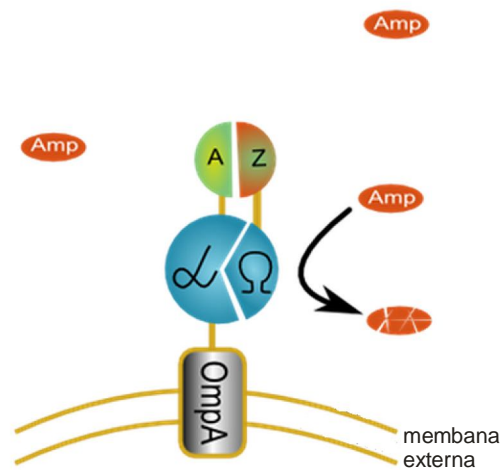
**Figura 13.** Esquema del sistema LE/RE. Los elementos de unión a Cre de los sitios *loxP* se esquematizan con cuadros, los elementos WT en blanco y los mutados en negro; la región espaciadora como flecha. La recombinación entre los dos sitios *loxP* híbridos resulta en la integración del contenido del plásmido en el cromosoma (Modificado de Albert *et al.*, 1995).

La existencia de distintas secuencias homólogas de *loxP* y la modularidad de su estructura hace posible la utilización del sistema Cre/*loxP* como base para generar herramientas de manipulación genética, como los sistemas ya mencionados.

## II.5 Alternativas celulares de la biología sintética para la generación y selección de proteínas con reconocimiento específico.

Una alternativa en el desarrollo de técnicas de reconocimiento proteico fue presentada por la Universidad de Varsovia en el iGEM 2008. El proyecto consistió en diseñar y probar un dispositivo bacteriano para generar y seleccionar proteínas capaces de interactuar con otra proteína blanco. La proteína blanco se fusionó al fragmento  $\Omega$  de la  $\beta$ -lactamasa. Esta proteína de fusión se produce y purifica por separado y se añade al medio. Por otro lado, la proteína de reconocimiento se fusiona al fragmento  $\alpha$  de la  $\beta$ -lactamasa y a un fragmento de la proteína OmpA de *E. coli*, que localiza a la proteína en la superficie de la membrana externa de la célula. De esta manera, al interactuar ambas proteínas la proteína blanco y la de reconocimiento, las subunidades de la  $\beta$ -lactamasa pueden interactuar y degradar el antibiótico presente en el medio (Figura 14) ([http://2008.igem.org/Wiki/Team:Warsaw/igem\\_project.htm](http://2008.igem.org/Wiki/Team:Warsaw/igem_project.htm)).

El sistema de generación se basa en la acción de la enzima AID (*Activation-Induced cytidine Deaminase*), responsable del proceso de hipermutación somática en los linfocitos (Muramatsu *et al.* 2000). Esta enzima se ha expresado en *E. coli*, generando un fenotipo mutagénico en genes con un alto nivel de expresión (Petersen-Mahrt *et al.* 2002). El fenotipo mutagénico que esta enzima causa, se debe a su acción como deaminasa sobre la citidina, convirtiendo a esta base en uridina. El mecanismo de reparación de ADN cambia el par GC, convertido a GU por un par TA (Bransteitter *et al.* 2003).



**Figura 14.** Esquema del sistema de selección de proteínas capaces de interactuar con otra proteína blanco del equipo de la Universidad de Varsovia (iGEM 2008). La proteína A de *Staphylococcus aureus* (A) se fusionó al fragmento  $\alpha$  de la  $\beta$ -lactamasa y con un fragmento de OmpA. Mientras que la proteína blanco, el “affibody”  $Z_{SPA1}$  (Z), se fusionó al fragmento  $\Omega$  de la  $\beta$ -lactamasa. La interacción entre estas dos proteínas permite la interacción funcional entre los fragmentos  $\alpha$  y  $\Omega$  y por lo tanto la degradación del antibiótico.

El equipo logró mostrar el funcionamiento por separado de ambos sistemas, usando para demostrar el sistema de selección la unión entre la proteína A de *Staphylococcus aureus* y el affibody  $Z_{SPA1}$ , que reconoce específicamente a la proteína A (Lendel *et al.* 2004). Para el sistema de generación transformaron a *E. coli* con un plásmido que expresaba AID, comprobando la acción mutagénica de la enzima mediante la aparición espontánea de bacterias resistentes a rifampicina.

## **Hipótesis**

El arreglo de secuencias codificantes para fragmentos de una plataforma proteica diseñada y secuencias señales de recombinación específica, junto con la expresión de una recombinasa específica, resultará en la generación de un conjunto de proteínas con potencial de reconocimiento específico.

## **Objetivo general**

Diseñar y construir un sistema de recombinación somática sintética en *E. coli*, basado en la estructura modular de las DARPinas y el sistema de recombinación Cre/lox.

## **Objetivos específicos**

- 1.- Diseño y construcción de las secuencias de ADN de los módulos que conformarán al sistema de recombinación sintética
- 2.- Diseño de los mecanismos de recombinación y de regulación del sistema sintético de expresión de DARPinas en la bacteria *E. coli*.
- 3.- Construcción del mecanismo de recombinación propuesto.

## Metodología

### Obtención y alineamiento de secuencias

Se obtuvieron las secuencias de aminoácidos de cinco DARPinas diferentes disponibles en el *Protein Data Bank* (PDB) con los siguientes números de identificación: 2QYJ, 2V5Q:D, 2P2C:R y 1SVX:A (las secuencias completas se encuentran en el Anexo 1). Estas secuencias se alinearon usando el programa ClustalX (Thompson *et al.*, 1997). El alineamiento resultante se ajustó manualmente utilizando el programa CLC Sequence Viewer.

Las secuencias del promotor *Plac*, la región de inserción en la membrana externa de la proteína transmembranal de *E. coli* OmpA y los terminadores de la transcripción se obtuvieron del *Registry of Standard Biological Parts*, sus identificadores y secuencias se pueden ver en el Anexo 2.

Las secuencias heterólogas *loxP* utilizadas en el diseño del sistema recombinante se tomaron de Siegel *et al.* (2001)

### Transcripción *in silico*, retrotranscripción y optimización de secuencias

Las secuencias de aminoácidos se retrotranscribieron para obtener secuencias de nucleóticos optimizadas para su expresión en *E. coli* utilizando la aplicación "Backtranseq" del programa libre Jemboss (Carver y Bleasby 2003), con la tabla de uso de codones del archivo "Eecoli.cut". La transcripción *in silico* de secuencias de nucleótidos a secuencias de aminoácidos se realizó con la aplicación "Transeq" del mismo programa.

### Predicción de estructura terciaria y visualización de estructuras tridimensionales

Las predicciones de estructura terciaria se realizaron con el servicio en línea I-TASSER que combina alineamientos de secuencias, modelado *ab initio* y refinamiento de estructura a nivel atómico para predecir estructuras proteicas (Roy *et al.* 2010). Las imágenes se generaron con el programa "UCSF Chimera" (Pettersen *et al.* 2004).

### **Preparación de *E. coli* TOP 10 quimiocompetentes**

Se partió de un glicerol de *E. coli* TOP 10 (Invitrogen) que se creció en un matraz con 20 mL de medio SOC estéril (Tryptona 2 %, extracto de levadura 0.5%, NaCl 10 mM, MgCl<sub>2</sub> 2.5 mM, MgSO<sub>4</sub> 10 mM, glucosa 20 mM) a 37° C con una agitación de 150 rpm hasta obtener una OD<sub>600</sub> de 0.4. El cultivo se centrifugó a 3,000 rpm durante 10 min. Se decantó el sobrenadante y la pastilla se resuspendió en 10 mL de una solución amortiguadora TB estéril (Pipes 10 mM, MnCl<sub>2</sub> 55 mM, CaCl<sub>2</sub> 15 mM, KCl 250 mM). Se incubó en hielo por 25 min para posteriormente centrifugar a 2,000 g durante 10 min. Se decantó el sobrenadante y la pastilla se resuspendió en 2 mL de buffer TB estéril con glicerol al 20%. Se hicieron alícuotas de 100 µL y se almacenaron a -72° C.

### **Transformación de *E. coli* TOP 10 quimiocompetentes**

Se descongelaron en hielo alícuotas de 100 µL de *E. coli* TOP 10 quimiocompetentes, a las cuales se agregaron 2 µL de ADN a una concentración de aproximadamente 100 ng/µL y se incubaron en hielo durante 30 min. Después se transfirieron 30 segundos a un baño a 42°C y posteriormente se incubaron en hielo durante 2 min. Las células se resuspendieron en 500 µL de medio SOC estéril y se incubaron a 37 °C durante 1 hora. Se platearon 100 µL en cajas Petri con medio LB-Agar (Tryptona 10 g/L, Extracto de levadura 5 g/L, NaCl 5 g/L, pH 7.5 ajustado con NaOH, Agar 15 g/L), y ampicilina (100 µg/mL) o kanamicina (100 µg/mL), dependiendo del vector transformado.

### **Síntesis de ADN**

Las secuencias generadas para formar parte del Sistema de Recombinación se sintetizaron comercialmente por Mr. Gene (TM), y se flanquearon con las secuencias del prefijo y sufijo en los extremos 5' y 3', respectivamente. Los plásmidos obtenidos se nombraron pMK-T/C1, pMK-RQ/V1 y pMK-RQ/V2+C2 y se describen a detalle en la tabla 2.



**Tabla 2.** Detalle de los plásmidos en que se recibieron las secuencias sintetizadas.

<b>Secuencia</b>	<b>Tamaño de la secuencia (pb)</b>	<b>Plásmido</b>	<b>Resistencia</b>
<b>C1</b>	715	pMK-T/C1	Kanamicina
<b>V1</b>	554	pMK-RQ/V1	Kanamicina
<b>V2+C2</b>	908	pMK-RQ/V2+C2	Kanamicina

Los oligonucleótidos usados para amplificar específicamente las distintas secuencias se sintetizaron en la Unidad de síntesis y secuenciación del Instituto de Biotecnología de la UNAM mediante el método de Fosfito-Triéster en Fase Sólida, en el que la cadena creciente de ADN se mantiene "anclada" a un soporte sólido depositado en una columna delimitada por dos filtros e insoluble en los reactivos que fluyen a través de la misma.

**Tabla 3.** Oligonucleótidos utilizados.

<b>Nombre</b>	<b>Especificidad</b>	<b>Dirección</b>	<b>Secuencia</b>
<b>AF</b>	<b>Prefijo</b>	<b>Forward</b>	GTTTCT TCGAATTCGCGCCGCTTCTAG
<b>AR</b>	<b>Sufijo</b>	<b>Reverse</b>	CTGCAGCGCCGCTACTAGTA
<b>A1F</b>	<b>C1 5'</b>	<b>Forward</b>	TACTGGAATTCTTGACACCATC
<b>A2F</b>	<b>V1 5'</b>	<b>Forward</b>	GGCCGCTTCTAGACAACCTCG
<b>A3R</b>	<b>V2+C2 3'</b>	<b>Reverse</b>	CCTCACTAGTGCGAAAAAACCC

### **Secuenciación de ADN**

La secuenciación se realizó en el Instituto de Biotecnología de la Universidad Nacional Autónoma de México con el equipo Perkin Elmer/Applied Biosystems Modelo 3730. Para cada muestra se enviaron 15 µL de muestra de ADN a una concentración de al menos 30 ng/µL, junto con 1 µL del oligonucleótido específico a una concentración de 10 pmol/µL. El método empleado está basado en la Secuenciación Cíclica con Terminadores Marcados-Taq-FS. Este método consiste en una reacción de amplificación del templado con una polimerasa que incorpora dideoxinucleótidos fluorescentes que impiden la incorporación de nucleótidos en la cadena sintetizada. Cada uno de los dideoxinucleótidos (A, T, C y G)

tiene un marcador fluorescente distinto. La muestra es separada mediante electroforesis en gel de poliacrilamida y la secuencia se identifica mediante la excitación por laser de los marcadores fluorescentes, cuya emisión se detecta por un fotomultiplicador. La información es recibida y codificada por el software del equipo de secuenciación (<http://csb.wfu.edu/brf/ABlguide>).

### **Purificación de ADN plasmídico**

El ADN plasmídico se purificó a partir de cultivos de 5 ml de medio LB con ampicilina (100 µg/mL) o kanamicina (100 µg/mL), dependiendo del vector transformado. Los cultivos se crecieron a 37° C durante 12 h, con una agitación de 150 rpm. Para la purificación se utilizó el kit de extracción de plásmido *Quiagen QIAprep Spin Miniprep*, el cuál se basa en lisis alcalina de células seguida por adsorción de ADN en silica en presencia de altas concentraciones de sales. Para purificar el ADN se siguieron los pasos: preparación y lavado de lisado celular; adsorción de ADN en la membrana QIAprep; y lavado y elución de ADN.

### **Digestión de ADN**

Las digestiones se realizaron con las enzimas EcoRI (BioLabs, New England), PstI, SpeI y XbaI (Fermentas), utilizando una unidad de enzima en 50 µl de reacción durante 12 horas a 37°C. En el caso de las dobles digestiones, se utilizó la solución amortiguadora "Tango" (Fermentas).

### **Ligación de ADN**

Las ligaciones se realizaron con la ligasa T4 (Fermentas) en reacciones de 20 µL, durante 12 horas a temperatura ambiente. La proporción entre fragmentos a ligar fue de 1:1 para fragmentos de PCR; y de entre 3:1 y 5:1 para las ligaciones entre un inserto y un vector.

### **Ensayo de recombinación *in vitro***

El ensayo se realizó basándose en el protocolo de Abremski y Hoess (1983). El plásmido

pJW168 fue donado para este trabajo por el Dr. Guillermo Gosset Lagarda, del Instituto de Biotecnología de la UNAM. Este plásmido contiene el gen de la recombinasa Cre, cuya expresión se induce mediante IPTG. Este plásmido fue transformado en *E. coli* TOP 10 quimiocompetentes de acuerdo al protocolo ya descrito, excepto que la temperatura de recuperación fue de 30° C. Se cultivaron 5 mL de medio LB con Ampicilina (100 µg/mL) a 30° C agitado a 150 rpm durante 12 h. Se inocularon 1.5 mL de este cultivo en 15 mL de medio LB con Ampicilina (100 µg/mL) a 30° C agitado a 150 rpm. Cuando el cultivo alcanzó 0.6 de densidad óptica a 280 nm se indujo la expresión de Cre mediante la adición de IPTG para una concentración final de 0.5 mM. La expresión se indujo durante 4 h. Al finalizar la inducción se centrifugó el cultivo a 3,000 rpm durante 10 min a 4° C. La pastilla se resuspendió en 200 µL de solución amortiguadora TSE (Tris-HCl 20 mM, pH 7.5, NaEDTA 1 mM, NaCl 50 mM) con los inhibidores de proteasas aprotinina y leupeptina. Se hicieron alícuotas de 50 µL y se almacenaron a -72° C.

Para el ensayo de recombinación se tomó una alícuota y se le aplicaron 5 pulsos de sonicación de 15 segundos con 45 segundos de enfriamiento en hielo entre cada pulso. La muestra sonicada se centrifugó a 17,900 g durante 30 minutos a 4° C. Para el ensayo de recombinación se tomaron 35 µL del sobrenadante y 5 µL de ADN de cada uno de los plásmidos pMK-T/C1, pMK-RQ/V1 y pMK-RQ/V2+C2. Esta reacción se incubó a 37° C durante 15 minutos y se inactivó la reacción a 72° C durante 10 minutos.

Para amplificar los productos de la reacción de recombinación se centrifugó la reacción inactivada a 17,900 g durante 10 minutos y se tomaron 5 µL para una reacción de PCR de 25 µL. Se utilizaron oligonucleótidos específicos para el prefijo (en sentido) y el sufijo (reverso). De esta reacción se tomaron 5 µL para una segunda reacción de 25 µL, utilizando oligonucleótidos específicos para la región C1 (en sentido) y para la región C2 (reverso). Esta reacción se corrió en un gel de agarosa 2% y se cortó una banda de un tamaño aproximado de 1000 pares de bases. Se purificó el ADN de esta banda, eluyendo en 30 µL de agua libre de RNAsas. De esta solución se tomaron 2 µL para una amplificación

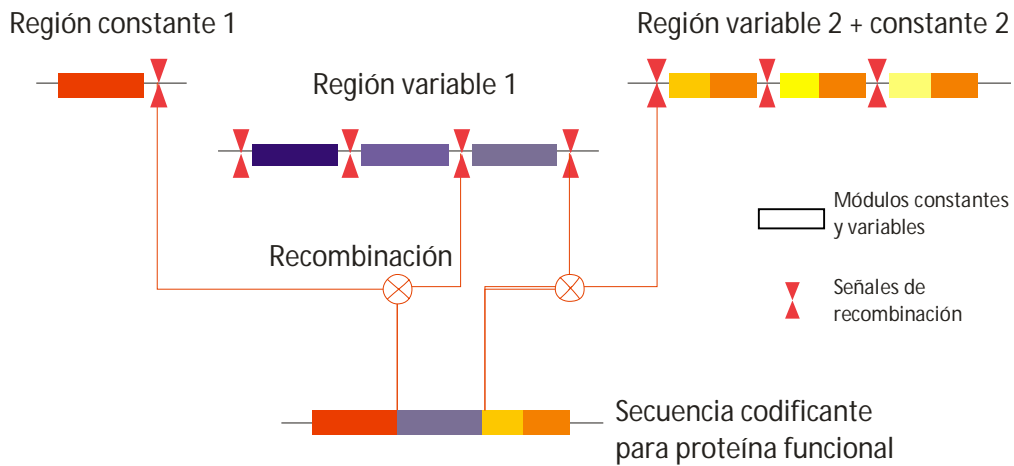
final en un volumen de 25  $\mu$ L, utilizando los oligonucleótidos A1F y A3R.

## Resultados

### Diseño de la estructura general del sistema de recombinación

Para diseñar un sistema sencillo capaz de generar una diversidad de secuencias, análogo al sistema altamente complejo de recombinación somática del sistema inmune, en este trabajo se eligió a las DARPinas como base estructural del sistema, debido a su modularidad y estabilidad estructural. Para lograr la funcionalidad de este sistema es crucial que durante el proceso de recombinación se generen secuencias de ADN correctas que puedan ser transcritas y posteriormente traducidas a proteínas funcionales.

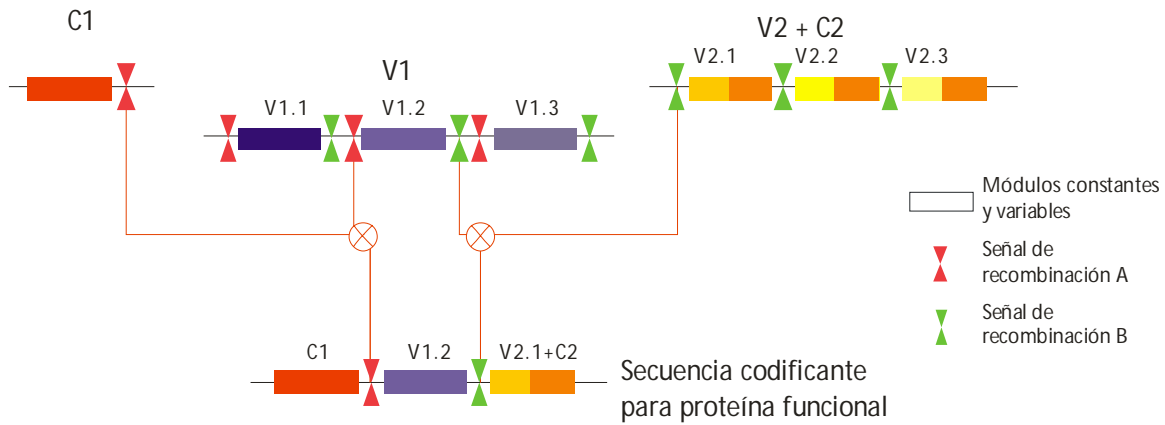
Además de que la base estructural debe ser sencilla, el mecanismo de recombinación debe ser fácil de implementar en *E. coli* y es deseable que no requiera de factores estructurales o funcionales adicionales. De los sistemas de recombinación se eligió a Cre/*loxP*, que no requiere de factores adicionales para funcionar, además de que existen secuencias *loxP* heterólogas que no recombinan entre sí (Siegel *et al.* 2001), y que pueden aprovecharse para el diseño del sistema. Para que el proceso de recombinación produzca secuencias funcionales de ADN, el arreglo de las secuencias nucleotídicas codificantes para los distintos módulos de las DARPinas y las secuencias señales de recombinación debe tener una cierta distribución a lo largo de la cadena, similar a la que se encuentra en los *loci* de las cadenas pesada y ligera (Figura 15).



**Figura 15.** Estructura general del Sistema de Recombinación. Las líneas rojas representan una posible recombinación.

Se decidió que el sistema constara de tres regiones, la región constante 1 (C1), la región variable 1 (V1) y la región variable 2 + constante 2 (V2+C2). La región C1 contendrá las secuencias necesarias para iniciar la transcripción, una secuencia de localización a membrana, el dominio N-terminal constante de las DARPinas y una secuencia señal de recombinación. La región V1 contendrá secuencias señales de recombinación intercaladas con secuencias codificantes para distintos dominios variables de DARPinas. La región V2+C2 contendrá repeticiones de una secuencia señal de recombinación, una secuencia codificante para un dominio variable y para el dominio C-terminal constante de DARPinas, así como una secuencia que permita la terminación de la transcripción. Tanto la región V1 como la región V2 tendrán tres módulos variables de DARPinas que diferirán en algunos de sus aminoácidos.

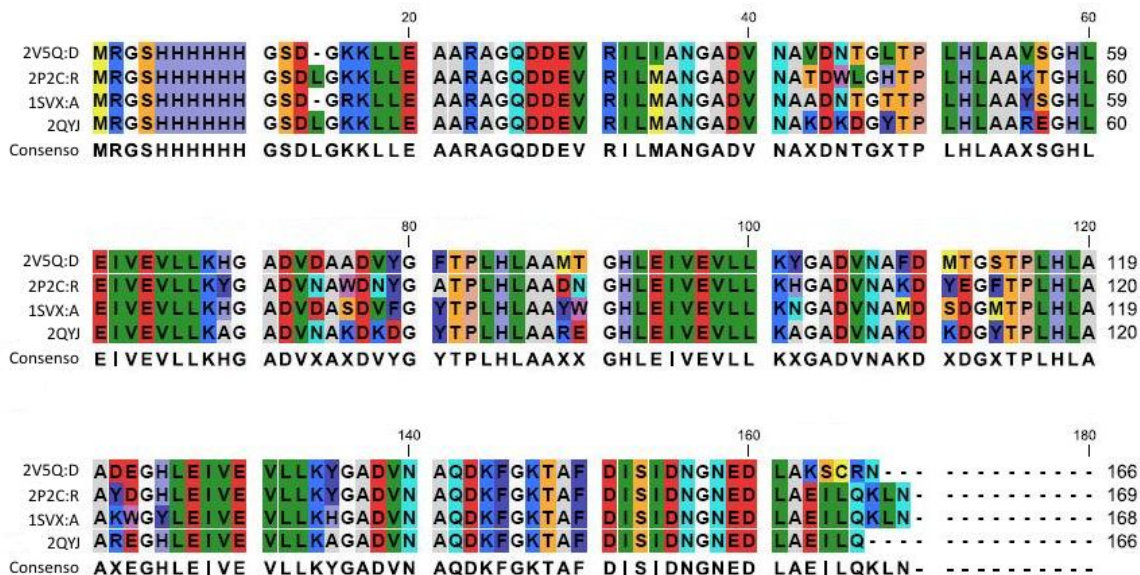
Dado el diseño aquí planteado, y debido a la distribución de las secuencias señales de recombinación, el producto esperado de la recombinación sería una secuencia codificante para una DARPina. Esta secuencia estará conformada por la región C1, al menos uno de los dominios variables de la región V1 y uno de la región V2+C2 (Figura 16).



**Figura 16.** Esquema del sistema de recombinación con la distribución de las secuencias señales en tres diferentes regiones de ADN. C1: Región constante 1, V1: Región variable 1, C2: Regiones variable 2 y constante 2.

### Identificación de módulos constantes y variables de las DARPinas

Se identificaron los módulos que constituyen a las DARPinas y los lugares en los que estos se pudieran separar estructuralmente, de manera que al insertar los aminoácidos correspondientes a las secuencias *loxP* no se afectara la estructura tridimensional característica de las DARPinas. Esto se debe a que durante la recombinación las secuencias *loxP* no se eliminan. Para identificar los módulos estructurales de las DARPinas se alinearon las secuencias de aminoácidos de cuatro DARPinas diferentes, con los identificadores PDB: 2QYJ, 2V5Q:D, 2P2C:R y 1SVX:A (Figura 17).



**Figura 17.** Alineamiento de las DARPinas 2QYJ, 2V5Q:D, 2P2C:R y 1SVX:A. Se muestra también la secuencia consenso. En colores similares se agrupa a los distintos residuos cuya polaridad es parecida.

A partir de este alineamiento, y con ayuda de la estructura tridimensional de la DARPina consenso (PDB ID: 2QYJ), cuyos módulos variables no tienen residuos sustituidos (Merz 2008), se identificaron los módulos constantes N-terminal (DC1), C-terminal (DC2) y la estructura del módulo variable (DV), así como los residuos que se sustituyen en este último. Las secuencias de aminoácidos obtenidas de este alineamiento pueden verse en la Tabla 2. Además se decidió remover del módulo constante DC1 la cola de histidinas y el “linker” formado por los aminoácidos GSDLG que se encontraban en el extremo amino-terminal, debido a que la cola de histidinas formará parte del extremo C-terminal.

**Tabla 4.** Secuencia de aminoácidos de los módulos constantes y del módulo variable de las DARPinas.

<b>Módulo</b>	<b>Núcleótidos (Respecto a la DARPina 2QYJ)</b>	<b>Secuencia de aminoácidos</b>
<b>Constante N-terminal (DC1)</b>	1-37	MRKKLLEAARAGODDEVRLMANG
<b>Constante C-terminal (DC2)</b>	136-186	GADVNAQDKFGKTAFDISIDNGNEDLAEILQ
<b>Variable (DV)</b>	38-70, 71-103, 104-135	ADVNAKDKDGYTPLHLAAXXGHLEIVEVLLKAG

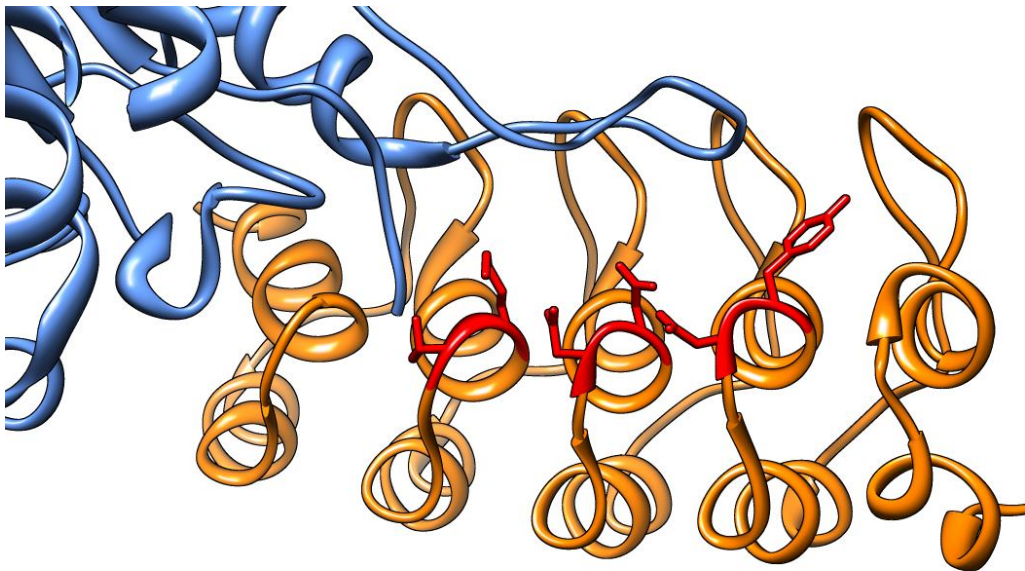
De los seis residuos que cambian entre los distintos módulos variables (Figura 18 A) Merz *et al.* 2008), se decidió variar solamente dos, debido a que se encuentran en la interfaz de reconocimiento (Figura 18 B). Esto se hizo con el fin de no alterar la estabilidad estructural de la proteína final.



**A**



**B**



**Figura 18.** Ubicación de los residuos variables en las DARPinas. **A.** DARPina consenso (PDB ID: 2QYJ), en rojo se muestran los residuos que se sustituyen en las diferentes DARPinas. **B.** Interacción de una DARPina específica (anaranjado) y su proteína blanco (PDB ID: 2V5Q), la cinasa Plk-1 (azul). En rojo se muestran los residuos que se seleccionaron dentro del módulo variable para ser sustituidos.

### Selección de secuencias *loxP* heterólogas

Dado que en el sistema Cre/*loxP* las secuencias *loxP* no se eliminan durante el proceso de recombinación, se comprobó que la secuencia codificante final tuviera los marcos de lectura abierto correctos (Tabla 5). Para esto se tradujeron *in silico* las secuencias *loxP* heterólogas descritas por Siegel *et al.*, (2001) en sus tres marcos de lectura para saber en cuál marco de lectura se podrían insertar las secuencias *loxP*.

**Tabla 5.** Traducción *in silico* de secuencias *loxP* heterólogas.

Secuencia	Marco de lectura 1	Marco de lectura 2	Marco de lectura 3
<i>loxP</i> WT	ITSYSIHYTEKL	*LRIAYIIRSY	NFV*HTLYEV
<i>loxP</i> 511	ITSYSIHYTEKL	*LRIVYIIRSY	NFV*YTLYEYV
<i>loxP</i> 511-I	ITSYNVYYTKL	*LRIMYTIRSY	NFV*CILYEYV
<i>loxP</i> FAS	TTSYIPFYTKL	QLRIYLSIRSC	NFVYTFLYEVV
<i>loxP</i> 2272	ITSYRIPYTEKL	*LRIGYLIRSY	NFV*DTLYEV
<i>loxP</i> 5171	ITSYSTHYTKL	*LRIVHIIRSY	NFV*YTLYEYV
* Codón de paro			

Además, se tomó en cuenta las tasas de recombinación entre las diferentes secuencias *loxP* heterólogas. Se excluyeron las secuencias que presentan una tasa de recombinación heteróloga mayor a 2% con cualquiera de las secuencias *loxP* reportadas por Siegel *et al.* (2001). Los pares de secuencias *loxP* heterólogas resultantes se muestran en la tabla 6.

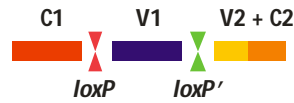
**Tabla 6.** Pares de secuencias heterólogas *loxP* resultantes.

Par de secuencias <i>loxP</i> heterólogas	% de recombinación heteróloga entre las secuencias
FAS/WT	0.2
WT/2272	0.5
FAS/2272	1.7

### **Predicción de estructura terciaria de los productos de recombinación**

Se revisó también que la estructura de la proteína final no se alterara en comparación con la estructura de la DARPina consenso, tomando en cuenta las secuencias de aminoácidos que resultarían de las secuencias *loxP*. Las secuencias de aminoácidos para las cuales se predijo su estructura terciaria tienen la siguiente estructura: DC1-*loxP*-DV1-*loxP'*-DV2-DC2 (Figura 19). *loxP* y *loxP'* corresponden a los distintos pares de sitios *loxP* heterólogos que se utilizaron (WT/FAS, WT/2272 y FAS/2272).

**A**



**B**

**WT/FAS**

**Dc1**      **loxP WT**      **V1**      **loxP FAS**      **V2**      **C2**

MRGSDLGKKLLEAARAGODDEVRLMANG TTSYIPFYTKLL ADVNAKDKDGYTPLHLAAYWGHLEIVEVLLKAG ITSRYRIPYTKLL ADVNAKDKDGMTPLHLAAKWGHLEIVEVLLKAG ADVNAODKFGKTAFDISIDNGNEDLAEILOGSHHHHHH

**C**

**WT/2272**

**Dc1**      **loxP WT**      **V1**      **loxP 2272**      **V2**      **C2**

MRGSDLGKKLLEAARAGODDEVRLMANG TTSYIPFYTKLL ADVNAKDKDGYTPLHLAAYWGHLEIVEVLLKAG ITSRYRIPYTKLL ADVNAKDKDGMTPLHLAAKWGHLEIVEVLLKAG ADVNAODKFGKTAFDISIDNGNEDLAEILOGSHHHHHH

**D**

**FAS/2272**





**Dc1**      **loxP FAS**      **V1**      **loxP 2272**      **V2**      **C2**

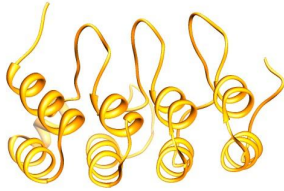

MRGSDLGKKLLEAARAGODDEVRLMANG TTSYIPFYTKLL ADVNAKDKDGYTPLHLAAYWGHLEIVEVLLKAG ITSRYRIPYTKLL ADVNAKDKDGMTPLHLAAKWGHLEIVEVLLKAG ADVNAODKFGKTAFDISIDNGNEDLAEILOGSHHHHHH

**Figura 19.** Estructura de las secuencias de proteínas finales con diferentes combinaciones de secuencias *loxP* heterólogas. **A.** Estructura general de las secuencias. **B, C y D.** Secuencias de aminoácidos de cada secuencia.

En la tabla 7, se muestra un modelo tridimensional de las proteínas finales para cada secuencia de aminoácidos de los cinco generados por el servicio en línea I-TASSER; así como su sobre-posición con la DARPina consenso. También se muestra el Valor TM, que es una medida de la similitud estructural entre dos estructuras. El modelo que se muestra es el de mayor Valor C, índice que mide la calidad de los modelos predichos por el servicio en línea. Ambos valores son generados por el software en línea I-TASSER. Los cinco modelos arrojados se muestran en el Anexo 3, junto con sus valores C.

**Tabla 7.** Comparación estructural entre las diferentes secuencias de proteínas finales y la DARPina consenso 2QYJ.

<b>WT/FAS</b>		
<b>Modelo</b>	<b>Sobre-posición con 2QYJ</b>	<b>Valor TM* (WT/FAS-2QYJ)</b>
		0.8
<b>WT/2272</b>		
<b>Modelo</b>	<b>Sobre-posición con 2QYJ</b>	<b>Valor TM* (WT/2272-2QYJ)</b>
		0.805

FAS/2272		
Modelo	Sobre-posición con 2QYJ	Valor TM* (FAS/2272-2QYJ)
		0.808
*Valor TM: Valor calculado de la similitud entre las topologías de dos estructuras proteicas. Asigna un mayor valor ponderado a los puntos cercanos entre las estructuras que puntos alejados.		

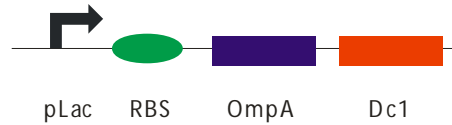
Dado que no existe diferencia significativa entre la estructura de la proteína final al usar cualquiera de las tres combinaciones de secuencias *loxP*, se decidió descartar la secuencia *loxP* WT y usar el par FAS/2272 por ser un par con una tasa de recombinación heteróloga de 1.7% y porque tanto FAS como 2272 presentan una baja tasa de recombinación con las demás secuencias reportadas por Siegel *et al.* 2001, esto podría ser útil si posteriormente se decide añadir otro módulo variable, para lo cual se requeriría una secuencia *loxP* heteróloga adicional, que podría ser la secuencia *loxP* 511.

### **Selección del promotor, señal de localización a membrana y sitio de unión a ribosoma**

La región C1 incluirá, además del módulo constante DC1 que se obtuvo de las DARPinas, las secuencias nucleotídicas necesarias para iniciar la transcripción, una secuencia de señalización a membrana; así como el prefijo y el sufijo necesarios para el ensamblaje estándar.

Las secuencias nucleotídicas usadas se muestran en el Anexo 2. El promotor elegido fue el pLac, que es inducible por IPTG (BioBrick BBa\_R0010). La secuencia de localización a membrana externa es un fragmento de la proteína de *E. coli* OmpA, que contiene el péptido señal de localización al lugar deseado (BioBrick BBa\_K103006). Esto se realizó con el fin de que, una vez que se da la recombinación, y la proteína expresada se despliegue

en la superficie de la bacteria. El sitio de unión a ribosoma (RBS) produce un nivel constitutivo de traducción (BioBrick BBa\_B0030). La estructura puede verse en la Figura 20.



**Figura 20.** Esquema de la región C1. **pLac.** Promotor. **RBS.** Sitio de unión a ribosoma. **OmpA.** Secuencia de localización a membrana externa. **Dc1.** Módulo constante N-terminal de DARPinas.

### Secuencias finales de las regiones C1, V1 y V2+C2

Como se mencionó, las secuencias de todas las regiones incluirán el prefijo y el sufijo de los BioBricks en los extremos 5' y 3', respectivamente; esto con el fin de facilitar su ensamblaje. Además de las secuencias BBa\_R0010, BBa\_K103006 y BBa\_B0030, también se encuentran incluidas en la región C1 dos "cicatrices". Estas se forman en el proceso de ensamblaje estándar entre dos BioBricks. Aunque no son funcionales se incluyen porque las regiones reguladoras de la transcripción que se han probado usando BioBricks tienen este tipo de cicatrices entre el promotor y el RBS, y entre el RBS y el sitio de inicio de la transcripción. Las secuencias finales son las siguientes:

<b>Región C1</b>	
<b>Prefijo</b>	GAATTCGCGGCCGCTTCTAGAG
<b>pLac</b>	CGTTGACACCATCGAATGGCGCAAACCTTTCGCGGTATGGCATGATAGCG CCCGG
<b>Cicatriz</b>	TACTAGAG
<b>RBS</b>	AAAGAGGAGAAA
<b>Cicatriz</b>	TACTAG
<b>OmpA</b>	ATGCATATGAAAGCTACTAAACTGGTACTGGGCGCGGTAATCCTGGGTTCTACTC TGCTGGCAGGTTGCTCCAGCAACGCTAAAATCGATCAGGGAATTAACCCGATGT TGGCTTTGAAATGGGTTACGACTGGTTAGGTCGTATGCCGTACAAAGGCAGCGTT

GAAAACGGTGCATACAAAGCTCAGGGCGTTCAACTGACCGCTAAACTGGGTTAC  
CCAATCACTGACGACCTGGACATCTACACTCGTCTGGGTGGCATGGTATGGCGTG  
CAGACACTAAATCCAACGTTTATGGTAAAAACCACGACACCGGGCGTTTCTCCGGT  
CTTCGCTGGCGGTGTTGAGTACGCGATCACTCCTGAAATCGCTACCCGTCTGGAA  
TACCAGTGGACCAACAACATCGGTGACGCACACACCATCGGCACTCGTCCGGACA  
ACGGCGGAGGTTCTGGAGGAGGGAGCTC

**DC1** ATGCGCGGCAGCGATCTGGGCAAAAACTGCTGGAAGCGGCGCGCGGGCCA  
GGATGATGAAGTGCGCATTCTGATGGCGAACGGC

**loxP FAS** ACAACTTCGTATATACCTTTCTATACGAAGTTGTTG

**Sufijo** TACTAGTAGCGGCCGCTGCAG

### Región V1

**Prefijo** GAATTCGCGGCCGCTTCTAGAG

**loxP FAS** ACAACTTCGTATATACCTTTCTATACGAAGTTGTTG

**V 1.1** GCGGATGTGAACGCGAAAGATAAAGATGGCTATACCCCGCTGCATCTGGCGGC  
GTATTGGGGCCATCTGGAAATTGTGGAAGTGCTGCTGAAAGCGGGC

**loxP 2272** ATAACTTCGTATAGGATACCTTATAACGAAGTTATTA

**loxP FAS** ACAACTTCGTATATACCTTTCTATACGAAGTTGTTG

**V 1.2** GCGGATGTGAACGCGAAAGATAAAGATGGCGGACCCCGCTGCATCTGGCGGC  
GGATAACGGCCATCTGGAAATTGTGGAAGTGCTGCTGAAAGCGGGC

**loxP 2272** ATAACTTCGTATAGGATACCTTATAACGAAGTTATTA

**loxP FAS** ACAACTTCGTATATACCTTTCTATACGAAGTTGTTG

**V 1.3** GCGGATGTGAACGCGAAAGATAAAGATGGCTTACCCCGCTGCATCTGGCGGCG  
ATGACCGGCCATCTGGAAATTGTGGAAGTGCTGCTGAAAGCGGGC

**loxP 2272** ATAACTTCGTATAGGATACCTTATAACGAAGTTATTA

**Sufijo** TACTAGTAGCGGCCGCTGCAG

### Región V2+C2

**Prefijo** GAATTCGCGGCCGCTTCTAGAG



**loxP 2272**   ATAACTTCGTATAGGATACCTTATACGAAGTTATTA

**V 2.1**

          GCGGATGTGAACGCGAAAGATAAAGATGGCATGACCCCGCTGCATCTGGCGGCG  
          AAATGGGGCCATCTGGAAATTGTGGAAGTGCTGCTGAAAGCGGGC

**C2**           GCGGATGTGAACGCGCAGGATAAATTTGGCAAACCGCGTTTGATATTAGCATT  
          GATAACGGCAACGAAGATCTGGCGGAAATTCTACAGGGCAGCCATCATCATCAT  
          CATCATTAATAA

**Terminador** CGCAAAAACCCCGCTTCGGCGGGGTTTTTTTCGC

**loxP 2272**   ATAACTTCGTATAGGATACCTTATACGAAGTTATTA

**V 2.2**

          GCGGATGTGAACGCGAAAGATAAAGATGGCTTTACCCCGCTGCATCTGGCGGCG  
          TATGATGGCCATCTGGAAATTGTGGAAGTGCTGCTGAAAGCGGGC

**C2**           GCGGATGTGAACGCGCAGGATAAATTTGGCAAACCGCGTTTGATATTAGCATT  
          GATAACGGCAACGAAGATCTGGCGGAAATTCTACAGGGCAGCCATCATCATCAT  
          CATCATTAATAA

**Terminador** CGCAAAAACCCCGCTTCGGCGGGGTTTTTTTCGC

**loxP 2272**   ATAACTTCGTATAGGATACCTTATACGAAGTTATTA

**V 2.3**

          GCGGATGTGAACGCGAAAGATAAAGATGGCAGCACCCCGCTGCATCTGGCGGCG  
          GGATGAAGGCCATCTGGAAATTGTGGAAGTGCTGCTGAAAGCGGGC

**C2**           GCGGATGTGAACGCGCAGGATAAATTTGGCAAACCGCGTTTGATATTAGCATT  
          GATAACGGCAACGAAGATCTGGCGGAAATTCTACAGGGCAGCCATCATCATCAT  
          CATCATTAATAA

**Terminador** CGCAAAAACCCCGCTTCGGCGGGGTTTTTTTCGC

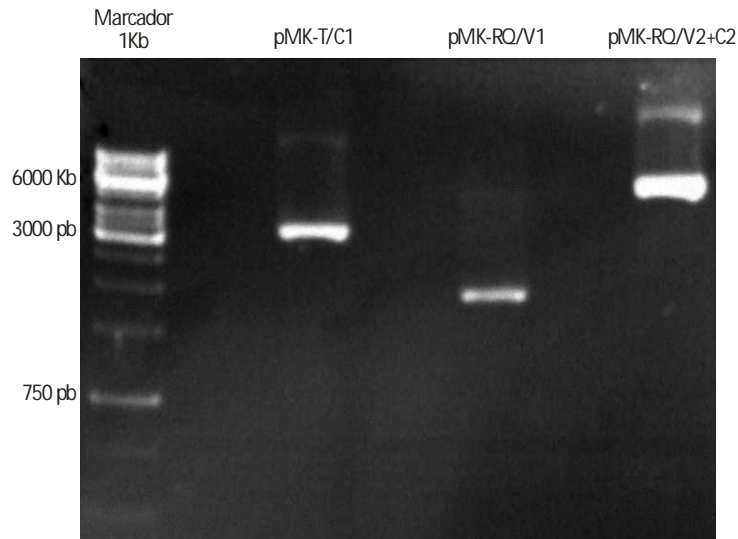
**Sufijo**       TACTAGTAGCGGCCGCTGCAG

### Síntesis de secuencias

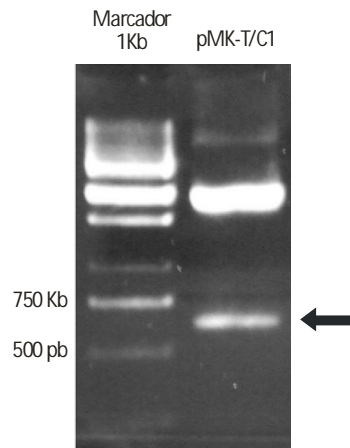
Las secuencias correspondientes a las regiones C1, V1 y V2+C2 se sintetizaron comercialmente (Figura 21A). El tamaño de las secuencias sintetizadas y el plásmido en el que se recibieron se muestran en la tabla 2.

Los plásmidos se transformaron en *E. coli* TOP 10 (Invitrogen) y se verificó la presencia del inserto deseado mediante purificación de plásmido de las colonias transformantes y PCR, utilizando los oligonucleótidos AF y AR, que son específicos para el prefijo y el sufijo (FIGURA 21 B, C Y D).

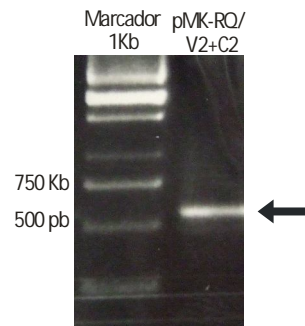
**A**



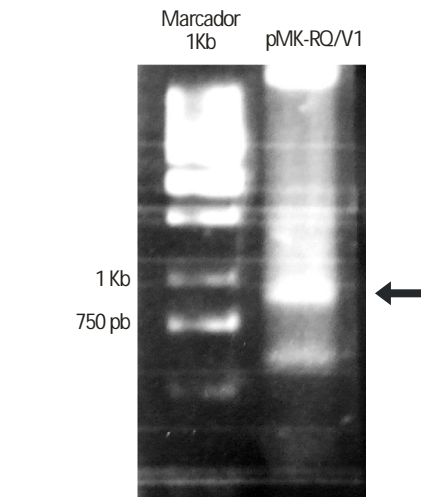
**B**



**C**



**D**



**Figura 21.** Comprobación mediante PCR de la amplificación de vectores construidos con secuencias sintetizadas comercialmente y transformados en *E. coli*. **A.** Gel de agarosa 1% de plásmido purificado de colonias transformantes. **B.** PCR de pMK-T/C1 para amplificar C1 (715 pb) . **C.** PCR de pMK-RQ/V1 para amplificar V1 (554 pb). **D.** PCR de pMK-RQ/V2+C2 para amplificar V2+C2 (908 pb). Las flechas indican los pesos moleculares esperados para los distintos amplicones.

Para estas amplificaciones se utilizaron los oligonucleótidos AF y AR.

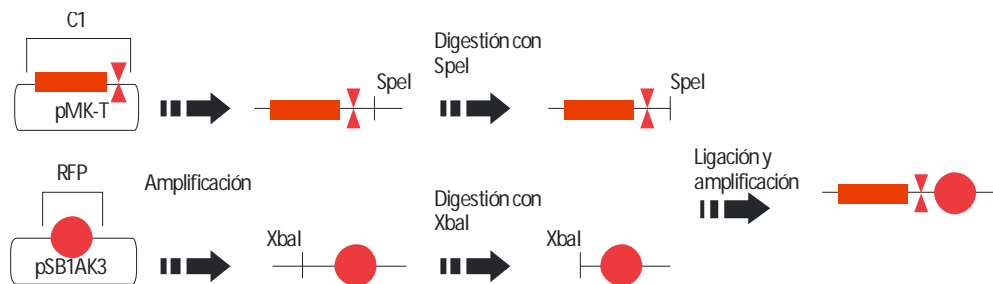
### Ensamblaje de secuencias

Se siguieron tres estrategias para probar el Sistema de Recombinación: 1) Ensamblaje de las tres regiones en un vector mediante ligación y amplificación *in vitro*; 2) Ensamblaje de las regiones C1 y V2+C2 en un vector; 3) Ensayo de recombinación *in vitro*. Los resultados de estas estrategias se describen a continuación:

#### 1) Ensamblaje de las tres regiones en un vector mediante ligación y amplificación *in vitro*

El objetivo de esta estrategia es incluir en un solo plásmido a las tres regiones, con el fin de probar *in vivo* el sistema de recombinación.

Todas las secuencias de las que se partió cuentan con la secuencia correspondiente al prefijo y al sufijo, por lo que dos secuencias pueden ensamblarse siguiendo los pasos del ensamblaje estándar. Esto permite juntar de manera sistemática dos secuencias mediante digestión y ligación *in vitro* (Figura 22).

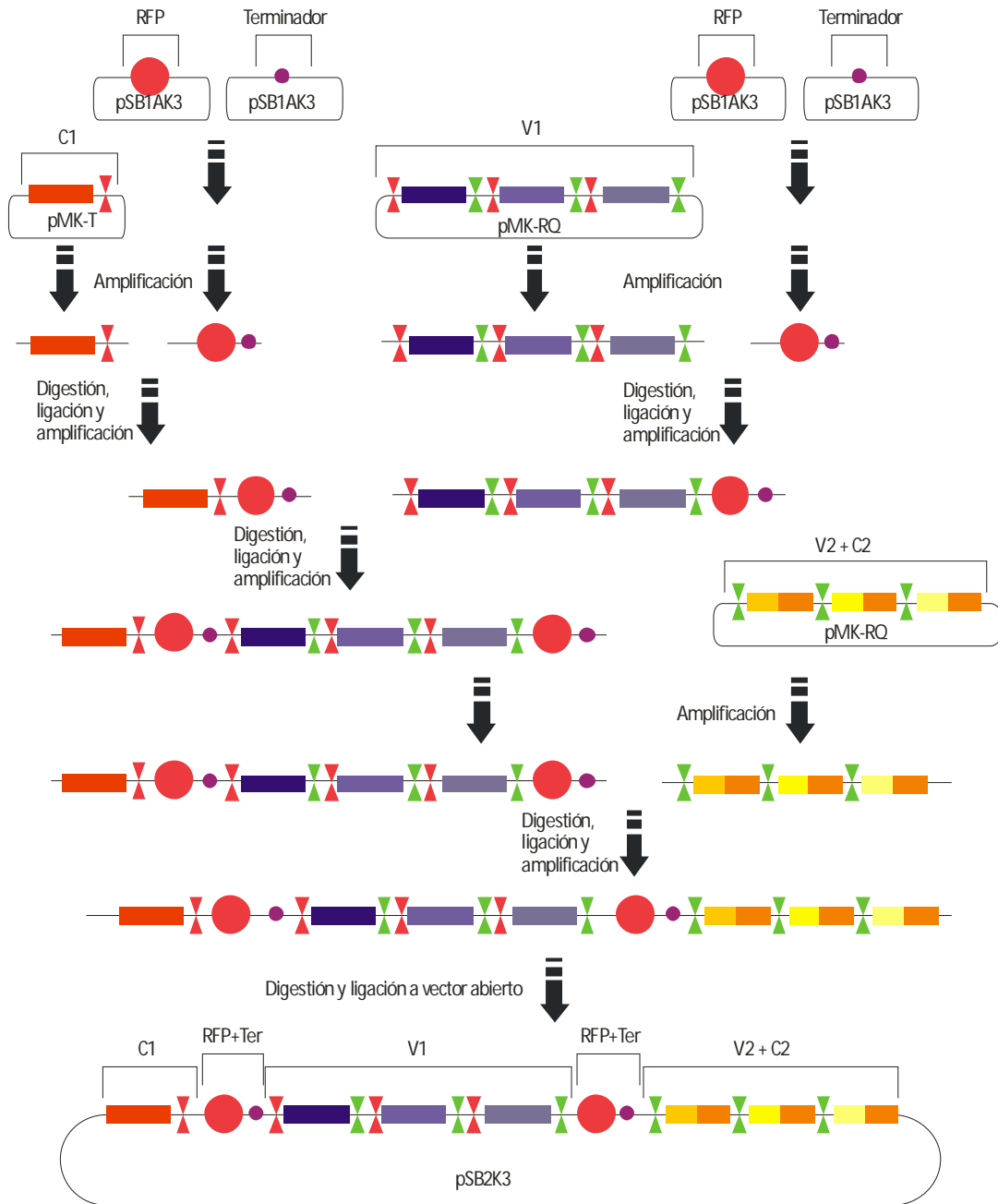


**Figura 22.** Ejemplo del proceso de ensamblaje *in vitro* para la unión de la región C1 y la RFP.

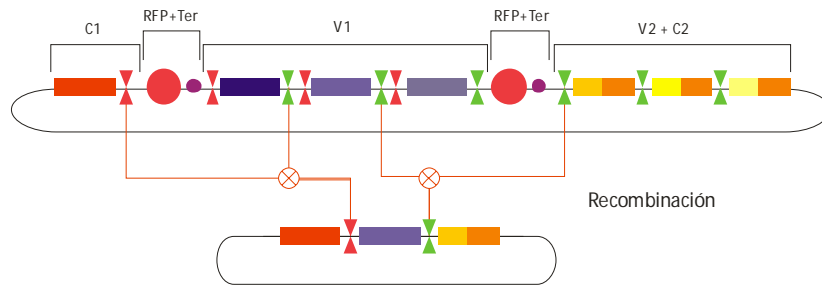
De esta manera se buscó ensamblar en una sola secuencia a las tres regiones del sistema de recombinación, separadas entre sí por una secuencia codificante para una RFP (Figura 23A). Esta secuencia funcionaría como reportera del proceso de recombinación, debido a que de ser exitoso el proceso de recombinación, la secuencia de la RFP sería escindida (Figura 23B). Por otro lado, también funcionaría como una secuencia espaciadora entre las secuencias *loxP* adyacentes. Se decidió que el vector final en el que estaría contenida la

secuencia ensamblada fuera el plásmido pSB2K3, un plásmido cuyo número de copias por célula puede regularse. El plásmido, además de contener el origen de replicación F', contiene también el origen de replicación de la fase lítica del bacteriófago P1 y al gen codificante (*repL*) para la proteína responsable de iniciar la replicación en dicho origen de replicación bajo el control de un promotor regulado por el inhibidor LacI (<http://partsregistry.org/Part:pSB2K3>). De esta manera se podría tener un mayor control del número de recombinaciones que pudieran suceder dentro de una célula.

A



**B**

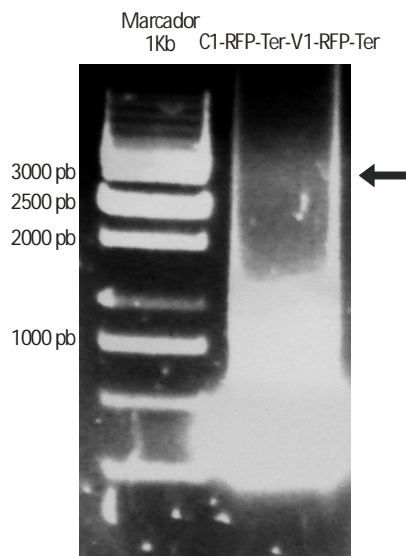


**Figura 23.** Ensamblaje de las tres regiones en un vector mediante ligación y amplificación *in vitro*.

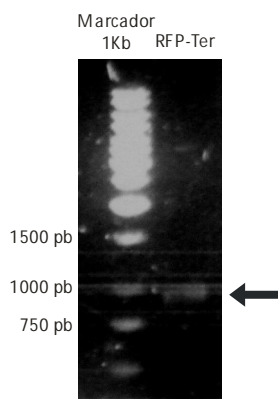
**A.** Esquema del ensamblaje de las tres regiones en un vector mediante ligación y amplificación *in vitro*. **B.** Esquema del proceso de recombinación de la secuencia final.

Siguiendo esta estrategia se logró ensamblar la secuencia C1-RFP-Ter-V1-RFP-Ter, que se comprobó por tamaño (Figura 24A). Además se purificó el amplicón correspondiente al tamaño esperado y se secuenció (Anexo 4) con el oligonucleótido A2F. Los pasos intermedios, esto es, las ligaciones RFP-Ter, C1-RFP-Ter y V1-RFP-Ter se muestran en la figura 24B, 24C y 24D.

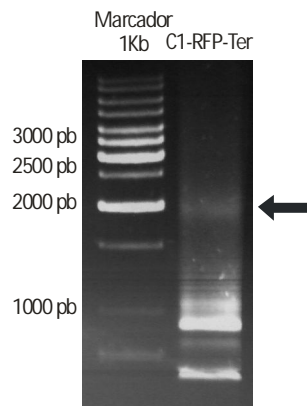
**A**



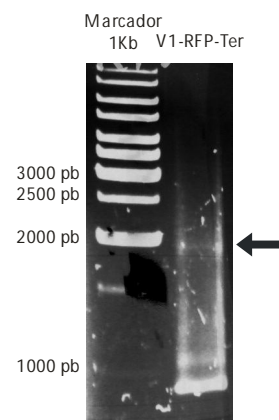
**B**



**C**



**D**



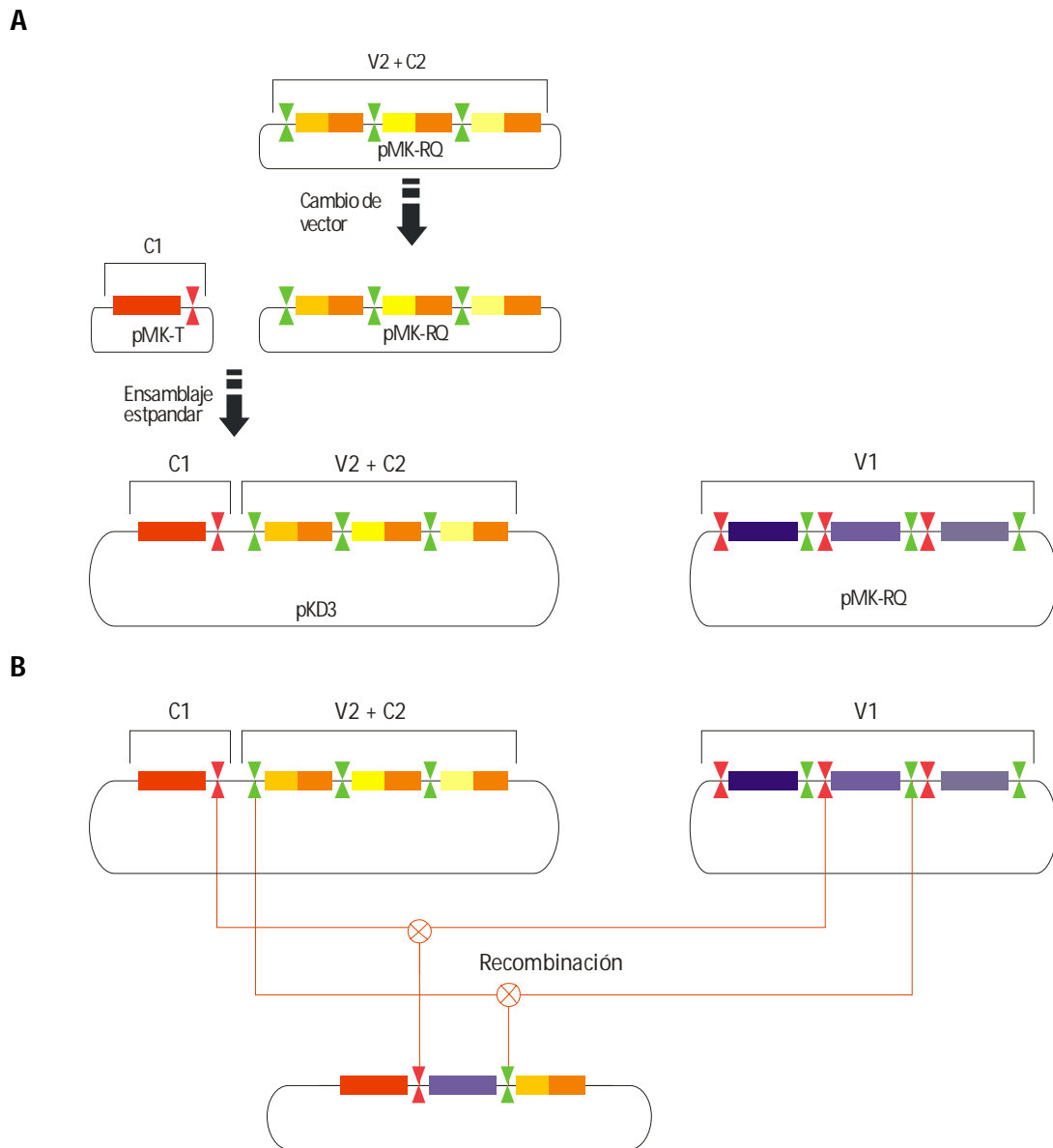
**Figura 24.** Amplificación de secuencias ensambladas *in vitro*. **A.** C1-RFP-Ter-V1-RFP-Ter. **B.** RFP-Ter. **C.** C1-RFP-Ter. **D.** V1-RFP-Ter. (Para todas las amplificaciones se utilizaron los oligonucleótidos AF y AR).

Amplificaciones de distintas ligaciones de la secuencia C1-RFP-V1-RFP con la región V2+C2, no produjeron bandas en el peso esperado; por lo que se intentó una estrategia diferente de ensamblaje que se describe a continuación: ligar las regiones C1 y V2+C2 en un vector que recombinara con la región V1 de un vector distinto.

## **2) Ensamblaje de las regiones C1 y V2+C2 en un vector**

El objetivo de esta estrategia de ensamblaje fue tener dos plásmidos, uno con las regiones C1 y V2+C2; y otro con la región V1. Durante el proceso de recombinación, los sitios *loxP* de uno de los módulos de la región V1 recombinaría con el sitio *loxP* de la región C1 y con uno de la región V2+C2 (Figura 25 A). Esto produciría al momento de la recombinación una secuencia codificante para una proteína funcional dentro de un vector. Se decidió que este vector fuera el plásmido pKD3, que tiene un marcador de resistencia a cloranfenicol, debido a que se planeó co-transformarlo con el plásmido pMK-RQ (con marcador de resistencia a kanamicina) que contiene la secuencia V1 y con el plásmido pJW168 (con marcador de resistencia a ampicilina), el cual contiene el gen codificante para la recombinasa Cre inducible por IPTG. De esta manera, al inducirse la recombinación se generaría una secuencia funcional dentro del plásmido pKD3 (Figura 25 B).

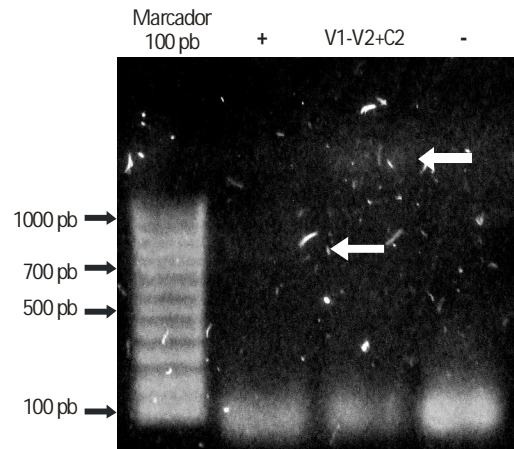




**Figura 25.** Esquema del Ensamblaje de las regiones C1 y V2+C2 en un vector. **A.** Esquema de los vectores con las secuencias finales. **B.** Esquema del proceso de recombinación de esta estrategia.

Se realizó exitosamente el cambio de vector de la secuencia V2+C2 mediante corte del plásmido pKD3 y de la secuencia V2+C2 con la enzima de restricción NotI. Sin embargo, no se obtuvieron colonias transformantes con la ligación entre las regiones C1 y V2+C2, aunque la amplificación de la reacción de ligación con oligonucleótidos específicos para el extremo 5' de la región C1 y para el extremo 3' de V2+C2 mostró que la ligación sí se llevó

a cabo (Figura 26). Esto puede deberse a que la secuencia resultante C1-V2+C2 afecta la capacidad de replicación del plásmido dentro de la célula, ya que el sitio de inserción de la secuencia se encuentra en la base 1447 del plásmido PKD3, a sólo un nucleótido del inicio del origen de replicación.

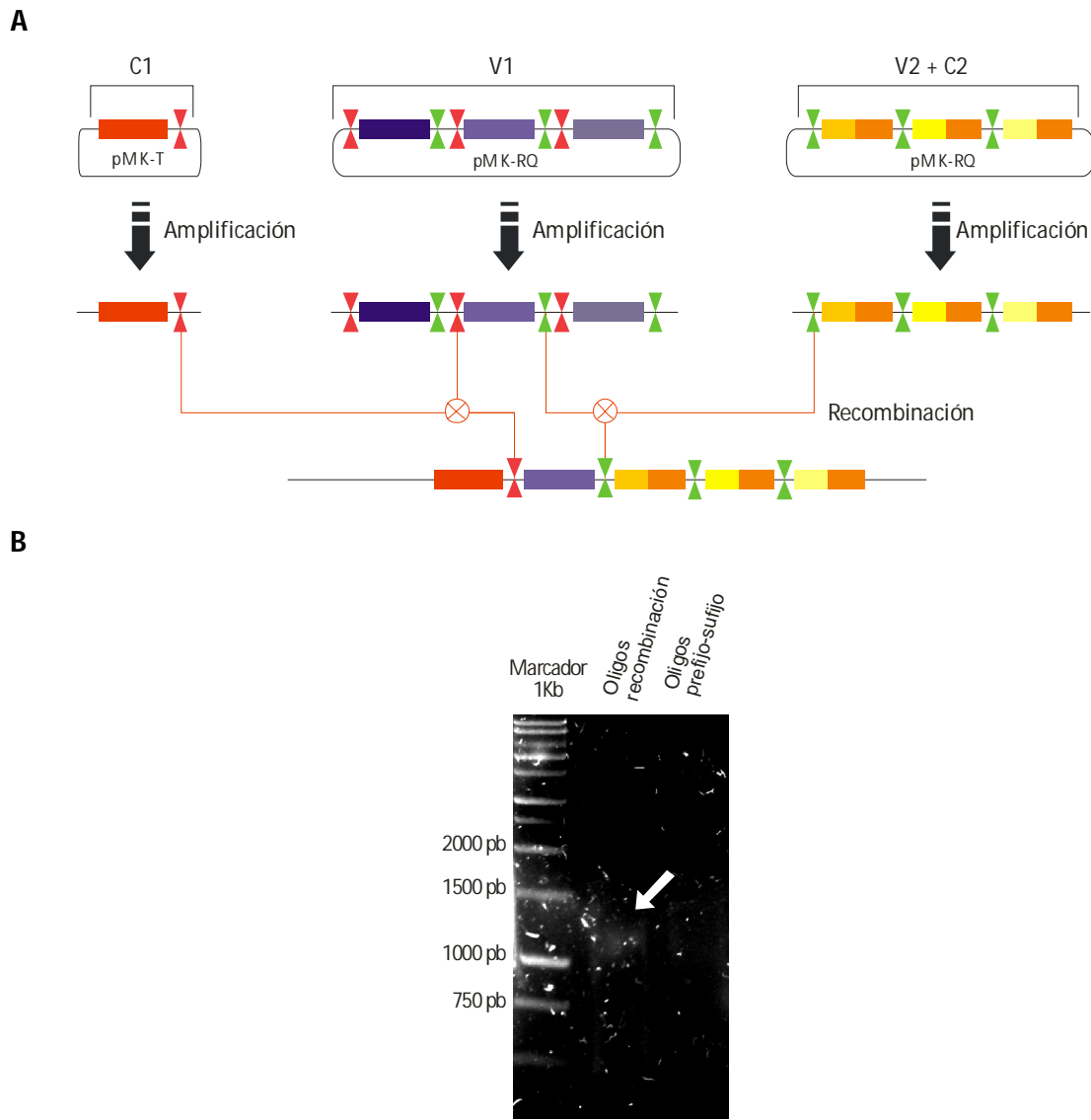


**Figura 26.** Amplificación de la ligación V1-V2+C2. Para la amplificación se utilizaron los oligonucleótidos AF y AR.

Debido a que no se pudieron obtener colonias transformantes se decidió hacer un ensayo de recombinación *in vitro*.

### 3) Ensayo de recombinación *in vitro*

Para realizar el ensayo de recombinación *in vitro* (Figura 27A) se siguió la metodología de Abremski y Hoess (1983), que se detalla en la sección de Materiales y métodos. La amplificación de los productos de recombinación se realizó utilizando oligonucleótidos específicos para una sección de la región C1, para el "forward", y para una sección de la región C2, el "reverse". Los pesos esperados de los productos de la recombinación eran de entre 1081 y 1252 pares de bases (Figura 27 B), dependiendo de si durante la recombinación quedó uno o dos módulos variables de la región V1. Esto se debe a que la cercanía entre dos secuencias *loxP* dentro de una sola región es de 99 pares de bases y eso podría evitar la recombinación.



**Figura 27.** Estrategia de recombinación in vitro. **A.** Esquema de la estrategia. **B.** Amplificación final específica del producto de recombinación utilizando oligonucleótidos específicos para la región C1 (oligonucleótido A1F) y para la región V2+C2 (oligonucleótido A3R). También se muestra la amplificación inicial (oligonucleótidos AF y AR).

El producto de esta amplificación se purificó y se secuenció. La banda que se amplificó no corresponde al producto de la recombinación. La secuencia corresponde a la secuencia C1 junto con una secuencia del vector. Esto puede deberse a un apareamiento inespecífico entre el oligonucleótido específico para la región 3' de la secuencia V2+C2 y el vector.

## Discusión

Aunque ya se ha propuesto un sistema procarionte para la selección y producción de proteínas capaces de interactuar con alguna molécula blanco ([http://2008.igem.org/Wiki/Team:Warsaw/igem\\_project.htm](http://2008.igem.org/Wiki/Team:Warsaw/igem_project.htm)), aquí se presenta el diseño de un nuevo sistema inspirado en el proceso de recombinación V(D)J, utilizando como plataforma estructural a las DARPinas.

El sistema que se propone combina las ventajas estructurales de las DARPinas con un proceso de recombinación de módulos estructurales para una mejor exploración del espacio potencial de secuencias de una DARPina, de manera análoga a como sucede durante la generación de inmunoglobulinas del sistema inmune. Además, este mecanismo es celular, lo que facilitaría y disminuiría los costos en la generación de bibliotecas con una gran diversidad de secuencias.

Para lograr abarcar el mayor número de secuencias posibles y por lo tanto aumentar la probabilidad de encontrar la secuencia adecuada, es necesario hacer una primera exploración, con un mecanismo similar al proceso de recombinación V(D)J, a diferencia del método propuesto por el equipo iGEM 2008 de la Universidad de Varsovia, que se basa exclusivamente en la modificación de nucleótidos mediada por la enzima AID. La principal ventaja del sistema aquí propuesto es que durante la recombinación se genera una biblioteca cuya base estructural se encuentra definida. De esta manera se asegura que la primera exploración del espacio de secuencias posibles no genere secuencias de aminoácidos estructuralmente inestables, como lo sugiere la predicción estructural del producto de la recombinación. Después de esta primera exploración de secuencias se podría hacer una segunda exploración, añadiendo al sistema el módulo de la enzima AID.

Además del mecanismo de generación, se propone también un método de selección, que consiste en la localización de la proteína producida a la membrana externa de *E. coli* mediante la fusión de la DARPina en su extremo amino-terminal a un fragmento de la

proteína OmpA. Esto direccionaría a la DARPina hacia la membrana externa de *E. coli*, con lo que la proteína se desplegaría en la superficie y permitiría una selección de las clonas que presentaran afinidad por un determinado ligando, similar a la estrategia reportada para seleccionar los fragmentos scFv mediante selección celular activada por fluorescencia en *Saccharomyces cerevisiae* (Feldhaus *et al.* 2003).

Al intentar construir el sistema propuesto a partir de las secuencias sintetizadas (C1, V1 y V2+C2) y las secuencias obtenidas del "Registry of Standard Biological Parts" se presentaron problemas durante el ensamblaje. En la primera estrategia, que consistió en ensamblar *in vitro* la secuencia final para posteriormente insertarla en un plásmido, se logró ensamblar la secuencia C1-RFP-Ter-V1-RFP-Ter. Sin embargo, no se pudo ensamblar la secuencia V2+C2 con la secuencia C1, posiblemente como consecuencia de la formación de estructuras que impidieron la ligación o la replicación del plásmido, pues se ha reportado que secuencias altamente repetitivas pueden dar lugar a la formación de estructuras secundarias en el ADN tanto *in vitro* (Gacy *et al.* 1995) como *in vivo* (Catasti *et al.* 1994, Gacy *et al.* 1995, Bzymek y Lovett 2001, Voineagu *et al.* 2008). Estas estructuras pueden ser las causantes de la dificultad para ensamblar la secuencia final en la primera estrategia, evitando que los extremos cohesivos de las dos distintas secuencias se aparearan y pudieran ligarse.

En la segunda estrategia, donde se intentó ensamblar las secuencias C1 y V2+C2 en el vector pKD3, la amplificación del producto de la ligación mostró que la ligación entre estas dos secuencias si se llevó a cabo, sin embargo no se obtuvieron colonias transformantes. Esto pudo deberse también a la formación de estructuras en el plásmido ya ligado, que evitó la replicación del mismo. Además, de que el sitio de restricción NotI del plásmido pKD3 donde se insertaron las secuencias se encuentra a un par de bases de distancia del origen de replicación del plásmido. De haberse formado una estructura secundaria de ADN cerca del origen de replicación, este impedimento estructural pudo haber inhibido el proceso de replicación, como ya se ha reportado en secuencias Alu (Voineagu *et al.* 2008),

o pudo haber producido inestabilidad en el ADN (Bzymek y Lovett, 2001).

El ensayo de recombinación *in vitro* no arrojó evidencia de que se haya producido la recombinación. Esto pudo haberse debido a una baja concentración de las secuencias sustrato de la recombinación, como se puede ver en la amplificación inicial con los oligonucleótidos específicos para el prefijo y sufijo, donde no se observaron amplicones. La baja concentración de los plásmidos, y por lo tanto una separación entre los diferentes sustratos de recombinación, pudo haber sido la causa de una baja eficiencia de recombinación. Ya se ha reportado que un alejamiento espacial de al menos 150 kb entre sitios *loxP* provoca una baja eficiencia de recombinación, incluso aunque estos se encuentren en la misma cadena de ADN (Coppoolse *et al.* 2005).

## Conclusiones

La secuencia de nucleótidos de una DARPina puede modificarse para contener *sitios loxP*, que, como sugiere sin modificar de manera significativa su estructura tridimensional, como se mostró con la predicción *in silico* de la estructura terciaria de una DARPina producto del mecanismo de recombinación propuesto.

Se diseñaron y sintetizaron las secuencias de ADN que formarían parte de un sistema de recombinación somática sintética procarionte. El sistema se basó en la estructura modular de las DARPinas y en el sistema de recombinación *Cre/loxP*.

Se logró el ensamblaje parcial del sistema de recombinación mediante ensamblaje estándar, sin embargo el sistema propuesto de recombinación no se logró ensamblar por completo. Tampoco se lograron ensamblar los módulos necesarios para probarlo mediante dos estrategias diferentes. La dificultad del ensamblaje pudo deberse a la alta repetitividad de las secuencias diseñadas.

El ensayo de recombinación *in vitro* arrojó resultados negativos, posiblemente por la baja

concentración de secuencias *loxP* heterólogas en el ensayo; lo que evitó que se diera una recombinación exitosa.

## Perspectivas

La repetitividad de las secuencias podría disminuirse retrotranscribiendo de manera ambigua las secuencias de aminoácidos codificantes para los motivos de DARPinas, esto es, utilizando diferentes codones a lo largo de la secuencia para un mismo aminoácido. Sin embargo, aún quedaría la repetitividad asociada al uso de numerosas secuencias *loxP*, las cuáles son imprescindibles para el funcionamiento del sistema de recombinación. De resultar inestables o difíciles de ensamblar estas secuencias debido a las repeticiones de los sitios *loxP*, resultaría imposible la construcción de sistemas de recombinación sintéticos controlados y específicos con la estructura como la que proponemos, debido a que todas las recombinasas específicas que no requieren de factores celulares adicionales, reconocen secuencias que deben repetirse numerosas veces a lo largo de la cadena para lograr un arreglo similar a los loci de las cadenas pesada o ligera en mamíferos. Esto es poco probable, debido a que arreglos similares de secuencias *loxP* se han probado exitosamente, como el sistema de marcaje fluorescente "Brainbow", que estocásticamente expresa combinaciones diferentes de proteínas fluorescentes al interior de cada célula (Weissman *et al.* 2011).

## Bibliografía

1. Abremski K, Hoess R (1983). Bacteriophage P1 site-specific recombination: purification and properties of the Cre recombinase protein. *Journal of Biological Chemistry* 259, 1509-1514.
2. Abul K, Abbas AH, Lichtman SP (2010) Cellular and molecular immunology. 6° ed Philadelphia, USA: Saunders/Elsevier. 566 p
3. Agrawal, A, Eastman, QM, Schatz, DG (1998). Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* 394, 744–751.
4. Albert H, Dale EC, Lee E, Ow DW (1995). Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. *The Plant Journal* 7, 649-659.
5. Amstutz P, Forrer P, Zahnd C, Plückthun A (2001). In vitro display technologies: novel developments and applications. *Current Opinion in Biotechnology*. 12, 400–405.
6. Anderson JC, Voigt CA, Arkin AP (2007). Environmental signal integration by a modular AND gate. *Molecular Systems Biology* 3, 133
7. Andrianantoandro E, Basu S, Karig DK, Weiss R (2006). Synthetic Biology: new engineering rules for an emerging discipline. *Molecular Systems Biology* 2: 2006.0028.
8. Arnold PH, Blake DG, Grindley NDF, Boocock MR, Stark WM (1999). Mutants of Tn3 resolvase which do not require accessory binding sites for recombination activity. *EMBO Journal* 18, 1407–1414
9. Austin S, Ziese M, Sternberg N (1981). A novel role for site-specific recombination in maintenance of bacterial replicons. *Cell* 25,729–736.
10. Balagaddé FK, Song H, Ozaki J, Collins CH, Barnet M, *et al.* (2008). A Synthetic *Escherichia coli* predator-prey ecosystem. *Molecular Systems Biology* 4, 187.
11. Binz HK, Stumpp MT, Forrer P, Amstutz P, Pluckthun A (2003). Designing repeat proteins: Well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *Journal of Molecular Biology* 332, 489–503.
12. Binz HK, Amstutz P, Plückthun A. (2005). Engineering novel binding proteins from nonimmunoglobulin domains. *Nature Biotechnology* 23, 1257–1268.



13. Bloom L, Calavaro, V (2009). FN3: a new protein scaffold reaches the clinic. *Drug Discovery Today* 14, 949-955.
14. Boder ET, Jiang W (2011). Engineering antibodies for cancer therapy. *Annual Review of Chemical and Biomolecular Engineering* 2, 53-75.
15. Boersma YL, Plückthun A (2011). DARPinS and other repeat protein scaffolds: advances in engineering and applications. *Current Opinion in Biotechnology* 22, 849-857
16. Bouhassira EE, Westerman K, Leboulch P (1997). Transcriptional behavior of LCR enhancer elements integrated at the same chromosomal locus by Recombinase-Mediated Cassette Exchange. *Blood* 90, 3332-3344.
17. Borrebaeck CA, Ohlin M (2002). Antibody evolution beyond nature. *Nature Biotechnology* 20, 1189-1190.
18. Bransteitter R, Pham P, Scharff MD, Goodman MF (2003). Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of Rnase. *PNAS* 4102-4107.
19. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P *et al.* (2001) Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nature Genetics* 29, 365 – 371
20. Broach JR, Guarascio VR, Jayaram M (1982). Recombination within the yeast plasmid 2  $\mu$  circle is site-specific. *Cell* 29,227–234.
21. Bushman W, Thompson JF, Vargas L, Landy A (1985). Control of directionality in lambda site-specific recombination. *Science* 230, 906–911.
22. Bzymek M, Lovett ST (2001). Instability of repetitive DNA sequences: The role of replication in multiple mechanisms. *PNAS* 98 8319-8325.
23. Carver T, Bleasby, A (2003). The design of Jemboss: a graphical user interface to EMBOSS. *Bioinformatics* 19, 1837-1843.
24. Canton B, Labno A, Endy D (2008). Refinement and standarization of synthetic biological parts. *Nature biotechnology* 26, 787-793.
25. Catasti P, Gupta G, Garcia AE, Ratliff R, Hong L (1994). Unusual Structures of the Tandem Repetitive DNA Sequences Located at Human Centromeres. *Biochemistry* 33, 3819-3830.

26. Conrath K *et al.* (2005). Antigen binding and solubility effects upon the veneering of a camel VHH in framework-2 to mimic a VH. *Journal of Molecular Biology* 350, 112–125.
27. Cooling MT, Rouilly V, Misirli G, Lawson J, Yu T, *et al.* (2010). Standard virtual biological parts: a repository of modular modeling components for synthetic biology. *Bioinformatics*. 26, 925-931.
28. Coppoolse ER, de Vroomen MJ, van Gennip F, Hersmus BJM, van Haaren MJJ (2005). Size does matter: Cre-mediated somatic deletion efficiency depends on the distance between the target lox-sites. *Plant Molecular Biology* 58, 687-698
29. Cortez-Retamozo V, Backmann N, Senter PD, Wernery U, Baetselier P, *et al.* (2004). Efficient cancer therapy with a nanobody-based conjugate. *Cancer Research* 64, 2853–2857.
30. De Genst E *et al.* (2004) Chemical basis for the affinity maturation of a camel single domain antibody. *The Journal of Biological Chemistry* 279, 53593–53601.
31. Elowitz MB, Leibler S (2000). A synthetic oscillatory network of transcriptional regulators. *Nature* 403, 335-338.
32. Endy, D (2005). Foundations for engineering biology. *Nature* 438, 449-453.
33. Feldhaus MJ, Siegel RW, Opresko LK, Coleman JR, Feldhaus JM (2003). Flow cytometric isolation of human antibodies from a nonimmune *Saccharomyces cerevisiae* surface display library. *Nature Biotechnology* 21, 163-170.
34. Gacy AM, Goellner G, Juranić N, Macura S, McMurray CT (1995) Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell* 81, 533-540.
35. Gebauer M, Skerra A (2009). Engineered protein scaffolds as next-generation antibody therapeutics. *Current Opinion in Chemical Biology* 13, 245-255.
36. Gibson DG, Glass JI, Lartigue C, Noskov VB, Chuang RY, *et al.* (2010). Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science* 329, 52-56.
37. Harmsen MM, De Haard HJ (2007). Properties, production, and applications of camelid single-domain antibody fragments. *Applied Microbiology and Biotechnology* 77, 13-22
38. Hartwell LH, Hopfield JJ, Leibler S, Murray AW (1999). From molecular to modular cell biology. *Nature* 402, C47-C52.

39. Heinemann M, Panke S (2006). Synthetic biology – putting engineering into biology. *Bioinformatics* 22, 2790-2799.
40. Hiom K, Melek M, Gellert M (1998). DNAtransposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell* 94,463–470.
41. Hirano N, Muroi T, Takahashi H, Haruki M (2011). Site-especific recombinases as tools for heterologous gene integration. *Applied Microbiology and Biotechnology*. 92, 227-239.
42. Holliger P, Hudson PJ (2005). Engineered antibody fragments and the rise of single domains. *Nature Biotechnology* 23, 1126-1136.
43. Hoogenboom HR, Chames P (2000). Natural and designer binding sites made by phage display technology. *Immunology Today* 21, 371-378.
44. Hosse RJ, Rothe A, Power BE(2006). A new generation of protein display scaffolds for molecular recognition. *Protein Science* 15, 14–27.
45. Huhlov A, Chester KA (2004). Engineered single chain antibody fragments for radioimmunotherapy. *The Quarterly Journal of Nuclear Medicine and Molecular Imaging* 48, 279–288.
46. Jeong KJ, Jang SH, Velmurugan N (2011). Recombinant antibodies: Engineering and production in yeast and bacterial hosts. *Biotechnology Journal* 6, 16-27.
47. Knappik A, Ge L, Honegger A, Pack P, Fischer M *et al.* (2000). Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. *Journal of Molecular Biology* 296, 57-86.
48. Kohler G, Milstein C (1975.) Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* 256, 495–497.
49. Lendel C, Dincbas-Renqvist V, Flores A, Wahlberg E, Dogan J, *et al.* (2004). Biophysical characterization of ZSPA-1—A phage-display selected binder to protein A. *Protein Science* 13, 2078-2088.
50. Litman GW, Rast JP, Fugmann SD (2010). The origins of vertebrate adaptive immunity. *Nature Reviews Immunology* 10, 543-553.
51. Lonberg N (2008). Fully human antibodies from transgenic mice and phage display platforms. *Current Opinion in Immunology* 20, 450-459
52. Mayfield SP, Franklin SE, Lerner RA (2003) Expression and assembly of a fully active antibody in algae. *Proceedings of the National Academy of Sciences* 100, 438-442

53. Merz T, Wetzel SK, Firbank S, Plückthun A, Grütter MG, *et al.* (2008) Stabilizing ionic interactions in a full-consensus ankyrin repeat protein. *Journal of Molecular Biology* 376, 232-40.
54. Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, *et al.* (2000). Class switch recombination and hypermutation require Activation-Induced cytidine Deaminase (AID), a potential RNA editing enzyme. *Cell* 102, 553-563.
55. Nagy A (2000). Cre recombinase: The universal reagent for genome tailoring. *Genesis* 26, 99-109.
56. Nord K, Gunneriusson E, Ringdahl J, Stahl S, Uhlen M, *et al.* (1997). Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain. *Nature Biotechnology* 15, 772–777.
57. Nord K, Gunneriusson E, Uhlen M, Nygren PA (2000). Ligands selected from combinatorial libraries of protein A for use in affinity capture of apolipoprotein A-1M and taq DNA polymerase. *Journal of Biotechnology* 80, 45–54.
58. Nygren PA (2008). Alternative binding proteins: Affibody binding proteins developed from a small three-helix bundle scaffold. *FEBS Journal* 275, 2668-2676.
59. Oettinger MA, Schatz DG, Gorka C, Baltimore D (1990). RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* 248, 1517–1523.
60. Pancer Z, Amemiya CT, Ehrhardt GRA, Ceitlin J, Gartland GL, *et al.* (2004). Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature* 430, 174–180.
61. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, *et al.* (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* 25, 1605-1612.
62. Petersen-Mahrt SK, Harris RS, Neuberger MS (2002). AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature* 418, 99-104.
63. Rice PA, Mouw KW, Montano SP, Boocock MR, Rowland SJ, *et al.* (2010). Orchestrating serine resolvases. *Biochemical Society Transactions* 38, 384–387.
64. Rothe A, Hosse RJ, Power BE (2006). In vitro display technologies reveal novel biofarmaceutics. *FASEB Journal* 20, 1599-1610.
65. Roy A, Kucukural A, Zhang Y (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, vol 5, 725-738

66. Ruigrok VJ, Levisson M, Eppink MH, Smidt H, van der Oost J (2011). Alternative affinity tools: more attractive than antibodies? *Biochemical Journal* 436, 1-13.
67. Saerens D, Ghassabeh GH, Muyldermans S (2008). Single-domain antibodies as building blocks for novel therapeutics. *Current Opinion in Pharmacology* 8, 600-608
68. Sandstrom K, Xu Z, Forsberg G, Nygren PA (2003). Inhibition of the CD28-CD80 co-stimulation signal by a CD28-binding affibody ligand developed by combinatorial protein engineering. *Protein Engineering* 16, 691–697.
69. Scolink PA (2009). mAbs - a business perspective. *mAbs* 1, 179-184.
70. Schatz DG, Oettinger MA, Baltimore D (1989). The V(D)J recombination activating gene, RAG-1. *Cell* 59, 1035–1048.
71. Shaner NC, Steinbach PA, Tsien RY (2005). A guide to choosing fluorescent proteins. *Nature Methods* 2, 905-909
72. Sheridan C (2007). Pharma consolidates its grip on post-antibody landscape. *Nature Biotechnology* 25, 365-366.
73. Siegel RW, Jain R, Bradbury A (2001). Using an in vivo phagemid system to identify non-compatible loxP sequences. *FEBS Letters* 505, 467-473.
74. Skerra A, Pluckthun A (1988). Assembly of a functional immunoglobulin Fv fragment in *Escherichia coli*. *Science* 240, 1038-1041.
75. Skerra A (2000). Engineered protein scaffolds for molecular recognition. *Journal of Molecular recognition* 13, 167-187.
76. Skerra A (2003). Imitating the humoral immune response. *Current Opinion in Chemical Biology* 7, 683-693.
77. Skerra A (2007). Alternative non-antibody scaffolds for molecular recognition. *Current Opinion in Biotechnology* 18, 295–304.
78. Skerra A (2008). Alternative binding proteins: Anticalins – harnessing the structural plasticity of the lipocalin ligand pocket to engineer novel binding activities. *FEBS Journal* 275, 2677-2683.
79. Smith SG, Mahon V, Lambert MA, Fagan RP (2007). A molecular Swiss army knife: OmpA structure, function and expression. *FEMS Microbiology Letters* 273, 1-11.
80. Stanfield RL, Dooley H, Flajnik MF, Wilson IA (2004). Crystal structure of a shark single-domain antibody V region in complex with lysozyme. *Science* 305, 1770–1773.

81. Steiner D, Forrer P, Plückthun A (2008). Efficient selection of DARPins with sub-nanomolar affinities using SRP phage display. *Journal of Molecular Biology* 382, 1211-1227.
82. Steinmeyer DE, McCormick EL (2008). The art of antibody process development. *Drug Discovery Today* 13, 613-618.
83. Sternberg N, Hamilton D (1981). Bacteriophage P1 site-specific recombination : I. Recombination between loxP sites. *Journal of Molecular Biology* 150, 467-486.
84. Strachan T, Read AP (1999). *Human Molecular Genetics*. 2nd edition. New York: Wiley-Liss.
85. Streltsov VA *et al.* (2004). Structural evidence for evolution of shark Ig new antigen receptor variable domain antibodies from a cell-surface receptor. *Proceedings of the National Academy of Sciences* 101, 12444–12449.
86. Stumpff MT, Binz HK, Amstutz P (2008). DARPins: a new generation of protein therapeutics. *Drug Discovery Today* 13, 695-701.
87. Sundaresan G *et al.* (2003). 124I-labeled engineered anti-CEA minibodies and diabodies allow high-contrast, antigen-specific small-animal PET imaging of xenografts in athymic mice. *The Journal of Nuclear Medicine* 44, 1962–1969.
88. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 25, 4876-4882.
89. Tran M, Zhou B, Petterson PL, Gonzalez MJ, Mayfield SP (2009) Synthesis and assembly of a full-length human monoclonal antibody in algal chloroplasts. *Biotechnology and Bioengineering* 104, 663-673.
90. Voineagu I, Narayanan V, Lobach KS, Mirkin SM (2008). Replication stalling at unstable inverted repeats: Interplay between DNA hairpins and fork stabilizing proteins. *PNAS* 105, 9936-9941.
91. Ward ES, Gussow D, Griffiths AD, Jones PT, Winter G (1989). Binding activities of a repertoire of single immunoglobulin variable domains secreted from *Escherichia coli*. *Nature* 341, 544–546
92. Weissman TA, Sanes JR, Lichtman JW, Livet J (2011). Generating and Imaging Multicolor Brainbow Mice. *Cold Spring Harbor Protocols* doi:10.1101/pdb.top114

93. Wolf DM, Arkin AP (2003). Motifs, modules and games in bacteria. *Current Opinion in Microbiology* 6, 125-134.
94. You L, Cox RS 3rd, Weiss R, Arnold FH (2004). Programmed population control by cell-cell communication and regulated killing. *Nature* 428, 868–871.

**Sitios de internet consultados (revisados el 17 de junio de 2012)**

1. [http://partsregistry.org/Main\\_Page](http://partsregistry.org/Main_Page)
2. <http://2010.igem.org>
3. <http://2010.igem.org/Team:Slovenia>
4. <http://www.syntheticbiology.org>
5. (<http://web.mit.edu/jagoler/www/biojade/>)
6. <http://synbiooss.sourceforge.net/>
7. <http://www.genocad.org>
8. <http://www.clothocad.org/>
9. [http://2008.igem.org/Wiki/Team:Warsaw/igem\\_project.htm](http://2008.igem.org/Wiki/Team:Warsaw/igem_project.htm)
10. <http://csb.wfu.edu/brf/ABlguide>

**Patentes consultadas**

1. Patente Internacional WO 02/20565 A2: Collections of repeat proteins comprising repeat modules

...nada, realmente nada, pero sucede que nada más nada no da nada sino que a veces da un poquito de algo, lo que se dice una nada que como todos saben ya no quiere decir una pura nada sino un cachito de cualquier cosa, y fijate que si la operación está bien hecha no hay cana ni juez que pueda meter leña...

*Libro de Manuel, Julio Cortázar*



**Anexo 1: Secuencia de aminoácidos de las DARPinas usadas para el diseño de los  
módulos constantes y variables**


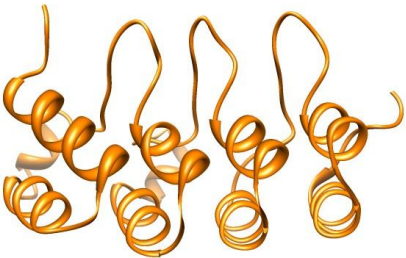


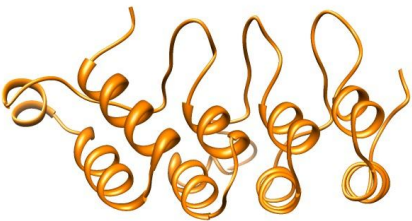
PDB ID	Secuencia de aminoácidos
2QYJ	MRGSHHHHHHGS DLGKKLLEAARAGODDEVRILMANGADVNAKDKDGYTPLHLAARE GHLEIVEVLLKAGADVNAKDKDGYTPLHLAAREGHLEIVEVLLKAGADVNAKDKDGYTPL HLAAREGHLEIVEVLLKAGADVNAQDKFGKTAFDISIDNGNEDLAEILQ
2V5Q:D	MRGSHHHHHHGS DLGKKLLEAARAGODDEVRILIANGADVNAVNTGLTPLHLAAVSG HLEIVEVLLKHGADVDAADVYGFTPLHLAAMTGHLEIVEVLLKYGADVNAFDMTGSTPLH LAADDEGHLEIVEVLLKYGADVNAQDKFGKTAFDISIDNGNEDLAKSCRN
2P2C:R	MRGSHHHHHHGS DLGKKLLEAARAGODDEVRILMANGADVNAATDWLGHTPLHLAAKT GHLEIVEVLLKYGADVNAWDNYGATPLHLAADNGHLEIVEVLLKHGADVNAKDYEGFTPL HLAAYDGHLEIVEVLLKYGADVNAQDKFGKTAFDISIDNGNEDLAEILQKLN
1SVX:A	MRGSHHHHHHGS DLGRKLLLEAARAGODDEVRILMANGADVNAADNTGTTPLHLAAYSG HLEIVEVLLKHGADVDAASDVFGYTPLHLAAYWGHLEIVEVLLKNGADVNA MDS DGM TPLH LAAKWGYLEIVEVLLKHGADVNAQDKFGKTAFDISIDNGNEDLAEILQKLN

## Anexo 2: Secuencias nucleotídicas utilizadas del “Registry of Standard Biological Parts”

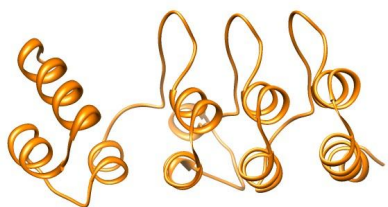

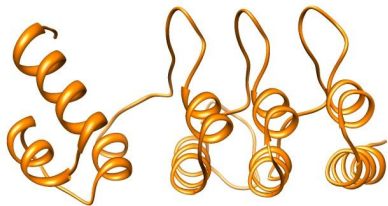

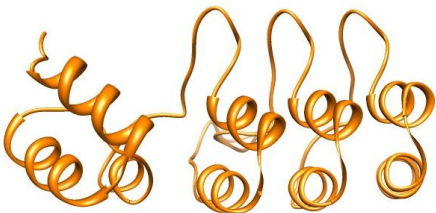
Nombre y descripción	Identificador	Secuencia
<b>pLac.</b> Promotor regulado por LacI, inducible con IPTG.	BBa_R0010	CAATACGCAAACCGCCTCTCCCCGCGGTTGGCCGATTCATTAATGCA GCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAA CGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCAGGCTTTACA CTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACA ATTCACACA
<b>OmpA.</b> Fragmento del péptido señal de lipoproteína con los AAs 46-159 de la proteína OmpA y un linker no estructurado (Glicina-serina-glicina).	BBa_K103006	CATATGAAAGCTACTAAACTGGTACTGGGCGCGGTAATCCTGGGTTCT ACTCTGCTGGCAGGTTGCTCCAGCAACGCTAAAATCGATCAGGGAATT AACCCGTATGTTGGCTTTGAAATGGGTTACGACTGGTTAGGTCGTATG CCGTACAAAGGCAGCGTTGAAAACGGTGCATACAAAGCTCAGGGCGT TCAACTGACCGCTAAACTGGGTTACCCAATCACTGACGACCTGGACAT CTACACTCGTCTGGGTGGCATGGTATGGCGTGCAGACACTAAATCCA ACGTTTATGGTAAAACCACGACACCGGGCTTTCTCCGGTCTTCGCTG GCGGTGTTGAGTACGCGATCACTCCTGAAATCGCTACCCGTCTGGAAT ACCAGTGGACCAACAACATCGGTGACGCACACACCATCGGCACTCGT CCGGACAACGGCGGAGGTTCTGGAGGAGGGAGCTC
<b>Terminador.</b> Consiste en dos terminadores que tienen una eficiencia de terminación de la transcripción 98%.	BBa_B0015	CCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTT CGTTTTATCTGTTGTTTGTGCGGTGAACGCTCTCTACTAGAGTCACACTG GCTCACCTTCGGGTGGGCCTTTCTGCGTTTATA
<b>RBS.</b> Sitio de unión a ribosoma. Proporciona un nivel constitutivo de traducción.	BBa_B0030	ATTAAGAGGAGAAA

**Anexo 3: Modelos estructurales para las proteínas finales con diferentes combinaciones de secuencias loxP heterólogas generados por I-TASSER**

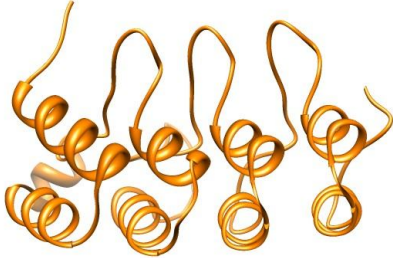


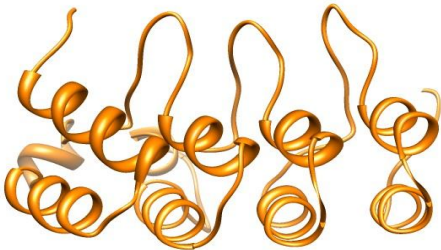
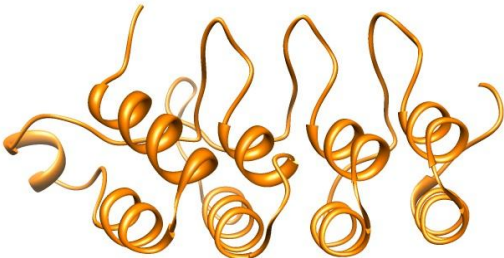
**Modelos para WT/FAS generados por I-TASSER**

# Modelo	Modelo	Valor C*
1		0.45
2		0.17
3		-1.95
4		-3.85
5		-4.37

Modelos para WT/2272 generados por I-TASSER

# Modelo	Modelo	Valor C*
1		0.41
2		-1.81
3		-0.68
4		-0.69
5		-2.31

Modelos para FAS/2272 generados por I-TASSER

# Modelo	Modelo	Valor C*
1		0.45
2		-2.87
3		-1.05
4		-0.23
5		-3.47

\*Valor C: Medida de la confiabilidad de los modelos predichos por I-TASSER. Se calcula con base en la significancia de los alineamientos de la secuencia realizados y en los parámetros de convergencia de las simulaciones de ensamblaje.

#### Anexo 4: Alineamiento de la secuencia obtenida con la secuencia V1

```

Seq1      -----CGTTCTCGAGTTGTTGGCGG
V1        GAATTCGCGGCCGCTTCTAGACAACCTTCGTATATACCTTTCTATACGAAGTTGTTGGCGG
                * :*. .*****

Seq1      ATGTGAACGCGAAAGATAAAGATGGCTATACCCCGCTGCATCTGGCGGCGTATTGGGGCC
V1        ATGTGAACGCGAAAGATAAAGATGGCTATACCCCGCTGCATCTGGCGGCGTATTGGGGCC
                *****

Seq1      ATCTGGAAATTGTGGAAGTGCTGCTGAAAAGCGGGCATAACTTCGTATAGGATACCTTATA
V1        ATCTGGAAATTGTGGAAGTGCTGCTGAAAAGCGGGCATAACTTCGTATAGGATACCTTATA
                *****

Seq1      CGAAGTTATTAACAACCTTCGTATATACCTTTCTATACGAAGTTGTTGGCGGATGTGAACG
V1        CGAAGTTATTAACAACCTTCGTATATACCTTTCTATACGAAGTTGTTGGCGGATGTGAACG
                *****

Seq1      CGAAAGATAAAGATGGCGCGACCCCGCTGCATCTGGCGGCGGATAACGGCCATCTGGAAA
V1        CGAAAGATAAAGATGGCGCGACCCCGCTGCATCTGGCGGCGGATAACGGCCATCTGGAAA
                *****

Seq1      TTGTGGAAGTGCTGCTGAAAGCGGGCATAACTTCGTATAGGATACCTTATACGAAGTTAT
V1        TTGTGGAAGTGCTGCTGAAAGCGGGCATAACTTCGTATAGGATACCTTATACGAAGTTAT
                *****

Seq1      TAACAACCTTCGTATATACCTTTCTATACGAAGTTGTTGGCGGATGTGAACGCGAAAGATA
V1        TAACAACCTTCGTATATACCTTTCTATACGAAGTTGTTGGCGGATGTGAACGCGAAAGATA
                *****

Seq1      AAGATGGCTTTACCCCGCTGCATCTGGCGGCGATGACCGGCCATCTGGAAATTGTGGAAG
V1        AAGATGGCTTTACCCCGCTGCATCTGGCGGCGATGACCGGCCATCTGGAAATTGTGGAAG
                *****

Seq1      TGCTGCTGAAAGCGGGCATAACTTCGTATAGGATACCTTATACGAAGTTATTATACTAGT
V1        TGCTGCTGAAAGCGGGCATAACTTCGTATAGGATACCTTATACGAAGTTATTATACTAGT
                *****

Seq1      AGCGGCCGCTGCAG
V1        AGCGGCCGCTGCAG
                *****

```

**Anexo 5: Artículo de revisión publicado en la revista BioTecnología**

## Biología Sintética: Diseñando Sistemas Biológicos con Piezas Genéticas

Daniel Aguilar-Salvador, Isabel Angeles-Santander, Mauricio A. Trujillo-Roldán,  
Norma A. Valdez-Cruz\*

*Departamento de Biología Molecular y Biotecnología, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, A.P. 70228, México D.F., C.P. 04510, México \* Correspondencia: adivaldez1@gmail.com*

### RESUMEN

La biología sintética es una nueva área de la biología y la tecnología que fusiona la biología molecular, la ingeniería genética y herramientas computacionales, para crear sistemas biológicos con funciones novedosas. Los sistemas creados sintéticamente son ya una realidad, y cada vez se acumulan más trabajos alrededor del mundo que muestran su factibilidad. En este campo no sólo se hacen pequeñas modificaciones en la información genética, sino que también se diseñan, manipulan, simulan e introducen circuitos genéticos a los organismos. Con este nuevo enfoque científico, se están abordando distintos problemas tecnológicos, como nuevas formas de síntesis y producción de biocombustibles, biofármacos, nanoestructuras, entre otras, permitiendo la innovación y desarrollos en la generación de energía, biomedicina y biología celular, así como en las ciencias de materiales. En este artículo se presenta una breve reseña de interesantes logros de la biología sintética. Entre ellos, la manipulación de *E. coli* para producir proteínas estables mediante recombinación homóloga simulando la biosíntesis de anticuerpos variables, basadas estructuralmente en las moléculas DARPinas derivadas de ankirinas, las cuales presentarán afinidad por diferentes sustratos.

**Palabras clave:** Biología sintética, Ingeniería genética, Biopartes, Anticuerpos, DARPinas.

### ABSTRACT

Synthetic biology is a new area of biology and technology that fuses molecular biology, genetic engineering and computational tools, to create biological systems with novel functions. Systems synthetically created are now a reality, and the accumulation of more works around the world, shows their feasibility. In this field, not only minor modifications to genetic information in organisms are made, genetic circuits are designed, manipulated, simulated, built and introduced to whole organisms. With this new scientific approach, various technological problems are being addressed, such as new forms of synthesis and production of biofuels, biopharmaceuticals, nanostructures, among others, allowing innovation and developments in power generation, biomedical and cell biology as well as in the materials science. This article provides a brief overview of interesting



achievements of synthetic biology. Among them, the manipulation of *E. coli* to produce stable proteins by homologous recombination (mimicking the biosynthesis of antibodies variability), based on the structure of DARPinas derived from ankyrins, which will have affinity for different substrates.

**Keywords:** Synthetic biology, Genetic engineering, Biobricks, Antibodies, DARPins.

## LA BIOLOGÍA SINTÉTICA

La promesa de la biología sintética es diseñar y construir diferentes sistemas biológicos de forma racional conjuntando piezas de información genética, lo que permitirá la modificación de la vida existente. Los sistemas que se diseñan en biología sintética comparten el hecho de no encontrarse en la naturaleza y de que algunos de ellos ni siquiera podrían sobrevivir fuera del laboratorio. Probablemente, en un futuro cercano, un científico de cualquier parte del mundo podrá sentarse ante la pantalla de su computadora y diseñar seres vivos en cuestión de días, o incluso de horas. Ya sea que desee fabricar fármacos complejos o combustible para aviones, podrá descargar de una base de datos la información genética necesaria para hacer un organismo “a la carte”. Como toda promesa, puede no cumplirse, aunque no parece tan lejana, pues ya se han construido circuitos genéticos que producen fármacos complejos y biocombustibles. Por ejemplo, el Doctor Jay Keasling de la Universidad de California, en 2003 diseñó e introdujo un circuito genético para producir en la bacteria *Escherichia coli* un precursor químico de la artemisinina (Martin *et al.*, 2003). Este es un compuesto que se usa como tratamiento contra la malaria y que de forma

convencional se extrae de la planta *Artemisia annua*. La *E. coli* fue modificada genéticamente introduciéndole información genética codificante para enzimas involucradas en la vía metabólica del mevalonato de la levadura *S. cerevisiae* y el gen codificante para la enzima amorfadieno sintasa de *Artemisia annua* (Martin *et al.*, 2003). De esta manera, se logró producir amorfa-4,11-dieno (un compuesto a partir del cual se puede sintetizar la artemisinina) en *E. coli*. Posteriormente, en 2006 el mismo grupo reportó la producción de ácido artemisinico en *S. cerevisiae*, otro precursor de la artemisinina, cuya molécula necesita menos pasos de modificación química comparado con el amorfa-4,11-dieno, para obtener artemisinina (Ro *et al.*, 2006). Esto se logró mediante la modificación de la regulación de la vía del mevalonato y la introducción de dos genes de *Artemisia annua* en la levadura. Por otro lado, la empresa LS9, con base en San Francisco, California reportó la modificación de *E. coli* mediante la incorporación de un circuito genético para producir alcanos y alquenos, principales constituyentes de los combustibles fósiles (gasolina, diesel y turbosina), mediante la inserción de la ruta biosintética de la cianobacteria *Synechococcus elongatus* (Schirmer *et al.*,

2010). La ruta biosintética incluye dos enzimas, una reductasa acarreadora de grupos acilo y una aldehído decarbonilasa, las cuales convierten los intermediarios del metabolismo de ácidos grasos en alcanos y alquenos. El hallazgo de dicha ruta, así como la posibilidad de ser transferida entre organismos, permitirá la conversión biológica de carbohidratos o azúcares en combustibles de bajo costo (Schirmer *et al.*, 2010). En el 2011, todos los genes involucrados en dicha ruta biosintética fueron construidos en formato de bioparte para su posterior uso en otros circuitos genéticos (<http://2011.igem.org/Team:Washington>).

La Biología Sintética, también ha generado alternativas para la biomedicina, mediante el diseño y construcción de diferentes microorganismos para atacar distintas infecciones virales y microbianas. Por ejemplo, para enfrentar el problema de las bacterias patógenas resistentes a antibióticos y que forman biopelículas, se han modificado virus que infectan a bacterias, para que eliminen a distintas poblaciones bacterianas (Lu y Collins 2007, 2009). Por ejemplo, fago lítico T7 fue modificado genéticamente para producir la enzima dispersina B, que degrada la matriz extracelular que mantiene agragadas a las células, mientras el fago infecta y lisa las células (Lu y Collins 2007). En otro estudio, se modificó genéticamente a la bacteria *E. coli* Nissle 1917, para que sintetizara el autoinductor CAI-1, el cual promueve la inhibición de los genes de virulencia de *Vibrio*

*cholerae*. La previa inoculación de la bacteria modificada en ratones disminuyó la colonización de *V. cholerae* y aumentó 85% la sobrevivencia de los ratones infectados (Duan y March 2010).

Algunas de las ideas más ambiciosas buscan crear construcciones genéticas que puedan detectar enfermedades y restauren las funciones saludables de las propias células del cuerpo. Un ejemplo de esto es el trabajo de Kemmer y colaboradores (2010), quienes diseñaron y probaron en ratones con expresión deficiente de la ureato-oxidasa (enzima que degrada el ácido úrico), un circuito genético codificante para producir una proteína que “sensa” el ácido úrico y, cuando hay un exceso, se produce la enzima ureato-oxidasa. Sin embargo, estos trabajos con aplicaciones directas en mamíferos aún son escasos, debido a que la mayoría de las herramientas para hacer biología sintética se han probado en bacterias y levaduras. Por lo que, se requiere que la biología sintética tenga un mayor desarrollo en células de mamífero, tejidos y en organismos (Constante *et al.*, 2011).

## HERRAMIENTAS DE LA BIOLOGÍA SINTÉTICA PARA ENSAMBLAR VIDA

La biología sintética además de la biología molecular requiere de la información genética y las diferentes técnicas de ensamblaje sistemático y estandarizado de distintas secuencias de ADN, para generar información genética nueva. Por esta razón, una contribución importante para el progreso

de la biología sintética fue el desarrollo de una base de datos en Internet llamada “Registry of Standard Biological Parts” (<http://partsregistry.org>, 2012), con información genética que a la fecha, incluye más de cinco mil biopartes estandarizadas o “BioBricks”. Cada uno de estas biopartes consta de una secuencia de ADN con características particulares que permiten ensamblarlas sistemáticamente. Casi todas las biopartes que se encuentran en esta base de datos fueron desarrolladas por los equipos de alumnos e investigadores del mundo que participan en el concurso “International Genetically Engineered Machine Competition” (iGEM), en el cual se plantean y desarrollan proyectos de biología sintética (<http://www.igem.org>, 2012). Las aplicaciones de los proyectos que han ganado este concurso incluyen propuestas en campos tan diversos como la biotecnología, la biomedicina, la nanotecnología y las ciencias computacionales (Goodman, 2008).

Hoy en día se genera una extensa cantidad de datos gracias a las tecnologías como genómica, transcriptómica, proteómica y metabolómica que sirven de base para el desarrollo de nuevos circuitos genéticos (Nandagopal y Elowitz, 2011). Actualmente, se hace un gran esfuerzo para construir numerosos genes en formato de bioparte, caracterizar la interacción de las mismas con otras moléculas, su regulación, expresión y funcionalidad en diferentes organismos. La mayoría de los proyectos de biología sintética

que se han desarrollado hasta ahora, se basan en el conocimiento actual del funcionamiento de los organismos vivos. Sin embargo, la información es numerosa y creciente, por lo que se requiere de herramientas computacionales y matemáticas para integrar y hacer manejable esta gran cantidad de información, que sean una base clara para diseñar y construir sistemas biológicos artificiales.

Algunas de las herramientas computacionales ya existentes son “From Metabolite to Metabolite” (<http://fmm.mbc.nctu.edu.tw/>, 2012), con la que se pueden diseñar nuevas rutas metabólicas, o el “Standard Virtual Biological Parts”, que contiene una colección de modelos computacionales que ayudan a predecir el funcionamiento de un circuito genético, facilitando el diseño de circuitos genéticos antes de decidir cuál construir en el laboratorio (Cooling *et al.*, 2010). Por otro lado, aplicaciones como BioJADE, SynBioSS, GenoCAD y ClothoCAD buscan conjuntar el acceso a bases de datos de biopartes, el ensamblaje de nuevos circuitos genéticos y el modelaje matemático y computacional de los mismos (<http://web.mit.edu/jagoler/www/biojade/>, 2012; <http://synbio.ss.sourceforge.net/>, 2012; <http://www.genocad.org>, 2012; <http://www.clothocad.org/>, 2012). Aún con la ayuda de éste tipo de herramientas, un circuito optimizado puede no funcionar como se espera una vez que se construye y se introduce a un organismo, esto debido a la

complejidad inherente a todos los sistemas biológicos. Una manera de reducir dicha complejidad sería construir una “célula mínima”, que sólo tuviera la información genética necesaria para vivir y reproducirse. Por ejemplo, *Mycoplasma laboratorium* fue creada con éste propósito, siendo el organismo con el genoma más pequeño con sólo 382 genes (The J. Craig Venter Institute *et al.*, 2007).

Después del diseño de los circuitos, hay que construir una cadena de ADN con la información genética necesaria. Actualmente, la síntesis comercial de ADN la realizan compañías como Blue Heron, Geneart, DNA 2.0 y GenScript, aunque a un precio elevado en comparación con las técnicas estándar de biología molecular cuando se quieren ensamblar cadenas largas de ADN (amplificación, restricción, ligación, clonación, etc.). Para facilitar el ensamblaje se han propuesto métodos más eficientes, como el “Golden-Gate” (Engler *et al.*, 2008), el “ensamblaje Gibson” (Gibson *et al.*, 2009) y el “DNA assembler” (Shao *et al.*, 2009). Con esta última técnica, por ejemplo, se ha logrado el ensamblaje de 8 genes de una vía metabólica, formando una cadena de 19 mil nucleótidos de ADN en un solo paso; aunque el record de ensamblaje lo tiene el genoma sintético de *Mycoplasma mycoides*, de más de un millón de pares de bases, el cual fue ensamblado mediante recombinación homóloga, y además fue insertado con éxito en una célula a la que previamente se le

había extraído su propio genoma (Gibson *et al.*, 2010).

Además de diseñar información genética nueva, algunos científicos buscan codificarla de una manera diferente. Ya se han creado ARNs de transferencia que pueden incorporar aminoácidos modificados (Liu *et al.*, 2007). También, se han probado sistemas transcripcionales *in vitro* que usan codones de cuatro bases en vez de tres, los cuales codifican para aminoácidos no naturales (Taira *et al.*, 2005). El principal potencial de estos trabajos sería la posibilidad de incorporar aminoácidos con propiedades químicas y estructurales diferentes a los naturales para construir proteínas con nuevas propiedades y funciones.

## LOS PROBLEMAS DE LA VIDA SINTÉTICA©

El Registro de Partes Biológicas Estandarizadas, al igual que las aplicaciones computacionales son gratuitas. Algunos investigadores opinan que las herramientas que permiten hacer biología sintética (biopartes, células mínimas, técnicas de laboratorio y programas computacionales) deberían estar disponibles para cualquiera que desee usarlas, y que sólo las aplicaciones específicas que se desarrollen deberían patentarse (Erickson *et al.*, 2011). Uno de los principales problemas con limitar el uso y distribución de las biopartes, y de cualquier herramienta que facilite la tarea de los biólogos sintéticos, es que cada grupo

tendría que desarrollar sus propias biopartes y herramientas, lo que se reflejaría en pérdida de tiempo y recursos (Peccoud *et al.*, 2011). Un ejemplo de ello es lo que sucede con varios de los proyectos que se publican en revistas científicas, donde sólo se presentan los datos finales y no las secuencias de ADN o estrategias que se utilizaron (Peccoud *et al.*, 2011). Por otro lado, la convergencia de la biología sintética con otros campos de la ciencia, en especial con las tecnologías de la información y la nanotecnología, dificulta la toma de decisiones sobre qué debería ser patentable (Parens *et al.*, 2009). Después de todo, a nivel molecular los sistemas biológicos pueden considerarse como máquinas moleculares, y la información genética aunque propia de los seres vivos, puede ser vista como otro tipo de información (Parens *et al.*, 2009).

Otro tema en discusión son los retos de bioseguridad que plantea el diseño de seres vivos (Schmidt, 2008; Schmidt *et al.*, 2008; Schmidt *et al.*, 2009). En todos los países se deben discutir y crear los marcos regulatorios adecuados para que estos nuevos organismos se construyan con fines benéficos, tomando todas las medidas precautorias (Schmidt *et al.*, 2008). Es por esto que tanto la comunidad científica como el público en general, necesitan entender el potencial y los límites de la biología sintética para aprovechar al máximo esta nueva disciplina. Con las aplicaciones que se desarrollen se mejorarán las expectativas de

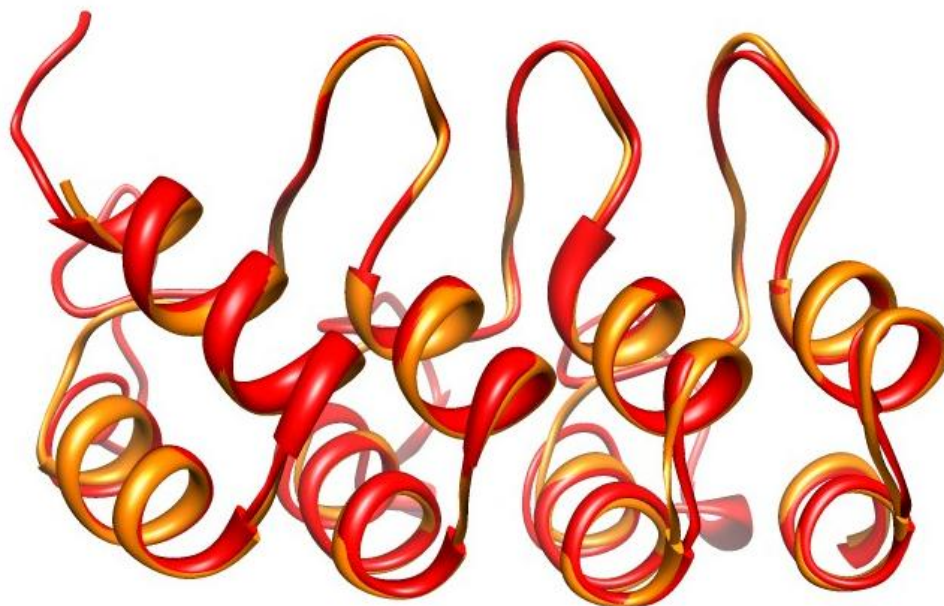
vida, la sobrevivida ante diferentes enfermedades y padecimientos. Y no sólo eso, también se encontrarán nuevas maneras de producir energías renovables y alimentos para nuestro mundo en crecimiento, así como nuevos materiales que impactarán en la vida que conocemos.

## **EL DISEÑO DE UN SISTEMA BIOLÓGICO SIMULADOR DE RECOMBINACIÓN HOMÓLOGA**

Simular en bacterias (como en *E. coli*) la recombinación somática que sucede durante la generación de anticuerpos en la maduración de linfocitos es un ejemplo claro de lo que se puede lograr con Biología Sintética. Es así como en nuestro grupo de investigación se ha logrado el ensamblaje de las secuencias nucleotídicas que tienen la información codificante para generar mediante recombinación proteínas basadas en la estructura de DARPinas. Las DARPinas son proteínas muy estables con estructura alfa enlazadas por “loops” de forma repetida (Stumpff *et al.*, 2008). Las DARPinas se derivan de repeticiones de ankirinas, que son una familia de proteínas de unión que se encuentran en mamíferos y particularmente en humanos (Lander *et al.*, 2001). Estas proteínas tienen por función biológica la unión a distintos blancos, debido a que poseen una superficie que interactúa con diferentes moléculas, además de que la repetición de los módulos que las conforman las hacen versátiles, variadas y diferentes en tamaño, sin que se altere su estructura

básica (Forrer *et al.*, 2004). A partir de las repeticiones de ankirinas, se han diseñado moléculas alternativas de reconocimiento por antígenos, entre ellas librerías de módulos de DARPinas (Forrer *et al.*, 2004). De los diferentes módulos obtenidos, se han construido proteínas con alta estabilidad, solubilidad y semejanza con la proteína de unión GA humana (Uniprot ID Q06547) (Stumpp *et al.*, 2008). Además, las DARPinas obtenidas son moléculas pequeñas, de aproximadamente una décima parte de un anticuerpo IgG convencional. Debido a sus propiedades, las DARPinas pueden ser usadas como andamiaje para el diseño de proteínas de reconocimiento de diferentes blancos, para su uso en biomedicina e investigación básica. Lo

anterior sustenta que en nuestro grupo estamos construyendo una secuencia sintética de ADN codificante para DARPinas, las cuales presentarán dominios conservados y distintos fragmentos variables que se introducirán mediante recombinación usando el sistema CRE/lox. El sistema sintético será incorporado en la Bacteria *E. coli*, para que produzca hasta 18 DARPinas diferentes. Usando los elementos estructurales de las DARPinas y los posibles cambios en diferentes aminoácidos que introducirá la recombinación, se realizó una predicción estructural *in silico* de las posibles 18 combinaciones esperadas (Fig. 1). Los modelos mostraron que no se alterará la estructura básica que le confiere estabilidad a las DARPinas (Fig. 1). De ser exitoso,



**Fig. 1.** Comparación estructural entre la DARPina consenso con PDB ID 2QYJ (anaranjado), y la predicción estructural de una de las proteínas producto del sistema de recombinación propuesto por nuestro grupo (rojo), usando el software I-TASSER (Ambrish *et al.*, 2010).

este mecanismo de recombinación en la bacteria, podría convertirse en el primer paso para desarrollar un método alternativo de producción de moléculas variables análogas a los anticuerpos monoclonales, con las ventajas estructurales de las DARPinas. Las DARPinas resultantes posteriormente deben ser caracterizadas y enfrentadas con distintos blancos para encontrar sus antígenos. Hasta ahora, este trabajo es de Ciencia Básica, y una vez que el sistema sintético sea funcional se pueden introducir de forma racional elementos de reconocimiento específico al andamiaje diseñado en este trabajo.

## CONCLUSIONES

La Biología sintética aún está en sus inicios, sin embargo diversos trabajos muestran un claro desarrollo en la generación de nuevas terapias médicas, la obtención de biocombustibles y la producción de materiales novedosos. La identificación de conexiones entre las vías metabólicas, regulatorias, y de respuesta de las células, permitirá la construcción de circuitos genéticos, cada vez más complejos. De ahí que el desarrollo de nuevas herramientas computacionales sea necesario para identificar nuevos componentes que predigan el comportamiento de los sistemas sintéticos diseñados.

Los ejemplos presentados en este escrito dan una visión del potencial de la biología sintética. Particularmente, el modelo bacteriano diseñado en nuestro grupo del

proceso de recombinación somática, propia de células de nuestro sistema inmune, podría servir como un sistema alternativo para la producción de moléculas análogas a los anticuerpos.

Aunque los desarrollos de la biología sintética tienen problemas de tipo técnico, práctico, legal y ético, que tendrán que ser solucionados, la biología sintética seguirá creciendo con la incorporación de más disciplinas al campo, y será una de las principales herramientas para entender, diseñar y crear nuevos sistemas biológicos.

## AGRADECIMIENTOS

Los autores expresan su agradecimiento al CONACyT Ciencia Básica 104951-Z

## REFERENCIAS

- Ambrish R, Alper K & Yang Z (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Prot.* 5: 725-738
- Constante, M, Grunberg R & Isalan M (2011) A biobrick library for cloning custom eukaryotic plasmids. *PLoS. One.* 6: 23685.
- Cooling, MT, Rouilly V, Misirli G, Lawson J, Yu T, *et al.* (2010) Standard virtual biological parts: a repository of modular modeling components for synthetic biology. *Bioinformatics.* 26: 925-931.
- Duan F & March JC (2010) Engineered bacterial communication prevents *Vibrio cholerae* virulence in an infant mouse

# Artículos

- model. *Proc. Natl. Acad. Sci. USA* 107:11260-11264.
- Erickson B, Singh R & Winters P (2011) Synthetic biology: regulating industry uses of new biotechnologies. *Science* 333: 1254-1256.
- Engler C, Kandzia R & Marillonnet S (2008) A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLoS ONE* 3: e3647.
- Forrer P, Binz HK, Stumpp MT & Plückthun A. (2004) Consensus design of repeat proteins. *Chembiochem* 5: 183–189.
- Gibson DG, Young L, Chuang RY, Venter JC & Hutchison CA 3rd, *et al.* (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6: 343-345.
- Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, *et al.* (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329: 52-56.
- Goodman C (2008). Engineering ingenuity at iGEM. *Nat. Chem. Biol.* 4: 13
- Kemmer C, Gitzinger M, Daoud-El Baba M, Djonov V, Stelling J *et al.* (2010) Self-sufficient control of urate homeostasis in mice by a synthetic circuit. *Nat. Biotechnol.* 28, 355–360.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860–921.
- Liu W, Brock A, Chen S, Chen S & Schultz PG (2007) Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat. Methods* 4:239-244.
- Lu TK & Collins JJ (2007). Dispersing biofilms with engineered enzymatic bacteriophage. *Proc. Natl. Acad. Sci. USA* 104: 11197-11202.
- Lu T K & Collins JJ (2009) Engineered bacteriophage targeting gene networks as adjuvants for antibiotic therapy. *Proc. Natl. Acad. Sci. USA* 106: 4629-4634.
- Martin VJJ, Pitera DJ, Withers ST, Newman JD & Keasling JD (2003) Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotech.* 21: 796-802.
- Nandagopal N & Elowitz MB (2011) Synthetic biology: integrated gene circuits. *Science* 333: 1244-1248.
- Parens E, Johnston J & Moses J (2009) Ethical Issues in Synthetic Biology: An Overview of the Debates, Woodrow Wilson International Center for Scholars, (disponible: [www.synbioproject.org/library/publications/archive/synbio3/](http://www.synbioproject.org/library/publications/archive/synbio3/)).
- Peccoud J, Anderson JC, Chandran D, Densmore D, Galdzicki M *et al.* (2011) Essential information for synthetic DNA sequences. *Nat. Biotechnol.* 29:22
- Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, *et al.* (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 440: 940-943.
- Shao Z, Zhao H & Zhao H (2009) DNA assembler, an in vivo genetic method for



# Artículos

- rapid construction of biochemical pathways. *Nucl. Acids. Res.* 37: e16.
- Schirmer A, Rude MA, Li X, Popova E & del Cardayre SB (2010) Microbial Biosynthesis of Alkanes. *Science* 329: 559-562.
- Schmidt M (2008) Diffusion of synthetic biology: a challenge to biosafety. *Syst. Synth. Biol.* 2: 1-6.
- Schmidt M, Torgersen H, Ganguli-Mitra A, Kelle A, Deplazes A & Biller-Andorno N. (2008) SYNBIOSAFE e-conference: online community discussion on the societal aspects of synthetic biology. *Syst. Synth. Biol.* 2: 7-17.
- Schmidt M, Ganguli-Mitra A, Torgersen H, Kelle A, Deplazes A *et al.* (2009) A priority paper for the societal and ethical aspects of synthetic biology. *Syst. Synth. Biol.* 3: 3-7.
- Stumpp MT, Binz HK & Amstutz P (2008) DARPin: a new generation of protein therapeutics. *Drug Discov. Today* 13: 695-701.
- Taira H, Fukushima M, Hohsaka T & Sisido M (2005) Fourbase codon-mediated incorporation of non-natural amino acids into proteins in a eukaryotic cell-free translation system. *J. Biosci. Bioeng.* 99: 473-476.
- The J. Craig Venter Institute, Glass J, Smith HO, Hutchinson CA, Alperovich NY & Assad-Garcia N (2007) Minimal Bacterial Genome. World patent WO2007047148.