



Escola Tècnica Superior
d'Enginyeria Agronòmica i del Medi Natural



UNIVERSITAT POLITÈCNICA DE VALÈNCIA
ETSIAMN
GRADO EN BIOTECNOLOGÍA



DISEÑO DE UN CIRCUITO GENÉTICO
DE ADAPTACIÓN Y ESTUDIO DE SUS
APLICACIONES EN LA INDUSTRIA
BIOTECNOLÓGICA

Curso 2019-2020

Autora: Raquel Andreu Vilarroig

Tutora: Nuria Andrés Colás

Cotutor: Alejandro Vignoni

Valencia, julio de 2020

Título: Diseño de un circuito genético de adaptación y estudio de sus aplicaciones en la industria biotecnológica.

Resumen: La Biología Sintética se define como la ingeniería de la biología, incluyendo el diseño y rediseño de componentes biológicos como partes de circuitos genéticos más complejos que al intercambiarse, pueden producir organismos nuevos que desempeñen funciones específicas. Para ello, sigue una metodología cíclica que consta de cuatro fases: Diseño, Construcción del ADN sintético, Comprobación Experimental y Aprendizaje.

En la naturaleza, las funciones celulares están controladas por la actividad coordinada de ciertos genes y proteínas que interactúan entre sí, activándose, reprimiéndose o catalizando otras reacciones biomoleculares. Muchos de estos circuitos genéticos tienen una característica conocida como *adaptación perfecta*, que es la habilidad de responder ante un estímulo y volver a su valor anterior a dicho estímulo. Esta característica resulta muy interesante desde el punto de vista de las ciencias biológicas, ya que el sistema puede comportarse como un detector de cambios celulares. Uno de los más estudiados es el circuito genético de retroalimentación incoherente de tipo 1 (I1-FFL).

En este trabajo, se diseñará un circuito I1-FFL de tres genes (*luxR*, *ci* y *gfp*) modelando matemáticamente sus interacciones biomoleculares. Luego, y se planteará la metodología para la construcción de los plásmidos que componen el circuito y se estimarán los valores de los parámetros del modelo a partir de datos experimentales para que genere comportamiento adaptativo. La caracterización del circuito genético I1-FFL resulta interesante porque, combinado con otros circuitos, puede ser de utilidad en diversas áreas de la Biotecnología, como la Biotecnología Blanca o Industrial.

Palabras clave: Biología Sintética, adaptación, modelado, selección de parámetros biológicos.

Alumna: Raquel Andreu Vilarroig

Tutora: Nuria Andrés Colás

Cotutor: Alejandro Vignoni

Valencia, julio de 2020

Title: Design of an adaptation genetic circuit and study of its applications in the biotechnology industry.

Abstract: Synthetic Biology is defined as the engineering of biology, including the design and redesign of biological components as parts of more complex genetic circuits which, when exchanged, can produce new organisms performing specific functions. To do this, it follows a cyclical methodology that consists of four phases: Design, Construction of the synthetic DNA, Test and Learn.

In nature, cellular functions are controlled by the coordinated activity of certain genes and proteins that interact with each other, activating, repressing or catalyzing other biomolecular reactions. Many of these genetic circuits have a characteristic known as perfect adaptation, which is the ability to respond to a stimulus and return to its value prior to that stimulus. This feature is very interesting from the point of view of the biological sciences, since the system can behave like a detector of cellular changes. One of the most studied is the type 1 Incoherent Feedforward Loop (I1-FFL).

In this work, an I1-FFL circuit of three genes (*luxR*, *ci* and *gfp*) will be designed by mathematically modelling their biomolecular interactions. Then, the methodology for the construction of the plasmids that form the circuit will be proposed and the values of the parameters of the model will be estimated from experimental data so that it generates adaptive behaviour. The characterization of the genetic circuit I1-FFL is interesting because, combined with other circuits, it can be useful in several areas of Biotechnology, such as White or Industrial Biotechnology.

Key words: Synthetic Biology, adaptation, modeling, selection of biological parameters.

Student: Raquel Andreu Vilarroig

Thesis director: Nuria Andrés Colás

Thesis co-director: Alejandro Vignoni

Valencia, July 2020

Títol: Disseny d'un circuit genètic d'adaptació i estudi de les seues aplicacions en l'indústria biotecnològica.

Resum: La Biologia sintètica es defineix com l'enginyeria de la biologia, incloent el disseny i el redisseny de components biològics com a parts de circuits genètics més complexos que quan a l'intercanviar-se, poden produir nous organismes que realitzen funcions específiques. Per aquest motiu, segueix una metodologia cíclica formada per quatre fases: Diseny, Construcció d'ADN sintètic, Comprovació Experimental i Aprenentatge.

A la natura, les funcions cel·lulars estan controlades per l'activitat coordinada de determinats gens i proteïnes que interactuen entre elles, activant, suprimint o catalitzant altres reaccions biomoleculares. Molts d'aquests circuits genètics tenen una característica coneguda com a *adaptació perfecta*, que és la capacitat de respondre a un estímul i tornar al seu valor previ a l'estímul. Aquesta característica és molt interessant des del punt de vista de les ciències de la vida, ja que el sistema pot comportar-se com un detector de canvis cel·lular. Un dels més estudiats és el circuit genètic de retroalimentació incoherent de tipus 1 (I1-FFL).

En aquest treball, es dissenyarà un circuit I1-FFL de tres gens (*luxr*, *cl* i *gfp*) modelant matemàticament les seves interaccions biomoleculares. A continuació, es proposarà la metodologia per a la construcció dels plasmidis que componen el circuit i s'estimaran els valors dels paràmetres del model a partir de dades experimentals per generar comportaments adaptatius. La caracterització del circuit genètic I1-FFL és interessant perquè, combinada amb altres circuits, pot ser útil en diverses àrees de la Biotecnologia, com la Biotecnologia Blanca o Industrial.

Paraules clau: Biologia Sintètica, adaptació, modelatge, selecció dels paràmetres biològics.

Alumna: Raquel Andreu Vilarroig

Tutora: Prof. Nuria Andrés Colás

Cotutor: D. Alejandro Vignoni

València, juliol de 2020

Agradecimientos

En primer lugar, quiero agradecer a Alejandro su gran ayuda y cercanía. También a Yadira, Nuria y Jesús, por su apoyo en todo momento. Debido a la pandemia, el reto que ya suponía el trabajo de por sí ha sido mayor. Gracias a vuestra paciencia, disponibilidad e interés en que aprendamos, lo hemos sacado adelante con éxito.

También quiero dar las gracias a Belén, Ana, Raquel, Ana Esther, Celia, Carmen, Iván y Javi, que me han acompañado desde el día uno durante estos cuatro años. Gracias por tantos momentos y por haberlo hecho todo ha sido mucho más fácil. Me llevo ocho grandes amigos, no podría haber encontrado otros mejores.

Por último, me gustaría agradecer a mis padres, hermanos y abuelos su enorme apoyo. Por alegraros de mis éxitos, animarme en los malos momentos, darme fuerzas, consejo, valores y ser un gran referente.

Este trabajo está dedicado a todos vosotros.

Índice

1. Introducción	1
1.1. La Biología Sintética	1
1.1.1. Diseño	2
1.1.2. Construcción	3
1.1.3. Comprobación experimental	4
1.1.4. Aprendizaje	5
1.2. El circuito I1-FFL - Comportamiento adaptativo	5
1.3. Aplicaciones	7
1.4. Limitaciones actuales	8
2. Objetivos	10
3. Materiales y métodos	11
3.1. Diseño	11
3.1.1. Modelado matemático	11
3.1.2. Simulación computacional	14
3.2. Construcción	15
3.2.1. Ensamblaje GoldenBraid	15
3.2.2. Transformación	18
3.2.3. PCR de colonias (cPCR)	18
3.2.4. Electroforesis en gel de agarosa	19
3.2.5. Cultivos celulares	19
3.2.6. Glicerinado bacteriano	20
3.2.7. Extracción de plásmidos: miniprep	20
3.3. Comprobación experimental	20
3.3.1. Medidas de fluorescencia y OD ₆₀₀ <i>in vivo</i>	21
3.3.2. Calibración	21
3.4. Aprendizaje	22
4. Resultados y Discusión	24
4.1. Diseño	24
4.1.1. Modelado matemático	24
4.1.2. Simulaciones del circuito I1-FFL	28
4.2. Construcción	31
4.3. Comprobación experimental	32
4.4. Aprendizaje	33
5. Conclusiones	35
6. Referencias bibliográficas	36
Anexo I: Artículos de revista y de congresos derivados de este trabajo	40

Listado de abreviaturas y acrónimos

ADN: Ácido desoxirribonucleico

AHL: N-acil L-homoserina lactona

ARN: Ácido Ribonucleico

ARNm: ARN mensajero

ARNP: Polimerasa de ARN

CDS: Secuencia codificante

C_N: Número de copias del plásmido

FT: Factor de transcripción

GFP: Proteína fluorescente verde

I1-FFL: Bucle de retroalimentación incoherente de tipo 1

LAM: Ley de acción de masas

MEFL: Moléculas equivalentes de fluoresceína

MOO: Optimización multi-objetivo

OD: Densidad óptica

ODE: Ecuaciones diferenciales ordinarias

QSSA: Aproximación quasi- estacionaria

RBS: Sitio de unión de los ribosomas

Índice de tablas

3.1. Parámetros y condiciones iniciales de la simulación.....	14
3.2. Partes básicas adaptadas a la gramática de GoldenBraid.....	15
3.3. Componentes de una reacción de GoldenBraid	16
3.4. Condiciones para llevar a cabo una reacción de GoldenBraid.....	17
3.5. Componentes para una cPCR.....	19
3.6. Condiciones para una cPCR.....	19
3.7. Condiciones para la toma de medidas de OD ₆₀₀ y fluorescencia de la GFP.....	21
4.1. Especies bioquímicas del circuito I1-FFL	25
4.2. Parámetros del modelo del circuito I1- FFL.....	27
4.3. Partes básicas que componen el circuito genético del que se han extraído los datos experimentales.....	32
4.4. Parámetros del modelo estimados a partir de los datos experimentales.....	33

Índice de figuras

1.1. Ciclo DBTL de la Biología Sintética.....	2
1.2. Representación gráfica y matemática de la adaptación de un circuito genético, su sensibilidad y precisión.....	6
1.3. Representación de una célula con el circuito I1-FFL incorporado.....	7
3.1. Esquema de la reacción de ensamblaje GoldenBraid de una unidad transcripcional utilizando las cuatro partes básicas y los vectores de destino de Nivel 1.....	16
3.2. Esquema de la construcción del circuito sintético I1-FFL, con sus respectivos niveles y plásmidos....	17
3.3. Curvas de calibración de absorbancia (izquierda) y fluorescencia (derecha).....	22
3.4. Esquema de la metodología seguida para la caracterización de las partes básicas	23
4.1. Simulaciones de las principales especies bioquímicas.....	29
4.2. Simulaciones del circuito genético implementado para distintas inducciones	30
4.3. Simulaciones de la concentración de GFP para distintos valores de la tasa de traducción de GFP.....	30
4.4. Esquema del circuito genético del que se han extraído los datos experimentales, implementado en una célula de E. coli.....	31
4.5. Calibración de los datos experimentales.....	32
4.6. Validación de la estimación de parámetros.....	34

1. Introducción

1.1. La Biología Sintética

La Biología Sintética se define como la ingeniería de la biología, incluyendo el diseño y rediseño de componentes biológicos como partes de circuitos genéticos más complejos que al intercambiarse, pueden producir organismos nuevos que desempeñen funciones específicas (ERASYNBIO, 2014) (Kojima y Fussenegger, 2019).

Para ello, es necesario el uso de tecnologías de Ingeniería Genética y herramientas de simulación *in silico*, por lo que se trata de un campo multidisciplinar que combina competencias de áreas como las Matemáticas, Computación y Bioquímica.

Como en cualquier Ingeniería, la Biología Sintética sigue una metodología cíclica, conocida como ciclo DBTL (Figura 1.1), que incluye el **diseño** de una solución a un problema (el prototipo inicial), la **construcción** física del mismo, la obtención de **datos experimentales** y, finalmente, la aplicación de herramientas de **aprendizaje** computacional para mejorar este prototipo. Se trata de una estrategia iterativa, ya que el ciclo se va repitiendo y modificando hasta conseguir el comportamiento deseado.

Los circuitos se diseñan o rediseñan a partir de los componentes presentes en otros ya existentes en la naturaleza con la finalidad de implementarlos en organismos vivos y que estos adquieran el comportamiento deseado.

El diseño comienza con el análisis de los componentes del circuito genético de interés y de sus interacciones, para obtener un modelo matemático que explique de forma precisa su comportamiento. Una vez realizado, se llevan a cabo simulaciones *in silico* para estudiar cómo se comportaría el circuito dentro de una célula si se modificasen unos parámetros u otros. Cuando el diseñador, en función de su objetivo, ha establecido los valores iniciales (provisionales) que considera más adecuados para estos parámetros, procede a la selección de los componentes con dichos valores para los parámetros y a la construcción del circuito en el laboratorio y su implementación en un microorganismo. Después, midiendo propiedades como la fluorescencia o la absorbancia, podrá obtener datos experimentales de este sistema y observar si el organismo realiza la función esperada. El gran número de factores que hay que tener en cuenta para diseñar un sistema biológico hacen de él un proceso complicado, por lo que es necesario realizar muchas iteraciones, optimizando los parámetros hasta conseguir aquellos valores que generen el comportamiento deseado. Finalmente, la simulación realizada con el modelo matemático y los datos experimentales se comparan para comprobar que, efectivamente, el modelo simula un comportamiento real.



Figura 1.1. Ciclo DBTL de la Biología Sintética, en el que se representan las cuatro fases fundamentales: Diseño (Design), Construcción (Build), Experimentación (Test) y Aprendizaje (Learn) (Inspirada en Boada Acosta, 2018).

Esta manera de trabajar se traduce en un ahorro de tiempo y dinero en el laboratorio pues, aunque los experimentos *in vivo* siempre están sujetos a condiciones cambiantes, el previo diseño permite predecir de la manera más exacta posible cómo se va a comportar el organismo una vez tenga incorporado el circuito. A continuación, se exponen de manera más detallada cada una de las fases.

1.1.1. Diseño

Muchas funciones celulares están controladas por la actividad coordinada de ciertos genes y proteínas que interactúan entre sí, activándose y reprimiéndose. Por este motivo, es necesario un diseño racional previo del circuito genético que se quiere construir, analizando bien dichas interacciones. Una parte importante del diseño es el **modelado matemático**, que consiste en plantear las ecuaciones que explican el comportamiento del sistema.

Para modelar un sistema biológico, normalmente se utilizan modelos cuantitativos cinéticos. Este tipo de modelos consisten en una serie de ecuaciones diferenciales que representan la evolución temporal de los distintos componentes o especies biológicas que intervienen en el sistema (Villaverde y Banga, 2014). Para ello, en primer lugar, deben definirse dichos componentes y las **reacciones bioquímicas** que explican el circuito. A continuación, aplicando la Ley de Acción de Masas, se obtienen los **balances dinámicos** para cada una de las especies. El resultado es un sistema de Ecuaciones Diferenciales Ordinarias (**ODEs**) de grandes dimensiones y que depende de muchos parámetros, por lo que será necesaria una reducción del modelo. Esta reducción permite obtener un modelo más manejable, pero hay que evitar que sea excesiva, pues llevaría a una falta de relevancia biológica (Boada Acosta, 2018).

Una vez obtenido el modelo final, éste se implementa en un software y, fijando unos valores iniciales para los parámetros del modelo, se ejecuta la **simulación**. Ésta sirve para analizar si, con esos valores, el modelo diseñado presenta el comportamiento adaptativo que se espera del sistema. También permite observar cómo influyen distintos valores de los parámetros en la concentración de las especies con respecto al tiempo.

1.1.2. Construcción

Los circuitos genéticos son sistemas en los que uno o más genes interactúan entre sí para llevar a cabo una función específica (Boada Acosta, 2018). Los circuitos se componen de unidades transcripcionales, que a su vez están formadas por cuatro partes biológicas estándar, también conocidas como **biopartes**:

- **Promotor**: zona del ADN a la que se une el enzima ARN polimerasa (ARNP) para comenzar el proceso de transcripción, que da lugar a un ARN mensajero (ARNm). Un promotor pueden ser de dos tipos:
 - Constitutivo: no depende de factores externos, por lo que el gen se transcribe a una tasa constante.
 - Regulador: necesita de la unión de un factor de transcripción (FT) para comenzar su actividad. Los FT pueden ser activadores, si su unión al promotor recluta a la ARNP y activa la transcripción, o represores, si evitan la unión de la ARNP y, por tanto, se inhibe la formación del ARNm.
- **RBS (Ribosome Binding Site)**: región del ARNm donde se unen los ribosomas para la traducción; es decir, la formación de la proteína. Los distintos RBS pueden tener más o menos afinidad por los ribosomas, lo que determinará la tasa de traducción.
- **CDS (Coding Sequence)**: corresponde a la secuencia de aminoácidos que formarán la proteína.
- **Terminador**: es la parte del ADN que detiene la transcripción pues, al alcanzar esta zona, la ARNP se disocia.

Para su uso en proyectos de Biología Sintética, estas cuatro biopartes están representadas mediante el lenguaje *Synthetic biology Open language* (SBOL) (Cox et al., 2018), que permite representar gráficamente circuitos genéticos de una manera estándar, facilitando el intercambio de información. Además, se encuentran individualizadas y caracterizadas por varios parámetros en librerías, en las cuales el diseñador puede escoger entre varias biopartes con diferentes propiedades y combinarlas como desee, en función de su objetivo. Una de las librerías más utilizadas es el *Registry of Standard Biological Parts* (Biobrick Foundation).

A grandes rasgos, la fase de Construcción consiste en formar sistemas cada vez más complejos, generando nuevas unidades transcripcionales y ensamblándolas en circuitos que, a su vez, se pueden combinar para crear cromosomas y genomas sintéticos. En primer lugar, las cuatro biopartes individualizadas deben ensamblarse en el orden adecuado (promotor, RBS, CDS y terminador) dentro de un **plásmido** llamado vector de destino. Así, este nuevo plásmido contendrá una unidad transcripcional. Finalmente, se pueden combinar dos unidades transcripcionales en un nuevo plásmido, formando una construcción más grande.

Para ello, es necesario el desarrollo de sistemas de alta eficiencia que permitan ensamblar las partes y comprobar si estos circuitos funcionan (Ellis et al., 2011). Dichos sistemas son los **métodos de ensamblaje**, que están clasificados en dos grupos: los basados en enzimas de restricción y los basados en secuencias solapantes (Weninger et al., 2016).

Los métodos basados en enzimas de restricción están, a su vez, clasificados en dos subgrupos: los que utilizan enzimas de restricción tipo IIS y de tipo II. Los basados en enzimas tipo IIS tienen una ventaja frente a los de tipo II: los sitios de reconocimiento y restricción no coinciden exactamente; es decir, el enzima corta fuera de su secuencia de reconocimiento generando extremos cohesivos (*overhangs*). Esto abre la posibilidad de diseñar la secuencia de nucleótidos de cadena simple que sobresale en los *overhangs*. De esta forma, podemos hacer compatibles las partes que queremos unir para ensamblar incluso más de dos partes en una sola reacción. La manera en que las partes a ensamblar están flanqueadas, teniendo en cuenta los sitios de reconocimiento y restricción de las enzimas utilizadas, se conoce como **standard de ensamblaje** (Chao et al., 2014).

De la gran variedad de métodos de ensamblaje que existen actualmente, es fundamental escoger bien con cuál se va a llevar a cabo la construcción, ya que su elección puede influir en la estructura final del circuito. El método escogido en este trabajo es el llamado GoldenBraid.

GoldenBraid es un *standard* de ensamblaje binario basado en la metodología Golden Gate que emplea enzimas de restricción de tipo IIS para generar circuitos sintéticos multigénicos de manera robusta (Sarrion-Perdigones et al., 2011). Consiste en construir plásmidos cada vez más complejos para, finalmente, formar circuitos sintéticos. Estas construcciones tienen 3 niveles de complejidad:

- Nivel 0: se refiere a un plásmido que contiene una de las 4 partes básicas.
- Nivel 1: plásmido que contiene todas las biopartes; es decir, una unidad transcripcional completa.
- Nivel 2: plásmido formado por dos o más unidades transcripcionales.

Estos niveles se construyen en orden; es decir, para llegar a uno es necesario haber alcanzado el anterior. Para el ensamblaje Nivel 1, se necesita un vector de destino con capacidad de albergar una unidad transcripcional completa. En la metodología GoldenBraid, se puede escoger entre dos tipos de vectores de destino Nivel 1: α_1 y α_2 . Para alcanzar el Nivel 2, se necesitan plásmidos de tipo Ω , y sólo se pueden unir en una construcción de Nivel 2 insertos que procedan de dos vectores de Nivel 1 distintos: uno debe proceder de un plásmido α_1 y otro, de α_2 (de ahí el concepto de ensamblaje binario).

Una vez construido el plásmido que contiene el circuito, éste se introduce en un microorganismo, normalmente *Escherichia coli* (*E. coli*), mediante transformación.

1.1.3. Comprobación experimental

La fase de Comprobación Experimental comienza tomando datos experimentales a partir de un cultivo de bacterias que expresan el circuito sintético. En nuestro caso, se toman medidas de **fluorescencia**, que permite medir el comportamiento de la proteína de interés (en nuestro caso es la Proteína Fluorescente Verde (GFP)), y **absorbancia**, para medir el crecimiento celular. Después, es necesario estandarizar las unidades; esto es, convertirlas a unas unidades tipificadas que permitan comparar los datos experimentales con el modelo matemático previamente diseñado, mediante un procedimiento de calibración. Estas unidades son moléculas de fluoresceína equivalente (**MEFL**) para la fluorescencia y **partículas** para la absorbancia (Boada et al., 2019). Finalmente, se obtiene la ratio entre MEFL y partículas (**MEFL/partículas**) como medida de la fluorescencia específica de cada célula. Así, las mediciones experimentales quedan

en un dominio común en el que pueden ser comparadas entre ellas y con los resultados obtenidos a partir de los modelos matemáticos.

1.1.4. Aprendizaje

El Aprendizaje es la última etapa del ciclo DBTL, añadida recientemente gracias al acceso a datos masivos y a la automatización de las tecnologías implicadas. Mediante algoritmos diseñados específicamente para cada función, esta fase permite el autoaprendizaje de nuevos conocimientos con la finalidad de incorporarlos en los siguientes diseños. Los datos son analizados con la intención de optimizar el proceso y repetir el ciclo de manera mejorada (Valero, 2018).

En nuestro caso, la fase de Aprendizaje se centra en **estimar los valores de los parámetros** que hacen que el circuito se comporte de la manera deseada. Para ello, se utiliza la metodología de **Optimización Multi-Objetivo (MOO)**, tal y como se explica en el apartado de Materiales y Métodos.

1.2. El circuito I1-FFL - Comportamiento adaptativo

Las redes de regulación transcripcional, que controlan la expresión de los genes en las células, son sistemas complejos formados por diversos nodos que interactúan entre sí. Para simplificar su estudio, se busca definir primero la dinámica de las unidades que las componen como base para entender la dinámica de toda la red (González Boscá, 2014).

Los avances en la recopilación y el análisis de datos están generando una cantidad de información sin precedentes sobre las redes de regulación transcripcional, y están demostrando que muchas de estas unidades o módulos de circuitos aparecen de manera muy frecuente y repetitiva en las redes genéticas de diversos organismos. Por tanto, se deduce que cumplen funciones importantes, lo que ha llevado a muchos grupos a dedicarse a su estudio. Estas estructuras se conocen como **motivos de red** (Shen-Orr et al., 2002).

Uno de los motivos más estudiados es el bucle de retroalimentación incoherente de tipo 1 o I1-FFL (Tyson et al., 2003; Goentoro et al., 2009; Ma et al., 2009). Se trata de un componente de circuitos más complejos que aparece cientos de veces de manera natural en organismos como *E. coli* (Eichenberger et al., 2004; Mangan et al., 2006; Milo et al., 2002) y levaduras (Lee et al., 2002; Milo et al., 2002). En células animales, el I1-FFL aparece también en redes de regulación génica en células madre embrionarias humanas (Boyer et al., 2005).

Además, este circuito resulta muy interesante desde el punto de vista de las ciencias biológicas porque presenta una importante propiedad conocida como **adaptación perfecta**, que es la habilidad de responder ante un estímulo y volver al valor anterior a dicho estímulo, aunque la señal persista (Boada Acosta, 2018). Esta propiedad está ligada a la homeostasis (Alon, 2007) e influye en otros procesos fundamentales de los organismos, como se verá más adelante.

La capacidad de adaptación se puede medir utilizando dos índices: sensibilidad y precisión (Figura 1.2). La sensibilidad mide la proporción entre el cambio de la salida frente al cambio de la entrada tras aplicar el estímulo. La precisión mide la diferencia entre el valor al que vuelve el sistema tras el cambio inicial frente al estímulo y su valor inicial; es decir, el error de la salida. En el contexto de un sistema biológico, entendemos “entrada” como el sustrato que estimula al sistema y “salida” como la sustancia producida.

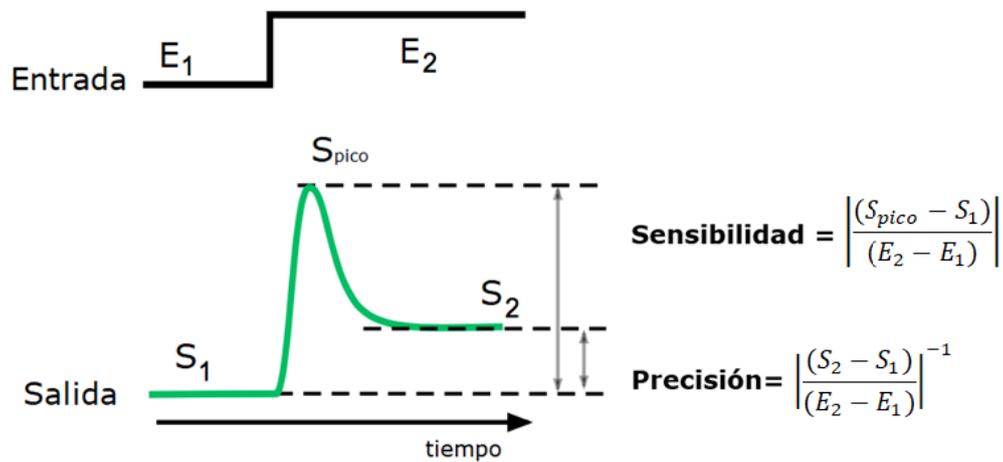


Figura 1.2. Representación gráfica y matemática de la adaptación de un circuito genético, su sensibilidad y precisión (Inspirada en Oleas, 2018).

Existen distintas implementaciones del circuito I1-FFL, tales como redes enzimáticas, genéticas o transcripcionales (Ma et al., 2009; Boada et al., 2016). En nuestro caso de estudio, la implementación se basa en un circuito de 3 genes: *gfp*, *luxR* y *cl*. A partir de estos genes se expresan las proteínas GFP (la salida del sistema), LuxR y *cl*, respectivamente (Figura 1.3).

En el circuito también interviene la molécula N-acil-L-homoserina lactona (AHL), que es la entrada del sistema. El AHL extracelular (AHL_{ext}) entra en la célula a través de la membrana por difusión y una vez internalizado, actúa como inductor del circuito. El AHL interno se une a la proteína LuxR, que se expresa constitutivamente, y forma el monómero (R.A), que a su vez dimeriza dando lugar a $(R.A)_2$. Por un lado, este dímero se une al promotor híbrido $P_{lux/cl}$ del gen *gfp*, activando la expresión de GFP. Por otro lado, $(R.A)_2$ se une al promotor P_{lux} del gen *cl* y estimula la producción de su correspondiente proteína, la cual se une también al promotor híbrido e inhibe la expresión de GFP. Por tanto, $(R.A)_2$ es un factor de transcripción que activa tanto la expresión de GFP como la de *cl*: activa la producción de GFP por una vía directa pero indirectamente la reprime, mediante la activación del inhibidor *cl*.

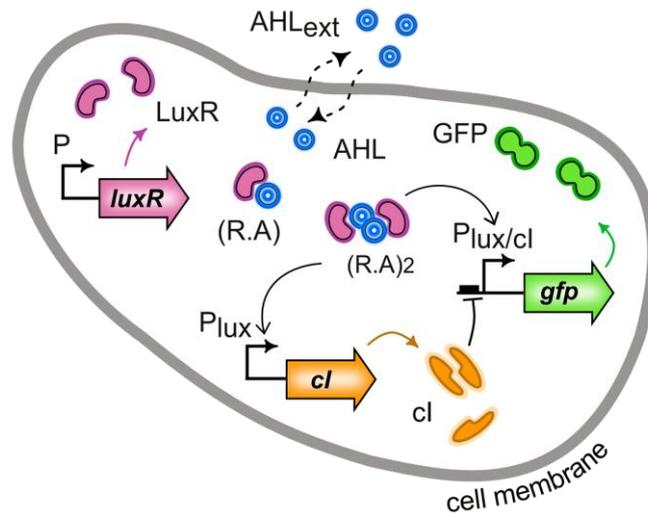


Figura 1.3. Representación de una célula con el circuito I1-FFL incorporado. La proteína LuxR (producida constitutivamente por el gen *luxR*) forma el monómero (R.A) con el inductor AHL. (R.A) dimeriza formando (R.A)₂, que actúa como factor de transcripción, activando a los genes *cl* y *gfp*. A su vez, la proteína *cl* inhibe la producción de GFP (Boada Acosta, 2018).

Concretamente, el circuito I1-FFL está diseñado como dos plásmidos para su implementación en la bacteria *E. coli*: un plásmido contendrá la unidad transcripcional del gen *cl* y el otro, las de los genes *luxR* y *gfp*. El hecho de separar los genes en plásmidos distintos permite una mayor modularidad y flexibilidad a la hora de cambiar los parámetros de uno de los genes; para ello tan solo habría que modificar su plásmido correspondiente.

1.3. Aplicaciones

Gracias al desarrollo de herramientas de Ingeniería Genética, las células pueden ser modificadas genéticamente y utilizarse como fábricas para convertir materias primas en biocombustibles, biopolímeros, productos químicos, aditivos alimentarios o productos farmacéuticos, entre otros (Keasling, 2010). Estas “fábricas de células” se desarrollan mediante el ciclo iterativo DBTL y se pueden llevar a cabo a través de diferentes estrategias, tales como la creación de rutas metabólicas de novo, la optimización de la expresión génica dentro de rutas ya existentes o el reemplazo de genes específicos (Julleson et al., 2015).

Como se ha comentado anteriormente, el circuito I1-FFL ha sido objeto de estudio por la alta frecuencia con la que aparece en los sistemas sensoriales de los organismos. Este circuito es fundamental, por ejemplo, en la quimiotaxis bacteriana (Alon et al., 1999; Sourjik y Berg, 2002), que es el proceso por el cual los cambios en la concentración de un sustrato en el medio circundante influyen en la actividad de propulsión de la célula, permitiéndole desplazarse hacia el lugar de mayor o menor concentración, según si es una sustancia atrayente o repelente. También interviene en la homeostasis de los organismos (Alon, 2007), propiedad que les permite mantener una condición interna estable compensando los cambios ambientales mediante el intercambio regulado de recursos y energía con el exterior. Todo esto es posible gracias a la existencia de una red de sistemas de control de retroalimentación que constituyen los mecanismos de autorregulación en las células (González Boscá, 2014). Estudiando estos procesos, se ha descubierto que el I1-FFL tiene un amplio abanico de funciones, actuando como

un mecanismo para generar pulsos, acelerar respuestas, detectar cambios en la intensidad de las señales, amortiguar el ruido o lograr adaptación perfecta (Reeves, 2019).

Sobre todo, el I1-FFL es un elemento con gran potencial en Biología Sintética cuando se combina con otros circuitos para formar un gran sistema. Trabajos como el de Ma et al. (2009) o Reeves (2019) demuestran que, en conjunto, los bucles de retroalimentación negativa (FFL) y los de retroalimentación incoherente (iFFL) son útiles para conseguir una expresión génica robusta. Por un lado, los FFL modifican la entrada del sistema en respuesta a cambios en la salida. Sin embargo, son difíciles de implementar y regular. Por su parte, en los iFFL la salida del sistema está regulada por la entrada, que la activa y reprime simultáneamente, independientemente de la salida (Tomazou y Stan, 2018). Además, los iFFL garantizan una adaptación perfecta a las perturbaciones de la entrada, y son relativamente simples de implementar en una célula.

Así pues, vemos que combinando distintos circuitos (entre ellos, el I1-FFL) se pueden crear sistemas más grandes y complejos a nivel de regulación, con aplicaciones en diversos campos de la Biotecnología, como la llamada Biotecnología Blanca o Industrial.

1.4. Limitaciones actuales

El impresionante avance en el campo de la Biología Sintética durante la última década ha sido en gran medida consecuencia de aplicar una visión propia de la ingeniería en un contexto diferente como es un sistema biológico. Los circuitos genéticos se conciben como interruptores que activan o desactivan la expresión de un gen, e incluso se pueden programar para actuar ante señales que reciben del entorno. Sin embargo, es esta inspiración en campos como la Electrónica y la Computación la que ha dado lugar a las principales limitaciones, pues los sistemas biológicos presentan ciertas diferencias con respecto a los sistemas electrónicos (Macía y Solé, 2011).

Por este motivo, en la actualidad todavía existe una brecha entre la posibilidad de diseñar un sistema biológico y su implementación real en el laboratorio. Esta brecha puede atribuirse a la **ausencia de partes biológicas bien caracterizadas** y a la **falta de mediciones reproducibles y estandarizadas**, entre otras causas (Way et al., 2014).

En este punto, es importante destacar el enfoque ascendente del ciclo DBTL (Schwille et al., 2018). Esto significa que el diseñador comienza desde abajo, combinando partes básicas, y va subiendo la complejidad hasta construir sistemas completos. Para ello, existe la necesidad esencial de tener unas partes bien caracterizadas (Church et al., 2014), lo que sigue siendo un gran obstáculo. A pesar de que muchas de las biopartes utilizadas en Biología Sintética sí están caracterizadas individualmente en librerías, no se conoce su comportamiento una vez integradas en el circuito. En la práctica, las partes biológicas están sujetas a la incertidumbre (Boada et al., 2016); es decir, dependen de un entorno celular cambiante.

De hecho, en la fase de modelado, uno de los retos principales es tener en cuenta todos los factores que alteran los niveles de expresión génica y que queden representados en el modelo matemático. Muchos cambios ambientales como la cepa del organismo hospedador utilizada, cambios en la composición del medio, la temperatura o la tasa de crecimiento son factores que pueden alterar el número de copias de un plásmido (C_N) y, en consecuencia, los niveles de expresión. Esto supone un problema a la hora de controlar los procesos industriales, pues una de las características que se busca es que los sistemas sean robustos frente a los cambios externos. La robustez no implica que no haya cambios en el sistema, sólo en su funcionalidad.

Muchas veces, el cambio es una parte esencial de la funcionalidad de un sistema: si el sistema cambia de la forma adecuada, se puede considerar robusto.

Por otro lado, la falta de mediciones reproducibles y estandarizadas es una importante limitación en la fase de Comprobación Experimental, pues no podemos comparar dos sistemas en términos de mediciones si no han sido creados con la misma metodología (Beal et al., 2018). Afortunadamente, hay muchos estudios enfocados en la resolución de este problema (Beal et al., 2019; Boada et al., 2019; Martin et al., 2019; Robinson et al., 2020).

En este trabajo, se pretende recorrer todas las fases del ciclo DBTL orientadas al diseño de un circuito I1-FFL que presente comportamiento adaptativo, teniendo en cuenta las limitaciones citadas. Para ello, se plantea una extensión de la metodología desarrollada por Boada et al. (2019), aportando así una posible solución al problema de caracterización de las partes. Por otro lado, se propone un modelo simplificado respecto a su versión anterior que, además, tiene en cuenta un nuevo parámetro clave: el número de copias del plásmido (C_N).

2. Objetivos

La finalidad principal del trabajo es transitar las distintas etapas del ciclo DBTL aplicadas a un circuito genético sintético I1-FFL que presente capacidad de adaptación.

Los objetivos específicos del proyecto son:

- Modelar matemáticamente un circuito I1-FFL.
- Analizar la influencia de los parámetros del circuito en su funcionamiento a través de simulaciones.
- Diseñar los plásmidos necesarios para la construcción del circuito genético deseado.
- Calibrar y estandarizar datos experimentales para compararlos con el modelo matemático.
- Optimizar los parámetros del modelo matemático del circuito a partir de datos experimentales.
- Discutir las posibles aplicaciones biotecnológicas del circuito diseñado.

3. Materiales y métodos

3.1. Diseño

3.1.1. Modelado matemático

El modelado matemático de los circuitos sintéticos permite predecir su comportamiento y seleccionar las partes biológicas que se van a utilizar en ellos.

En el caso del circuito I1-FFL, nos interesa predecir su comportamiento medio en el cultivo, no el comportamiento individual de cada célula. Por tanto, se diseña un modelo determinístico que no tiene en cuenta la naturaleza probabilística de las reacciones sino que asume que, para cada especie, la cantidad de moléculas transformadas depende solo de la cantidad actual de moléculas, las velocidades de reacción, y la estequiometría de las reacciones. Esta aproximación, se puede tomar como punto de partida para obtener modelos más complejos (Picó et al., 2015).

El modelado se realiza mediante tres pasos fundamentales:

1. Obtención de las reacciones bioquímicas
2. Obtención de los balances dinámicos de las especies (LAM)
3. Reducción del modelo (QSSA)

Partiendo del esquema del circuito (Figura 1.3), en el que están representados todos los componentes del sistema y cómo interactúan entre sí, se determinan las reacciones bioquímicas que ocurren en él. A partir de estas reacciones, utilizando la Ley de Acción de Masas (LAM), se obtiene el balance dinámico de las especies como un sistema de Ecuaciones Diferenciales Ordinarias (ODEs), que describen la evolución temporal de las principales especies bioquímicas del circuito: ADN, ARN, proteínas y factores de transcripción. Este modelo debe tener en cuenta todos los factores que puedan afectar al circuito, entre ellos el número de plásmidos en una célula (C_N). Como se comenta en la Introducción, los cambios ambientales muchas veces alteran el número de plásmidos y, en consecuencia, los niveles de expresión génica. Por tanto, la incorporación de este parámetro es un paso en el control del circuito y, a futuro, en el control de sistemas de producción.

3.1.1.1. Balances dinámicos. Ley de acción de masas

Las reacciones bioquímicas que ocurren en el interior de las células son estocásticas por naturaleza, por lo que los modelos deberían ser formulados en términos de la probabilidad de que cada una de estas reacciones tenga lugar. En cambio, los modelos deterministas, al no tener en cuenta la probabilística de las reacciones, permiten plantear de una manera más simple el circuito y facilitar su estudio. Teniendo esto en cuenta, utilizamos la LAM para expresar la velocidad de cada una de las reacciones que explican nuestro circuito. Esta ley establece que la velocidad de una reacción (bio)-química es proporcional al producto de las concentraciones de las especies reactantes elevadas a una potencia dada por la estequiometría de la reacción (Chellaboina et al., 2009).

De esta manera, a partir de las reacciones bioquímicas previamente descritas, podemos establecer matemáticamente el balance dinámico de las especies como ODEs. Cada ecuación está representada con su correspondiente estequiometría y asociada a una constante cinética. El conjunto de estas ecuaciones constituye lo que se conoce como modelo ODE. Los valores de las constantes cinéticas fueron obtenidos de estudios anteriores y a los que no se conocían, se les dio un valor provisional.

Sin embargo, el modelo que se obtiene al aplicar la LAM es de grandes dimensiones. Cuanto más detallado sea un modelo; es decir, cuantas más ecuaciones incluya, más se asemejará a la realidad. Hay que tener en cuenta que el circuito está compuesto por múltiples procesos: la unión/desunión de las especies entre sí, transcripción, traducción... Si se atiende a todos ellos, junto con el resto de los procesos celulares paralelos al circuito, el modelado se complica demasiado. Para simplificarlo, es necesaria una reducción del modelo.

3.1.1.2. Reducción del modelo. Aproximación Quasi - Estacionaria

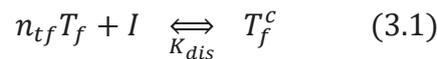
La reducción tiene como objetivo disminuir la complejidad del modelo y el número de parámetros para conseguir un modelo manejable, con el que se pueda trabajar. Es muy importante que la reducción no sea excesiva porque llevaría a una falta de relevancia biológica (Boada Acosta, 2018).

Uno de los métodos para llevar a cabo la reducción es la Aproximación Quasi-Estacionaria (QSSA). La QSSA considera que existe una separación en la escala temporal entre las distintas dinámicas de las especies (Zagaris et al., 2004; Mélykúti et al., 2014). La idea bajo este método es que las reacciones que son más veloces que otras alcanzan el estado estacionario muy rápidamente, por lo que podemos despreciar su dinámica y asumir que se encuentran en régimen permanente (Picó et al., 2015). Esta aproximación se aplica a las ecuaciones que representan a los promotores regulados por factores de transcripción (FTs), como se explica a continuación.

Reducción de las reacciones de Formación de complejos FT-inductor

Para cada promotor, la activación o represión mediada por un FT se modela con la función de Hill con la modificación propuesta por Trabelsi et al. (2018). Ésta tiene en cuenta que el número de FTs y sitios de unión aumentan o disminuyen con el número de copias del plásmido (C_N).

Considerando que un FT forma multímeros (monómeros, dímeros, etc), se comienza con el siguiente equilibrio:



donde n_{tf} es el número de moléculas de FT (en este caso, la proteína T_f) en unidades de concentración, I es la concentración del inductor, T_f^c es la concentración del multímero formado de la reacción, y K_{dis} es la constante de velocidad para la reacción de disociación del complejo T_f^c .

Ignorando el orden de unión de las especies, porque no es importante para el equilibrio final del producto sino únicamente para la cinética (no considerada aquí); y dadas las escalas de tiempo, se aplica la LAM y se obtiene la ecuación:

$$T_f^c = \frac{I \cdot T_f^{n_{tf}}}{K_{dis}} \quad (3.2)$$

donde K_{dis} depende del inductor y de la proteína considerada en cada reacción.

La ecuación clásica de Hill (Ecuación 3.3) representa la fracción de saturación (P_{fold}); es decir, la proporción de moléculas de T_f unidas al inductor frente a las moléculas de T_f totales.

$$P_{fold} = \frac{(T_f^c)^n}{(K_d)^n + (T_f^c)^n} \quad (3.3)$$

donde n es la constante de cooperación y representa la máxima inducción o el rango dinámico y K_d es la constante de equilibrio de disociación (o simplemente, constante de disociación); es decir, la concentración del complejo inductor de T_f necesaria para la inducción media-máxima.

Sin embargo, queremos considerar el hecho de que el C_N cambia el número de sitios de unión para T_f (proporcional al C_N en nuestra construcción, ya que podría haber cooperación y, por lo tanto, más de un sitio de unión por plásmido). La Ecuación 3.3 se modifica teniendo en cuenta el hecho de que, para alcanzar la saturación media-máxima de un número mayor de sitios de unión, el número de complejos de unión también tiene que ser mucho mayor:

$$P_{fold} = \frac{(T_f^c)^n}{(K_d \cdot C_N)^n + (T_f^c)^n} \quad (3.4)$$

Reemplazando la Ecuación 3.2 en la 3.4, se obtiene la nueva función de Hill:

$$P_{fold} = \frac{I^n}{\left(\frac{K_d \cdot K_{dis} \cdot C_N}{T_f^{n_{tf}}}\right)^n + I^n} \quad (3.5)$$

La Ecuación (3.5) es la que se utilizara como base para modelar los distintos promotores que intervienen en el circuito genético.

3.1.1.3. Asunciones

Así, las asunciones que se han tenido en cuenta para diseñar y simular el modelo son:

- a. Tan solo se tienen en cuenta las reacciones metabólicas que intervienen en el circuito, se asume que el resto están incluidas en la ecuación del crecimiento celular.
- b. El volumen celular se considera constante, por lo que no se tiene en cuenta su variación a lo largo del ciclo celular para el cálculo de las relaciones de transporte a través de la membrana.
- c. Hay suficiente cantidad de RNAP y ribosomas para transcribir los genes en todo momento, por lo que no son un factor limitante.
- d. La transcripción es más rápida que la traducción, por lo que se lleva a cabo una Aproximación Quasi-Estacionaria (QSSA) para el mRNA y solo se tienen en cuenta las dinámicas de las proteínas.
- e. La traducción no es un proceso sencillo (Alberts et al., 2009). Se trata de una reacción irreversible con una tasa de traducción media que tiene en cuenta que la unión de los ribosomas al sitio de unión del ribosoma (RBS) es reversible, y que varios ribosomas pueden traducir una sola copia mensajera de ARN simultáneamente.

f. La degradación y dilución debida al crecimiento celular se consideran en todas las especies, incluido el mARN, con sus correspondientes tasas de degradación d_j y tasa de crecimiento específico μ .

3.1.2. Simulación computacional

Una vez obtenido el modelo reducido, se lleva a cabo la simulación de las principales especies bioquímicas utilizando el software MATLAB®, en el cual se implementan tres archivos:

- Archivo Ejecutable
- Archivo Parámetros
- Archivo Función

El Ejecutable llama a la función ODE23 para cargar los parámetros en otro archivo. ODE23 es una función de MATLAB® que permite resolver sistemas de ecuaciones diferenciales ordinarias. El archivo Función contiene las ecuaciones del modelo reducido, y utilizará el archivo Parámetros con sus respectivos valores. El archivo Ejecutable, llama a la resolución del sistema de ecuaciones tres veces, obteniéndose los resultados de la simulación en tres partes que se detallan a continuación.

La primera parte de la simulación representa la situación inicial, que comienza con la transformación de las células con los plásmidos que contienen el circuito. En esta situación, la concentración de las especies es nula y va creciendo hasta llegar a un valor de equilibrio. Los parámetros y condiciones iniciales utilizados en la primera parte se recogen en la Tabla 3.1.

Tabla 3.1. Parámetros y condiciones iniciales de la simulación.

Variable	Descripción	Valor	Unidades
Varianza	Establece la variabilidad de los parámetros	0	adimensional
ODmax	Densidad óptica máxima	2	
ODinitial	Densidad óptica inicial	0.1	
Tfin	Tiempo de simulación hasta el estado estacionario	360	min
Step	Paso de simulación	0.1	min
Initial	Vector con la cantidad inicial de todas las especies	0	moléculas

La segunda parte de la simulación comienza con la inoculación del cultivo, haciendo una dilución a partir de un cultivo anterior. Las condiciones iniciales de la simulación en esta segunda parte corresponden con las condiciones de equilibrio a las que han llegado las especies en la primera parte. Sólo se cambia la cantidad de células para reflejar la dilución y poder comenzar la segunda parte con una cantidad de células conocida, de manera que se pueda corresponder con el inicio de un experimento.

La tercera parte se inicia simulando una inducción de AHL pasado un tiempo desde que se realizó la dilución. El circuito responde incrementando la expresión de las proteínas. Las condiciones iniciales de esta tercera parte se corresponden con los equilibrios de la segunda, salvo el valor del inductor.

Además, para ver cómo reacciona el circuito ante distintas concentraciones de inductor y distintos valores de la tasa de traducción de GFP (p_G), se realizan nuevas simulaciones.

3.2. Construcción

La construcción del circuito comienza escogiendo las partes básicas o biopartes que se van a utilizar para formar cada unidad transcripcional (explicadas en la Introducción). Estas partes se encuentran individualizadas y caracterizadas en librerías, como el *Registry of Standard Biological Parts* (Biobrick Foundation, 2006). Después, deberán ensamblarse en el orden adecuado (promotor, RBS, CDS y terminador) dentro de un plásmido llamado **vector de destino**. Recordemos que los vectores de destino de Nivel 1; es decir, los que solo albergan una unidad transcripcional, pueden ser de tipo α_1 o α_2 . Para combinar distintas unidades transcripcionales (Nivel 2), se utilizan vectores de destino de tipo Ω .

3.2.1. Ensamblaje GoldenBraid

Utilizando la metodología GoldenBraid, se lleva a cabo el ensamblaje de los plásmidos en cuestión, pasando por los tres niveles de complejidad correspondientes en su construcción.

Nivel 0

Las partes básicas previstas para utilizar en este trabajo (Tabla 3.2) se obtienen de la colección creada por el equipo Valencia UPV para la competición iGEM 2018, donde se encuentran ensambladas en plásmidos pUPD2. Estos plásmidos están diseñados de manera que cada bioparte está ya adaptada a la gramática del GoldenBraid y preparada para combinarse con otras. Las partes correspondientes a $P_{lux/cl}$ y al gen *cl*, no se encuentran en esta colección, y su construcción estaba planificada en este trabajo.

Tabla 3.2. Partes básicas adaptadas a la gramática de GoldenBraid.

Nombre	Tipo	Descripción	Fuente
P	Promotor	Promotor constitutivo de <i>luxR</i>	BBa_K2656004
P_{lux}	Promotor	Promotor inducible del gen <i>cl</i>	BBa_K2656003
P_{lux/cl}	Promotor	Promotor híbrido del gen <i>gfp</i>	Este trabajo
	RBS	RBS fuerte	BBa_K2656011
	RBS	RBS de fuerza media	BBa_K2656010
luxR	CDS	Secuencia codificante de LuxR	BBa_K2656016
cl	CDS	Secuencia codificante de <i>cl</i>	Este trabajo
gfp	CDS	Secuencia codificante de GFP	BBa_K2656023
	Terminador	Terminador transcripcional doble	BBa_K2656026

Nivel 1

Una vez se tienen las construcciones de Nivel 0, hay que unir las cuatro biopartes para formar una unidad transcripcional. La metodología GoldenBraid pretende, como hemos comentado anteriormente, hacerlo en una sola reacción. Esto es posible gracias a que los *overhangs* de las partes contiguas son complementarios, pero no lo son al resto de partes. Así, en una sola reacción se pueden ensamblar todas las partes en el orden correcto.

Con los insertos (biopartes) y vectores de destino escogidos, se procede a realizar la reacción de ensamblaje GoldenBraid (Figura 3.1), que se compone de dos reacciones:

- Digestión: con las enzimas de restricción BsaI, se extraen las partes básicas del plásmido pUPD2. También el vector de destino es digerido con BsaI, de manera que queden extremos complementarios al promotor y al terminador.
- Ligación: Las cuatro biopartes se introducen en el vector de destino correspondiente ($\alpha 1$ o $\alpha 2$). Mientras que la unidad transcripcional de *cl* se puede ensamblar en cualquiera de los dos, es importante que las de *luxR* y *gfp* vayan en vectores distintos. Para la construcción del plásmido que contiene dos unidades transcripcionales, hay que recordar que el GoldenBraid sólo permite unir en una construcción de Nivel 2 insertos que procedan de dos vectores de Nivel 1 distintos: uno debe proceder de un plásmido $\alpha 1$ y otro, de $\alpha 2$.

Los componentes y condiciones necesarios para llevar a cabo la reacción están descritos en las Tabla 3.3 y 3.4, respectivamente.

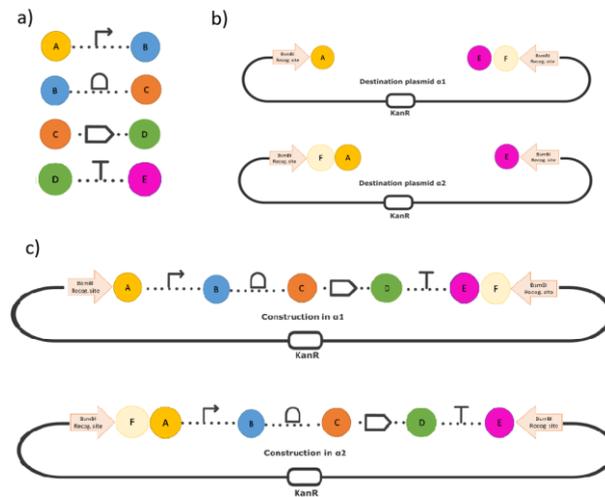


Figura 3.1. Esquema de la reacción de ensamblaje GoldenBraid de una unidad transcripcional utilizando las cuatro partes básicas y los vectores de destino de Nivel 1, ambos digeridos con BsaI. Cada letra (asociada a un color) representa la secuencia de nucleótidos que sobresale en los extremos cohesivos, siendo A: GGAG, B: TACT, C: AATG, D: GCTT, E: CGCT y F: GTCA. Las secuencias complementarias tienen la misma letra y color. KanR es resistencia a kanamicina. a) Partes básicas del ADN extraídas del plásmido pUPD2 con el enzima BsaI. b) Vectores de destino de nivel 1 ($\alpha 1$ y $\alpha 2$) digeridos con BsaI. c) Resultado del ensamblaje en ambos vectores de destino. (Requena, 2019)

Tabla 3.3. Componentes de una reacción de GoldenBraid.

Componente	Volumen (μ l)
Enzima de restricción tipo IIS (BsaI o BsmBI)	1
T4 ADN ligasa	1
T4 ADN ligasa buffer 10x	1.5
BSA 10x	1.5
ADN	75 ng de cada plásmido
ddH ₂ O	hasta 15 μ l
Total	15

Tabla 3.4. Condiciones para llevar a cabo una reacción de GoldenBraid.

Temperatura (°C)	Tiempo (min)	Número de ciclos
37	10	1
37	3	25
50	10	1
80	10	1

Los plásmidos diseñados para el circuito I1-FFL en Nivel 1 incluyen el pRAKA11, que es un plásmido definitivo, con la unidad transcripcional que expresa el gen *cl* con su correspondiente promotor (BBa_K2656003), un RBS de fuerza baja (BBa_K2656010) y un terminador (BBa_K2656026). El vector de destino utilizado para crear este plásmido es el $\alpha 1$ (Figura 3.2, a).

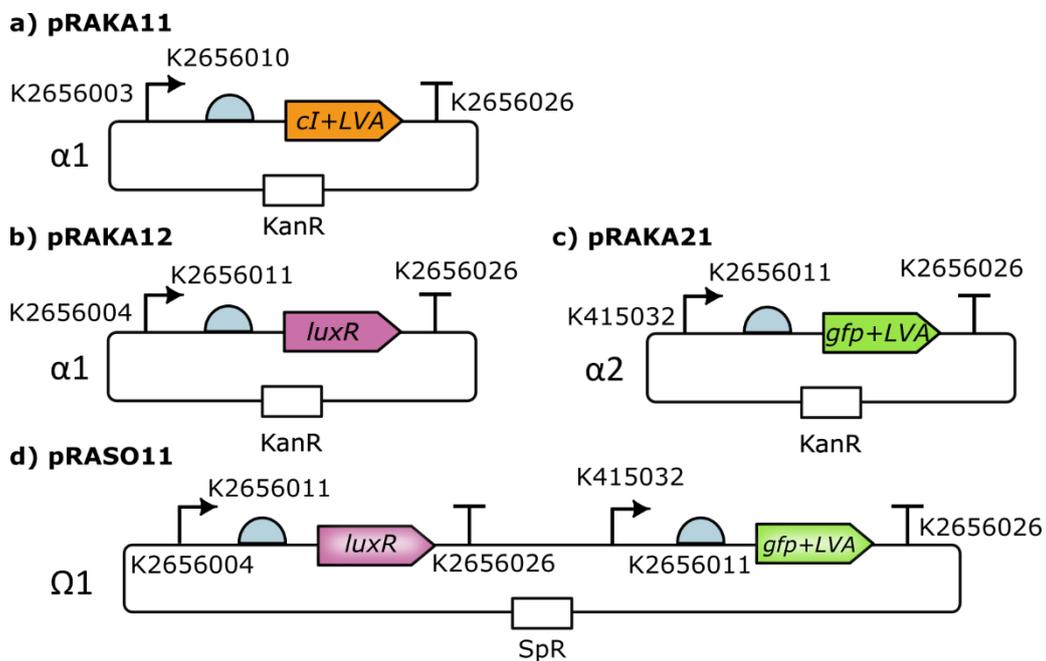


Figura 3.2. Esquema de la construcción del circuito sintético I1-FFL, con sus respectivos niveles y plásmidos.

Además, los plásmidos pRAKA12 (Figura 3.2, b) y pRAKA21 (Figura 3.2, c) contiene las unidades transcripcionales de los genes *luxR* y *gfp*. El plásmido pRAKA12 contiene el gen codificante para la proteína LuxR (BBa_K2656016) que se expresa a una tasa constante bajo el control de un promotor constitutivo de fuerza media (BBa_K2656004) y un RBS de fuerza media (BBa_K2656011). Tras la CDS, se encuentra el terminador (BBa_K2656026). El vector de destino utilizado para crear este plásmido es el $\alpha 1$.

Por otro lado, el plásmido pRAKA21 contiene los componentes para la expresión de la proteína GFP: un promotor híbrido (BBa_K415032), el RBS (BBa_K2656011) y la CDS que codifica para la proteína (BBa_K2656023). Finalmente, aparece el mismo terminador utilizado anteriormente. El vector de destino utilizado para crear este plásmido es el $\alpha 2$.

Nivel 2

Para el circuito I1-FFL, solo es necesario llegar a este nivel en la construcción del plásmido pRASO11 (Figura 3.2, d), ya que en él queremos incluir dos unidades transcripcionales (*luxR* y *gfp*) provenientes de los plásmidos pRAKA12 y pRAKA21.

Una vez tenemos estas dos unidades transcripcionales una en un plásmido α 1 y la otra en un α 2, podemos combinarlas en un plásmido de destino Ω 1. De la misma manera, utilizando los mismos componentes y condiciones descritos en las Tablas 3.3 y 3.4, se lleva a cabo la reacción de ensamblaje GoldenBraid. Los vectores de destino llevan resistencia a un antibiótico para que sólo puedan crecer las bacterias que hayan incorporado el plásmido: los vectores α tienen resistencia a kanamicina y los Ω , a espectinomicina.

3.2.2. Transformación

Una vez obtenidos los plásmidos finales (pRAKA11 y pRASO11), se lleva a cabo la transformación; es decir, la inserción de un plásmido de interés en una bacteria. En este caso, se utilizan *E. coli* de cepa 10G[®] electrocompetentes.

La transformación se realiza por electroporación, una técnica que consiste en generar poros en la membrana plasmática aplicando un determinado voltaje durante un corto período de tiempo. De esta manera, se consigue que las bacterias incorporen el plásmido antes de que los poros se cierren. Para ello, hay que enfriar un alícuota de 50 μ L células electrocompetentes a -80°C en hielo y añadir de 1 a 5 mL del plásmido de interés. A continuación, las células se homogeneizan, se introducen en una cubeta de electroporación fría y seca y ésta se coloca en el electroporador. Tras la electroporación, a un voltaje de 1700V, se extrae la cubeta y se le añaden 700 mL de medio de supervivencia SOC. Después, el contenido de la cubeta se introduce en un tubo de cultivo y se incuba a 37°C y 250 rpm durante 1 hora. Finalmente, se cultiva en medio sólido con el antibiótico correspondiente y se incuba durante la noche a una temperatura de 37°C .

Para seleccionar las bacterias que han incorporado el plásmido se utiliza un gen selector, como el que codifica para la Proteína Fluorescente Roja (RFP). Este gen está presente en el vector de destino y está localizado en el lugar que, tras la reacción de ensamblaje, ocupará el inserto. Por tanto, si el plásmido incorpora el inserto, el gen selector se libera y no debería observarse coloración roja. Por el contrario, si el ensamblaje no es correcto, la bacteria expresará la RFP y se verá roja.

3.2.3. PCR de colonias (cPCR)

La cPCR sirve para amplificar una región de interés del ADN de un cultivo bacteriano. En este caso, se pretende amplificar la región donde se encuentra el inserto de los vectores creados por GoldenBraid. En primer lugar, se selecciona una colonia aislada del medio de cultivo y se introduce en 50 mL de agua desionizada estéril. El proceso se lleva a cabo utilizando el kit *Taq DNA polymerase kit with Standard Taq buffer* (New England Biolabs), cuyos componentes y condiciones se encuentran recogidos en las Tablas 3.5 y 3.6, respectivamente.

Los cebadores necesarios son VerTUdir (directo) y VerTUrev (reverso), diseñados para amplificar la región de interés de los vectores pARK (α) Y pRMSO (Ω).

Tabla 3.5. Componentes para una cPCR.

Componente	Volumen (μ l)
10x Standard Taq buffer	2
10 mM dNTPs	2
Cebador Directo 10 μ M	0.4
Cebador Reverso 10 μ M	0.4
Taq DNA polymerase	0.1
Colonia inoculada en agua	5
ddH ₂ O	10.1
Volumen total	20

Tabla 3.6. Condiciones para una cPCR

Nombre	Temperatura	Tiempo	Número de ciclos
Desnaturalización inicial	95 °C	2 min	1
Desnaturalización	95 °C	20 s	30
Alineamiento	56 °C	1 min	
Extensión	68 °C	1 min/kb	
Extensión final	68 °C	5 min	1

3.2.4. Electroforesis en gel de agarosa

El proceso de electroforesis en gel permite separar distintas muestras de ADN en función de su tamaño.

La preparación de 80 mL de gel a una porosidad del 1% se realiza introduciendo 0.8 gramos de agarosa con 80 mL de buffer SB en un matraz Erlenmeyer. A continuación, el matraz se lleva a ebullición en el microondas y, cuando se vuelve cristalino, se deja enfriar hasta los 50 °C. Luego, se vierte el contenido en la cubeta de electroforesis y se deja reposar hasta que adquiera una consistencia de gel, colocando los peines en la posición adecuada. Una vez listo, cargar las muestras con Loading Dye Purple (6X) (New England Biolabs) y O'GeneRuler 1Kb ADN Ladder (Fisher Scientific). Tras seleccionar el voltaje deseado, se corre el gel y se visualiza en un transiluminador UV.

3.2.5. Cultivos celulares

Las células de interés se cultivan tanto en medios con los nutrientes y antibióticos necesarios para su crecimiento. En este trabajo, se utilizan medios que pueden ser tanto líquidos como sólidos.

- Preparación de medios de cultivo líquidos
 - LB Broth
Añadir 10 g de triptona, 5 g de extracto de levadura, 5g de NaCl y 1 L de agua destilada en una botella. Homogeneizar y esterilizar en un autoclave. Dejar enfriar. Añadir 1 mL de solución antibiótica 1000x, si queremos un medio selectivo.
 - Super optimal broth with catabolite repression (SOC)

Primero, se prepara el SOB, para ello: añadir 20 g de triptona, 5 g de extracto de levadura, 2 mL de 5M NaCl, 2.5 mL de 1M KCl, 10 mL de 1M MgCl₂, 10 mL de 1M MgSO₄ en una botella. Ajustar a 1 L con agua destilada, homogeneizar y esterilizar en autoclave.

Para hacer SOC con SOB, dejar que SOB enfríe y, en condiciones asépticas, añadir 20 mL de glucosa 1M esterilizada.

- M9 minimal media

Disolver 12.5 g de M9 Medium Broth Powder (VWR Amresco Life Science) en 1 L de agua destilada y desionizada. Homogeneizar y esterilizar en autoclave. Dejar enfriar y, en condiciones asépticas, añadir 2 mL de 1M MgSO₄ esterilizado, 10 mL de glucosa esterilizada al 20 % y 0,1 mL de 1M CaCl₂ esterilizado. Si se usa con bacterias resistentes a un antibiótico, agregar 1 mL de solución madre de 1000x de ese antibiótico.

- Preparación de medios de cultivo sólido- cultivo en placa Petri

Todos los cultivos en placas Petri se realizaron con LB (Lennox) y agar como medio de cultivo. Para hacerlos selectivos, se usaron soluciones madre 1000x de los antibióticos. Añadir 10 g de triptona, 5 g de extracto de levadura, 5 g de NaCl, 20 g de Agar-agar y 1 L de agua destilada a una botella. Homogeneizar y esterilizar en autoclave. Dejar enfriar y añadir 1 mL 1000x solución antibiótica. Finalmente, en condiciones asépticas, extender 25 mL del contenido de la botella en cada placa Petri. Dejar reposar hasta que el medio gelifique.

3.2.6. Glicerinado bacteriano

Su función es almacenar las bacterias en un medio seguro, protegidas de temperaturas inferiores a 0 °C.

Para prepararlo, se añaden 1440 µL de cultivo bacteriano de interés (cultivado *overnight*) y 360 µL de glicerol al 80% en un microvial, manteniendo la esterilidad. Homogeneizar la mezcla y almacenar a -80 °C.

3.2.7. Extracción de plásmidos: miniprep

Mediante este procedimiento, realizado con *el QIAprep Spin Miniprep Kit* (Qiagen), se extraen o purifican los plásmidos de los cultivos bacterianos. En primer lugar, se centrifuga un cultivo bacteriano (previamente transformado con un plásmido de interés y crecido *overnight* en LB) para separar las células del medio de cultivo y lisarlas, dejando el plásmido libre. Después, el contenido se pasa por una columna de afinidad que retiene el plásmido de interés. Finalmente, éste se aísla utilizando un tampón de elución.

3.3. Comprobación experimental

Para obtener información sobre el comportamiento de los organismos que han incorporado el circuito, es necesario medir ciertas magnitudes, que en este caso son la fluorescencia y la absorbancia. Después, estos datos se calibran y se estandarizan las unidades. Así pues, en este apartado podemos distinguir dos partes: la toma de datos experimentales y su calibración.

3.3.1. Medidas de fluorescencia y OD₆₀₀ *in vivo*

Las medidas experimentales se realizan en Biotek Cytation™ 3, un lector de fluorescencia y densidad óptica (OD). El lector toma medidas a distintos tiempos mientras las células de interés están siendo incubadas a ciertas condiciones (Tabla 3.7) en una placa de 96 pocillos.

Se aíslan dos colonias de bacterias diferentes que han incorporado el plásmido de interés. Para ello, se dejan crecer en dos cultivos de LB con su correspondiente antibiótico y se incuban *overnight* a 37 °C y 250 rpm. A continuación, se realizan 3 réplicas de 3 mL de cultivo en M9 con su correspondiente antibiótico a una OD₆₀₀ inicial = 0.04, con cada uno de los cultivos *overnight* y se incuban a 37 °C, 250 rpm durante 4 horas. Para cada réplica, se realizan diluciones de 3 mL a una OD = 0.1. A continuación, se carga una placa de 96 pocillos con 10 muestras de 200 µl de cada cultivo y se rellenan algunos pocillos con M9, para usarlos como blanco.

Tabla 3.7. Condiciones para la toma de medidas de OD₆₀₀ y fluorescencia de la GFP.

Tiempo entre medidas	5 min
Temperatura	37 °C
Agitación	Doble órbita (continuamente)
Longitud de onda de absorbanca	600 nm
Longitud de onda de excitación	485 nm
Longitud de onda de emisión	528 nm

3.3.2. Calibración

El objetivo de la calibración es procesar los datos experimentales “crudos” obtenidos de un cultivo bacteriano que ha incorporado el circuito para convertirlos a una unidad estándar. Esta unidad se utiliza para comparar los datos experimentales con el modelo y así llevar a cabo la caracterización de las partes. Para ello, se sigue el protocolo propuesto por Boada et al. (2019).

En este protocolo se toman medidas de referencia a partir de una solución de fluoresceína (para la fluorescencia) y unas partículas de sílice similares a las células de *E. coli* (para la absorbanca). Con estas medidas, se obtienen unas curvas de calibración (Figura 3.3) que permiten relacionar las unidades que devuelve el lector (unidades arbitrarias) con unas unidades calibradas: MEFL para la fluorescencia y partículas para la absorbanca. Finalmente, se calcula su cociente, MEFL/partícula, que sirve como unidad estándar para la fluorescencia específica de cada célula y se utiliza para comparar los datos experimentales y el modelo computacional.

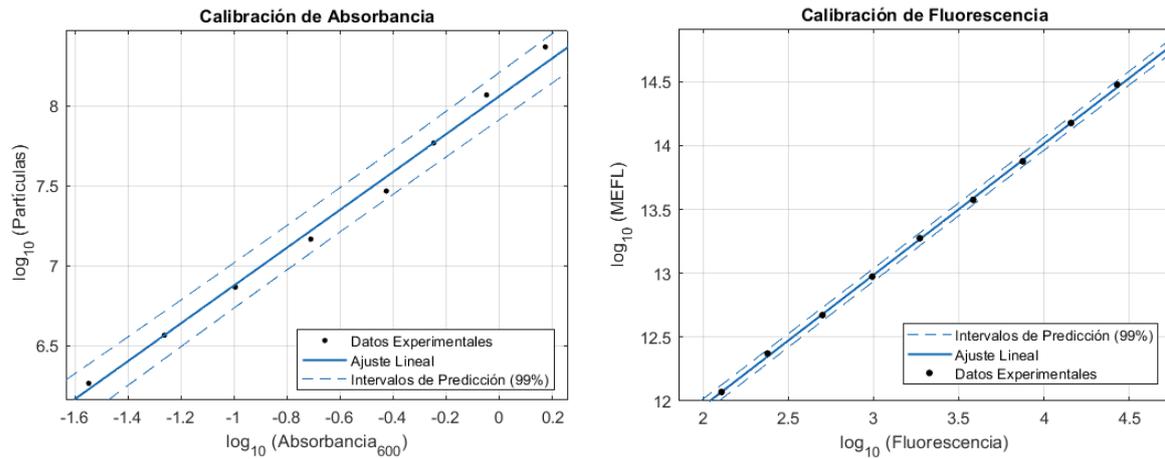


Figura 3.3. Curvas de calibración de absorbancia (izquierda) y fluorescencia (derecha).

3.4. Aprendizaje

Como se ha comentado anteriormente, una de las limitaciones actuales que encuentra la Biología Sintética es la ausencia de caracterización de las partes biológicas implementadas en un circuito, dentro de un contexto celular cambiante. Para afrontarla, el modelo tiene que ir mejorando a base de iteraciones.

Una manera de hacerlo es comprobando que el modelo matemático diseñado presenta un comportamiento similar a los datos experimentales; es decir, minimizando el error entre el modelo simulado y los datos del laboratorio. Sin embargo, la mayoría de las veces los requerimientos del circuito son excluyentes entre sí, esto es, la optimización de uno conduce a empeorar algún otro. La existencia de múltiples requerimientos y necesidades en un problema de diseño nos lleva a hablar de un problema multi-objetivo (Marler y Arora, 2004). Por tanto, la mejor manera de enfocarlo es con un procedimiento de Optimización Multi-Objetivo (MOO). En la MOO, todos los objetivos son importantes para el diseñador, por lo que todos son optimizados simultáneamente y la solución suele no ser única, sino un conjunto de soluciones conocido como conjunto de Pareto. En dicho conjunto, todas las soluciones son óptimas y ninguna es mejor que otra, solo difieren entre sí por el grado de compromiso que guardan entre los objetivos optimizados (Boada et al., 2014).

Así pues, la metodología seguida, resumida en la Figura 3.4, es la siguiente:

- -1) Prerrequisitos (Figura 3.4.-1): el punto de partida son el modelo matemático y las partes biológicas en las que se basa, que constituyen los parámetros que queremos caracterizar.
- 0) Preprocesamiento (Figura 3.4.0): se trata del paso previo a la optimización; es decir, la toma de medidas de las magnitudes deseadas (fluorescencia y absorbancia) y la calibración de los resultados. Además, en este paso también se definen unos intervalos iniciales para los parámetros del modelo que se van a estimar, y unos valores fijos para los parámetros que consideramos constantes.

Una vez tenemos los prerrequisitos y el preprocesamiento de los datos, se aplica la metodología MOO para la caracterización de las partes biológicas. La idea es realizar una

estimación de los parámetros *in situ* (Alarcón-Ruiz et al., 2020); es decir, en el contexto del circuito. La MOO consta de tres pasos:

- 1) Definición del problema multi-objetivo (MOP) (Figura 3.4.1): Consiste en especificar los objetivos que queremos minimizar. Éstos se representan con una **función de coste** o función objetivo, que depende de los parámetros que nos interesan. Así pues, en este paso se definen el **número de objetivos, $J(\theta)$** , y las **variables de decisión (parámetros), θ** . En este caso, establecemos dos objetivos por experimento (por cada nivel de inducción de AHL):
 - o **$J_1(\theta)$** es el error entre la fluorescencia predicha y la experimental (medida en MEFL/partícula).
 - o **$J_2(\theta)$** es el error entre el número de células predichas y las experimentales (medida en partículas).

Las variables de decisión son aquellos parámetros que se pueden modificar en el laboratorio y queremos estimar para conocer su valor óptimo. Estas son:

- o la tasa de transcripción (asociada con la fuerza del promotor)
- o la tasa de traducción (asociada con la fuerza del RBS)
- o el número de plásmidos dentro de una célula

- 2) Proceso de Optimización Multi-Objetivo (MOOP) (Figura 3.4.2): En esta parte, se ejecuta el algoritmo spMODE (Reynoso-Meza et al., 2013), que nos devuelve el **Frente de Pareto, $J_i^*(\theta)$** ; es decir, el conjunto de soluciones igualmente óptimas que minimizan los objetivos. El correspondiente **conjunto de Pareto** incluye los valores de los parámetros asociados con el Frente de Pareto.
- 3) Toma de decisiones multicriterio (MCDM) (Figura 3.4.3): en función del criterio del diseñador, se selecciona la **solución preferida** del Frente y conjunto de Pareto correspondiente a la parte que queremos caracterizar, con la ayuda de diagramas y agrupando las soluciones en *clusters*.

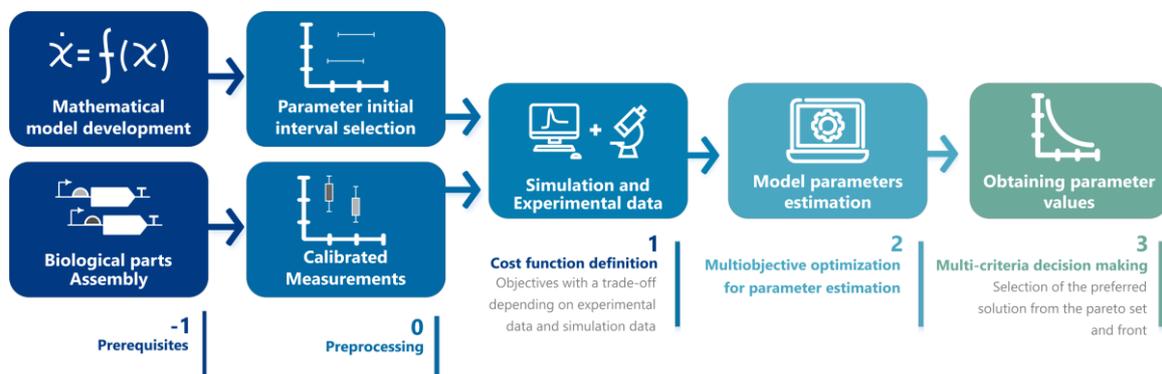


Figura 3.4. Esquema de la metodología seguida para la caracterización de las partes básicas: -1) Prerrequisitos, 0) Preprocesamiento de los datos, 1) Definición de la función de costes, 2) Optimización multi-objetivo y 3) Toma de decisiones multicriterio. (Alarcón-Ruiz, 2020).

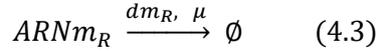
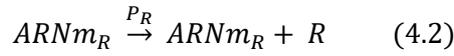
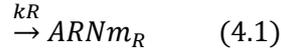
4. Resultados y Discusión

4.1. Diseño

4.4.1. Modelado matemático

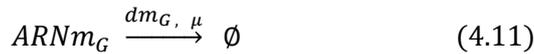
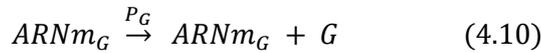
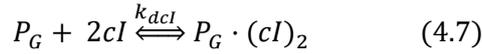
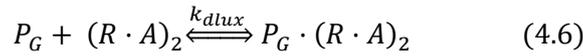
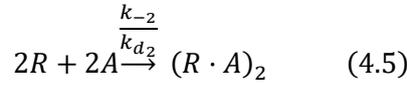
Tras establecer la estructura y funcionamiento del circuito sintético de interés (Figura 1.3), planteamos las ecuaciones bioquímicas teniendo en cuenta las reacciones y procesos biológicos. Las especies que intervienen en las ecuaciones están recogidas en la Tabla 4.1. Las constantes cinéticas vienen identificadas en la Tabla 4.2.

Gen *luxR*:



Donde la reacción (4.1) corresponde a la transcripción del gen *luxR*, produciendo su correspondiente ARNm ($ARNm_R$). Posteriormente ocurre la traducción (4.2), sintetizando la proteína LuxR (R) y manteniendo el $ARNm_R$. Finalmente, tanto el $ARNm_R$ (4.3) como la proteína (4.4) se degradan. La degradación se representa como \emptyset .

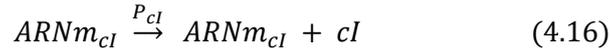
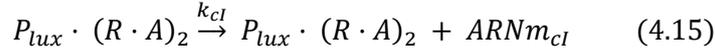
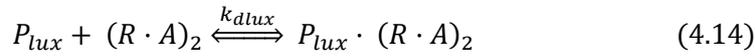
Gen *gfp*:



Donde la reacción (4.5) responde a la formación del factor de transcripción $(R \cdot A)_2$, el dímero formado por las especies LuxR y AHL. La reacción (4.6) responde a la unión y desunión del factor $(R \cdot A)_2$ con el promotor híbrido P_G , mientras que (4.7) responde a la unión y desunión del factor $(cI)_2$. La reacción (4.8) representa la transcripción basal de *gfp* mediante el promotor híbrido inducible P_G , cuando la unión del factor $(R \cdot A)_2$ no se ha producido. Seguidamente, (4.9) se

identifica con la transcripción y (4.10), con la traducción. Finalmente, las dos últimas, indican la degradación tanto del ARNm_G (4.11) como de la proteína R (4.12).

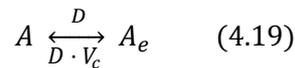
Gen cl:



Donde la reacción (4.13) representa la tasa basal de transcripción a partir del promotor inducible P_{lux} . Seguidamente, se indica la unión de $(R \cdot A)_2$ con P_{lux} (4.14) y la síntesis de ARNm_{cl} inducida (4.15). La reacción (4.16) marca la traducción, con la formación de la proteína cl. Finalmente, las ecuaciones (4.17) y (4.18) simbolizan la degradación de los componentes.

En las reacciones, no se muestran los detalles del complejo $(R \cdot A)_2$ porque en la ecuación de Hill del promotor (subsección 3.1.1.2. *Reducción de las reacciones de Formación de complejos FT-inductor*) ya se tienen en cuenta el número de moléculas que aporta cada una de las especies y el C_N .

Proceso de difusión extra/intracelular de AHLe:



Donde la reacción 4.19 representa la entrada y salida de AHLe de la célula, y donde $V_c = \frac{V_{cell}}{V_{ext}}$, por lo que depende del volumen interno y externo.

Tabla 4.1. Especies bioquímicas del circuito I1-FFL.

Especies	Descripción
1. R	Proteína LuxR
2. ARNm _R	ARNm de <i>luxR</i>
3. A	AHL intracelular
4. (R.A) ₂	Dímero formado por LuxR y AHL
5. P _G	Promotor híbrido $P_{lux/cl}$ libre
6. cl	Proteína cl
7. (cl) ₂	Dímero formado por cl
8. ARNm _G	ARNm de <i>gfp</i>
9. G	Proteína GFP
10. P _{lux}	Promotor P_{lux} libre
11. ARNm _{cl}	ARNm de <i>cl</i>
12. A _e	AHL extracelular

El modelado sigue la metodología descrita en 3.1.1. *Modelado matemático*: sobre las reacciones (4.1) – (4.19), se aplica la LAM y se formula las ecuaciones ODE correspondientes para cada una de las especies del circuito I1-FFL. Asimismo, este proceso se realiza utilizando las funciones de Hill que representan a los promotores, como se explica en la sección 3.1.1.2. *Reducción de las reacciones de Formación de complejos FT-inductor*, lo que conlleva un proceso de reducción con la QSSA. Así, se obtiene el modelo final presentado a continuación que representa la evolución temporal de la concentración de las especies LuxR, AHL, cI, GFP, A_e y el número de células (N):

$$\begin{aligned} \frac{d[R]}{dt} = \dot{x}_1 &= \frac{p_R C_N k_R}{d_{mR} + \mu} - (d_R + \mu) [R] \\ \frac{d[A]}{dt} = \dot{x}_2 &= D (V_c [A_e] - [A]) - (d_A + \mu) [A] \\ \frac{d[cI]}{dt} = \dot{x}_3 &= \frac{p_{cI} C_N k_{cI}}{d_{m_{cI}} + \mu} \left(\alpha + (1 - \alpha) \frac{\frac{1}{k_{d1lux}} \left(\frac{[R][A]}{k_{d2} C_N} \right)^2}{1 + \frac{1}{k_{d1lux}} \left(\frac{[R][A]}{k_{d2} C_N} \right)^2} \right) - (d_{cI} + \mu) [cI] \\ \frac{d[GFP]}{dt} = \dot{x}_4 &= \frac{p_G C_N k_G}{d_{m_G} + \mu} \left(\alpha + (1 - \alpha) \frac{\frac{1}{k_{d1lux}} \left(\frac{[R][A]}{k_{d2} C_N} \right)^2}{1 + \frac{1}{k_{d1lux}} \left(\frac{[R][A]}{k_{d2} C_N} \right)^2} \frac{1}{1 + \frac{[cI]^2}{k_{d_{cI}} C_N}} \right) - (d_G + \mu) [G] \\ \frac{dN}{dt} = \dot{x}_5 &= \mu N \left(1 - \frac{N}{N_{\max}} \right) \\ \frac{d[A_e]}{dt} = \dot{x}_6 &= D \left(-N V_c [A_e] + \sum_{i=1}^N [A] \right) - d_{A_e} [A_e] \end{aligned}$$

Cabe destacar que el modelo diseñado en este trabajo presenta ciertas mejoras con respecto a versiones anteriores realizadas en este mismo laboratorio. Por un lado, se trata de un modelo más reducido y sencillo de resolver, pues sólo se han tenido en cuenta las dinámicas de las proteínas, tal y como se indica en la sección de 3.1.1.3. *Asunciones*. Además, constituye el primer modelo de dicho circuito que tiene en cuenta el C_N en la representación de los promotores (Trabelsi et al., 2018). Este factor es muy relevante, ya que no se necesitan la misma cantidad de moléculas para ocupar todos los promotores cuando se tienen 10 copias del plásmido en cuestión que cuando hay 500 copias: a mayor C_N, más moléculas de FT se necesitan y, por lo tanto, más grande será la K_d aparente del promotor.

En la Tabla 4.2 se recogen valores de los parámetros del modelo utilizado, indicando el origen de dichos valores, ya sea de la literatura o del conocimiento previo del laboratorio donde se realizó mi TFG. Con el modelo diseñado y los valores de los parámetros, se llevan a cabo simulaciones computacionales para observar del comportamiento del circuito.

Tabla 4.2. Parámetros del modelo del circuito I1- FFL.

Parámetro	Descripción	Valor	Unidad	Referencia
Generales				
μ	Tasa de crecimiento específico	0.0077	min^{-1}	Obtenido experimentalmente en el laboratorio
V_c	Ratio entre el volumen externo celular y volumen externo normal en un sistema de microfluídica	5.5×10^{-12}	L	Boada Acosta, 2018
N_{\max}	Máximo número de células	32×10^7	células	Boada Acosta, 2018
C_N	Número de copias del plásmido	10	copias	Boada Acosta, 2018
α	Tasa de expresión basal	0.1	adimensional	Boada Acosta, 2018
LuxR				
p_R	Tasa de traducción	95.7	min^{-1}	Conocimiento previo del laboratorio
dm_R	Tasa de degradación del ARNm	0.23	min^{-1}	Milo et al., 2010
k_R	Tasa de transcripción	0.468	min^{-1}	Boada et al., 2019
d_R	Tasa de degradación de la proteína	0.07	min^{-1}	Boada Acosta, 2018
cl				
p_{cl}	Tasa de traducción	9.6	min^{-1}	Conocimiento previo del laboratorio
dm_{cl}	Tasa de degradación del ARNm	0.23	min^{-1}	Milo et al., 2010
K_{cl}	Tasa de transcripción	0.468	min^{-1}	Boada et al., 2019
d_{cl}	Tasa de degradación de la proteína	0.05	min^{-1}	Conocimiento previo del laboratorio
Promotor P_{lux}				
k_{dlux}	Constante de disociación del dímero $(R.A)_2$ al promotor	500	moléculas	Conocimiento previo del laboratorio
Dímero $(R.A)_2$				
k_{d2}	Constante de disociación entre los monómeros $(R.A)$	2165	moléculas	Boada et al., 2019
Promotor híbrido $P_{lux/cl}$				
k_{dcl}	Constante de disociación de cl al promotor híbrido	4800	moléculas	Boada et al., 2019
k_{dlux}	Constante de disociación del dímero $(R.A)_2$ al promotor	1000	moléculas	Conocimiento previo del laboratorio
GFP				
p_G	Tasa de traducción	71.5	min^{-1}	Conocimiento previo del laboratorio
dm_G	Tasa de degradación del ARNm	0.23	min^{-1}	Milo et al., 2010
k_G	Tasa de transcripción	0.468	min^{-1}	Boada et al., 2019
d_G	Tasa de degradación de la proteína	0.1	min^{-1}	Conocimiento previo del laboratorio
AHL				
D	Tasa de difusión del AHL a través de la membrana	4	min^{-1}	Conocimiento previo del laboratorio
d_A	Tasa de degradación del AHL intracelular	4×10^{-4}	min^{-1}	Milo et al., 2010
d_{Ae}	Tasa de degradación del AHL extracelular	4.81×10^{-5}	min^{-1}	Horswill et al., 2007

4.1.2. Simulaciones del circuito I1-FFL

Tras obtener el modelo ODE que representa el comportamiento del circuito, se puede pasar a realizar experimentos computacionales; es decir, simulaciones. Para ello se utiliza la metodología presentada en el apartado 3.1.2. *Simulación computacional*.

La primera parte de la simulación (Figura 4.1, A) representa la situación inicial, aquella en la que recién se han transformado las células y los plásmidos que contienen el circuito se encuentran “por primera vez” en el entorno celular. En esta situación, la concentración de todas las especies es nula (excepto N, ya que las células están creciendo en el cultivo). Los parámetros y condiciones iniciales utilizados se recogen en la Tabla 3.1. En esta primera parte, las cantidades de moléculas de las especies se incrementan desde cero (o la cantidad inicial de células) hasta llegar a un valor de equilibrio.

La segunda parte de la simulación (Figura 4.1, B) representa la situación que ocurre en el laboratorio cuando se realizan los experimentos en los que se inocula el cultivo, haciendo una dilución a partir de un cultivo anterior. En este caso, la dilución se realiza 4 horas después de la transformación de las células. Las condiciones iniciales de la simulación en esta parte B corresponden con las condiciones de equilibrio a las que han llegado las especies en la parte A. Sólo se cambia la cantidad de células para reflejar la dilución y poder comenzar la parte B con una cantidad de células conocida, de manera que se pueda corresponder con el inicio de un experimento. Las concentraciones iniciales de las proteínas en la parte B, entonces, se deben a que los genes tienen una tasa de expresión basal (α); es decir, siempre se produce una pequeña cantidad de proteína sin necesidad de inducción.

Pasado un determinado tiempo desde la dilución, se realiza una inducción que se computa en la tercera parte de la simulación (Figura 4.1, C). El circuito responde incrementando la expresión de las proteínas. Todas las condiciones iniciales de la parte C se corresponden con los equilibrios de la parte B, salvo el valor del inductor. De esta manera, se puede observar si el modelo diseñado presenta el comportamiento adaptativo deseado.

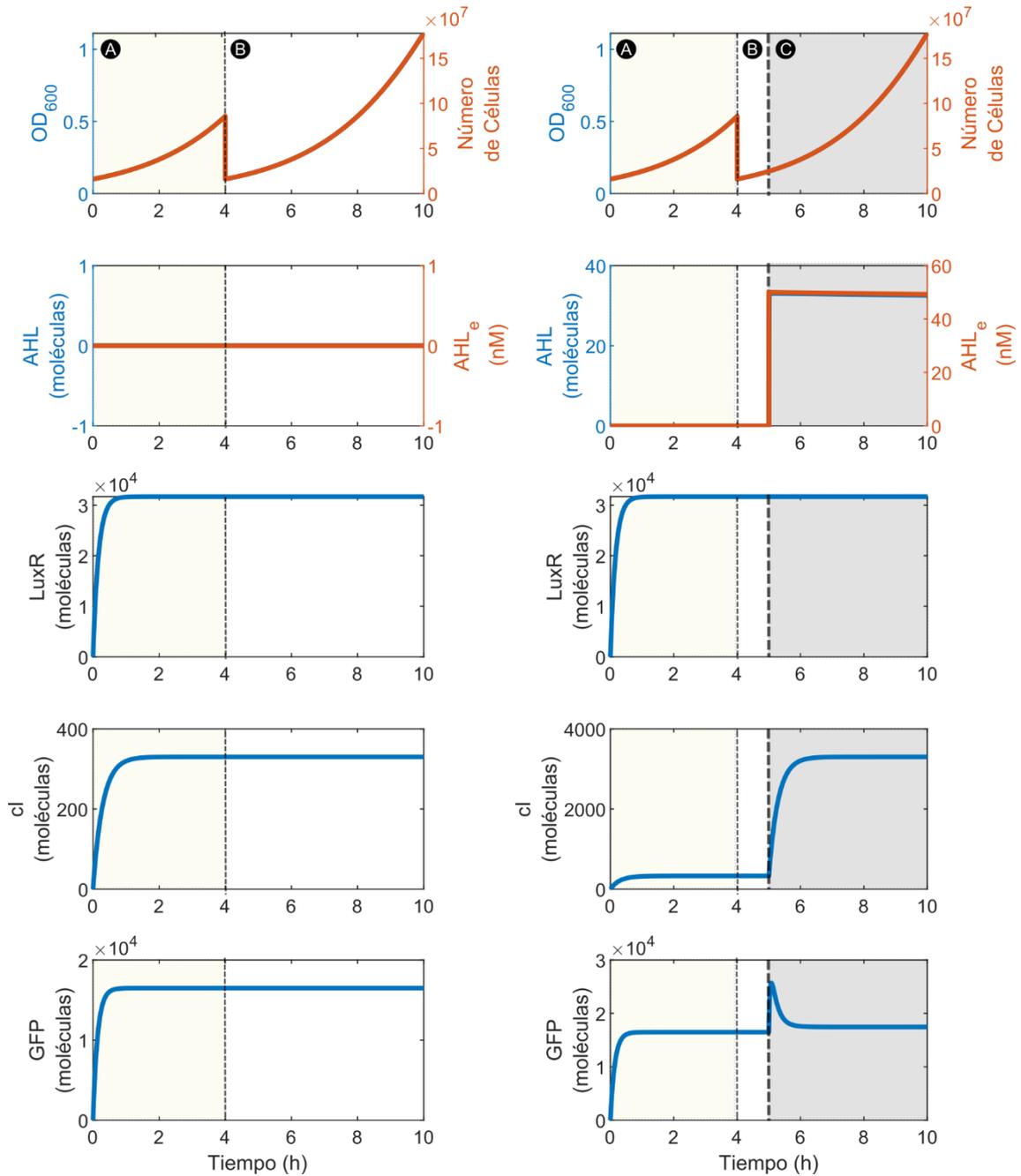


Figura 4.1. Resultados de las simulaciones de las especies del circuito: N (número de células), AHL y AHL_e, LuxR, cl y GFP. Los paneles de la izquierda simulan un experimento en el que no se ha realizado inducción. Los paneles de la derecha simulan una inducción de 50 nM de AHL a tiempo = 5 horas.

Con el objetivo de analizar el comportamiento del circuito ante cambios de la inducción con AHL_e, se han realizado simulaciones cambiando el nivel de inducción con los siguientes valores de 1, 5 y 30 nM de AHL_e media hora después del cultivo en placa de las células (tiempo de inducción = 30 min). En la Figura 4.2 se muestran la variación de la cantidad de moléculas de GFP con respecto al tiempo. Se puede observar cómo el circuito genético responde al estímulo de la inducción de AHL, generando un pico de expresión de GFP. Sin embargo, la expresión de esta proteína no cambia para distintos valores de AHL.

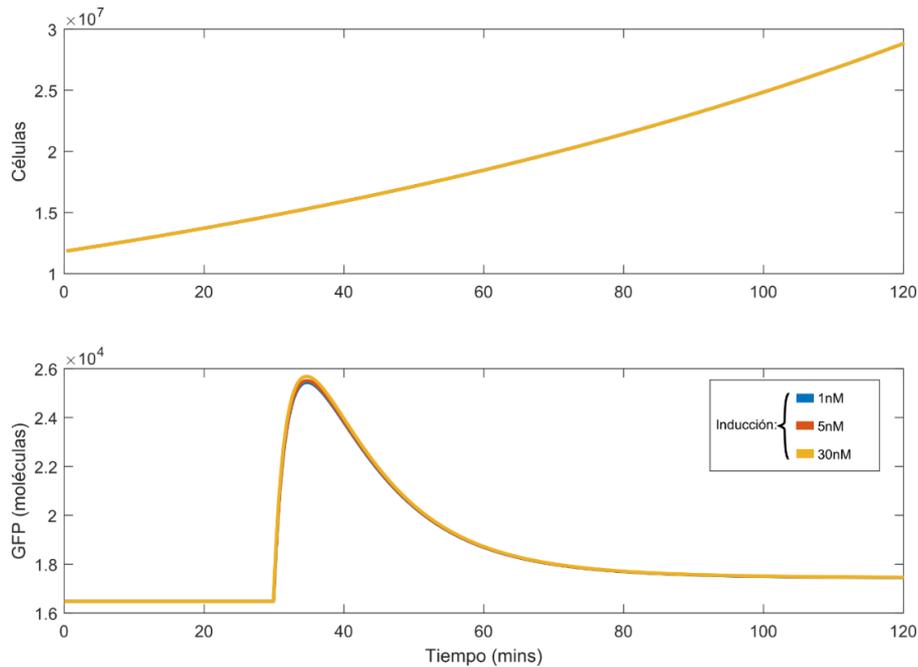


Figura 4.2. Simulaciones del circuito genético implementado para distintas inducciones. El panel superior representa el crecimiento celular y el panel inferior, la expresión de GFP.

Asimismo, se realizaron distintas simulaciones para analizar el comportamiento del circuito con distintos valores de tasa de traducción p_G : 40, 70 y 150 min^{-1} . En la Figura 4.3, se representa la variación de la concentración de GFP frente al tiempo. Comparando estas simulaciones, vemos que cuanto menor es p_G , menos moléculas de proteína se producen (la curva comienza en valores más bajos y la altura del pico también es menor). En cambio, a medida que p_G aumenta, también lo hace la cantidad de proteína producida. Sin embargo, el aumento de la sensibilidad (aumento del pico) lleva a una disminución de la precisión (un aumento del error entre el valor final y el inicial). Por tanto, podemos comprobar que sensibilidad y precisión son mutuamente excluyentes.

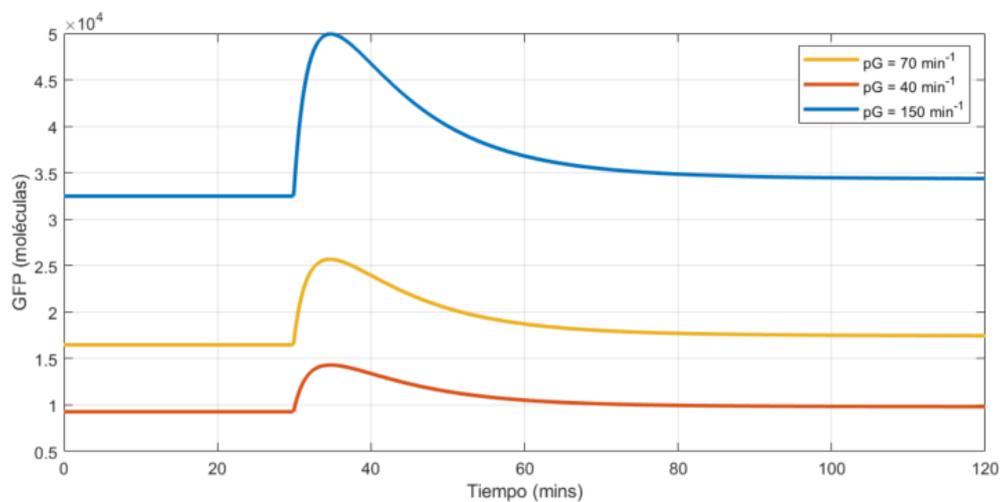


Figura 4.3. Simulaciones del circuito genético implementado con distintos valores de la tasa de traducción de GFP. El panel superior representa el crecimiento celular y el panel inferior, la expresión de GFP.

4.2. Construcción

Debido a la declaración del Estado de Alarma (14 de marzo de 2020) por la pandemia del coronavirus (COVID-19), cerró el Laboratorio de Biología Sintética y Control de Biosistemas (SB2CL), donde se estaba llevando a cabo el proyecto. Por tanto, no se pudo construir el circuito genético en cuestión ni se pudieron obtener datos experimentales propios.

Para poder avanzar a las siguientes fases del ciclo DBTL, se utilizaron datos experimentales de una versión anterior del circuito I1-FFL, que están representados en la Figura 4.4. Estos datos fueron obtenidos previamente en este mismo laboratorio, pero todavía no habían sido analizados. Las partes básicas con las que se construyó este circuito se recogen en la Tabla 4.3. En este caso, se utilizó una metodología diferente al GoldenBraid para el ensamblado de las partes y por este motivo, la fuente es distinta a las partes de la Tabla 3.2.

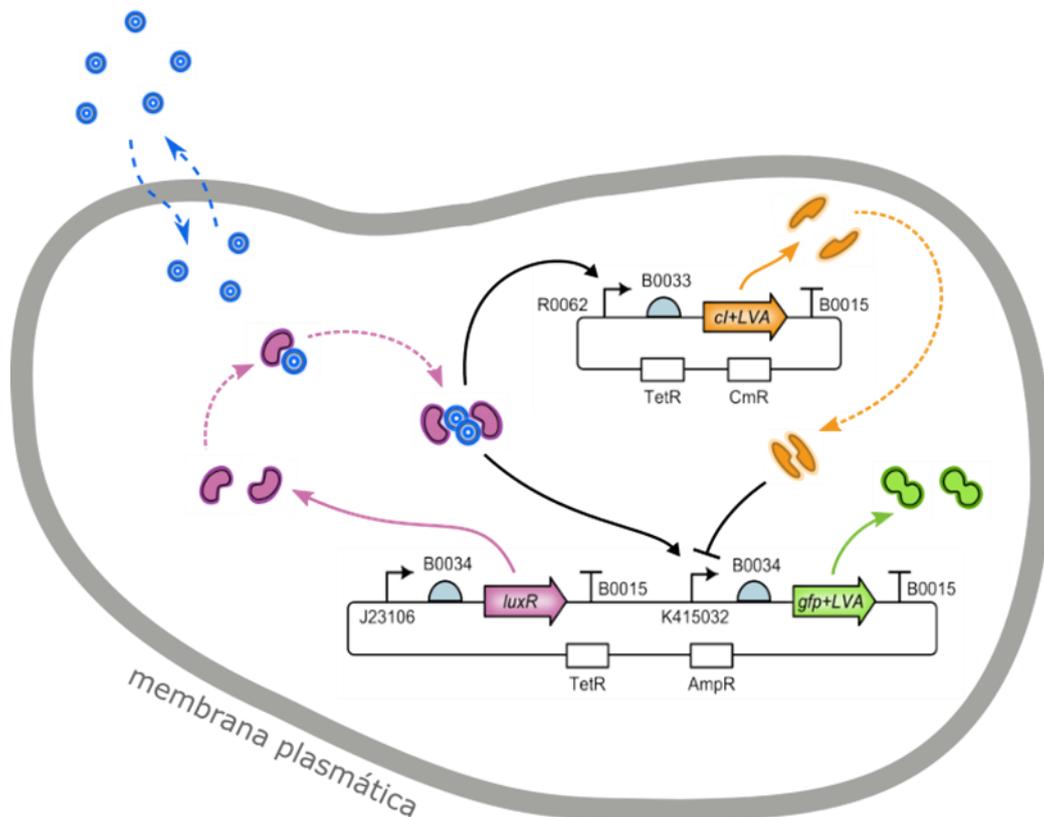


Figura 4.4. Esquema del circuito genético del que se han extraído los datos experimentales, implementado en una célula de *E. coli*.

Tabla 4.3. Partes básicas que componen el circuito genético del que se han extraído los datos experimentales.

Nombre	Tipo	Descripción	Fuente
P	Promotor	Promotor constitutivo de <i>luxR</i>	BBa_J23106
P_{lux}	Promotor	Promotor inducible del gen <i>cl</i>	BBa_R0062
P_{lux/cl}	Promotor	Promotor híbrido del gen <i>gfp</i>	BBa_K415032
	RBS	RBS fuerte	BBa_B0034
	RBS	RBS de fuerza media	BBa_B0033
luxR	CDS	Secuencia codificante de LuxR	BBa_C0062
cl	CDS	Secuencia codificante de <i>cl</i>	BBa_K327018
gfp	CDS	Secuencia codificante de GFP	BBa_K082003
	Terminador	Terminador transcripcional doble	BBa_B0015

4.3. Comprobación experimental

Debido a la situación relacionada con la pandemia de la COVID-19, no se pudieron obtener datos experimentales propios. Por tanto, este apartado se realizó con datos tomados previamente en el laboratorio de una versión anterior del circuito genético. La manipulación y calibración de los datos fue realizada por mí para este trabajo.

A partir de las curvas de calibración de fluorescencia y absorbancia (Figura 3.3), se calibraron los datos experimentales en cuestión. En la Figura 4.5 se representan los datos antes de su procesamiento (gráficas superiores) y una vez convertidos a sus respectivas unidades estándar: partículas para absorbancia y MEFL/partícula para fluorescencia (gráficas inferiores). Así, los datos experimentales quedan preparados para ser comparados con las simulaciones computacionales.

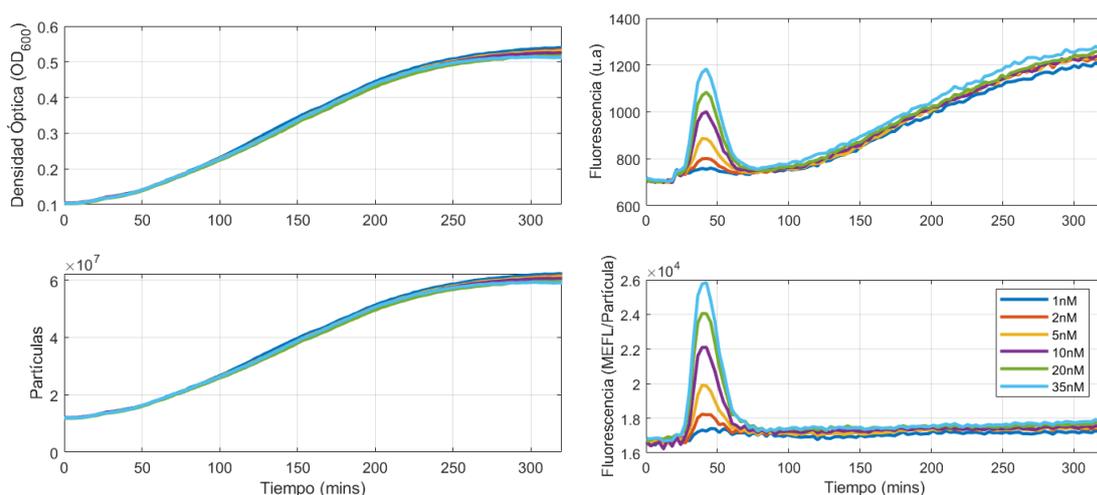


Figura 4.5. Calibración de los datos experimentales. Las gráficas superiores representan las medidas de absorbancia (derecha) y fluorescencia (izquierda) sin procesar. La absorbancia viene dada en OD_{600} y la fluorescencia, en unidades arbitrarias (u.a). Las gráficas inferiores muestran estos datos en unidades estándar: partículas para la absorbancia y MEFL/partícula para la fluorescencia. Cada color representa una inducción distinta de AHL.

4.4. Aprendizaje

A continuación, pasamos a la etapa de aprendizaje y mejora del modelo matemático diseñado en la primera fase del ciclo a partir de los datos experimentales calibrados obtenidos en la etapa anterior. Para ello procedemos con el protocolo comentado en la sección 3.4. *Aprendizaje*, obteniendo los valores de los parámetros del modelo que minimizan el error entre los datos experimentales y los obtenidos a partir de las simulaciones para las mismas condiciones.

De las seis condiciones experimentales, se escogen tres para realizar la estimación de parámetros. Luego, se validan los resultados comparando las otras tres inducciones con los resultados de las simulaciones cuando se induce con los correspondientes valores de AHL. Las tres inducciones seleccionadas para efectuar la estimación de los parámetros del modelo son tres inducciones a distintas concentraciones de AHL: 1, 5 y 35 nM.

Los parámetros elegidos para ser estimados son los relacionados con partes que pueden ser modificadas en el laboratorio como RBS, promotor y degradación de las proteínas, junto con parámetros desconocidos de los promotores, como son sus constantes de disociación. Con esto se lleva a cabo la MOO, de la cual se obtienen los valores de los parámetros que minimizan el error entre el comportamiento del modelo *in silico* y su implementación en el laboratorio. Los valores obtenidos tras ejecutar el algoritmo spMODE se encuentran en la Tabla 4.4.

Tabla 4.4. Parámetros del modelo estimados a partir de los datos experimentales.

Parámetro	Descripción	Valor previo a la estimación	Valor estimado	Unidad
μ	Tasa de crecimiento específico	0.0077	0.0074	min^{-1}
p_R	Tasa de traducción	95.7	31.5	min^{-1}
p_{ci}	Tasa de traducción	9.6	8.1	min^{-1}
d_{ci}	Tasa de degradación de la proteína	0.05	0.03	min^{-1}
k_{dlux}	Constante de disociación del dímero (R.A) ₂ al promotor	500	2165	moléculas
k_{dlux}	Constante de disociación del dímero (R.A) ₂ al promotor	1000	4219	moléculas
p_G	Tasa de traducción	71.5	65.3	min^{-1}
d_G	Tasa de degradación de la proteína	0.1	0.1	min^{-1}
D	Tasa de difusión del AHL a través de la membrana	4	0.2	min^{-1}

Se puede observar que, si bien los órdenes de magnitud son similares, los valores utilizados en las simulaciones previas a la estimación no eran capaces de captar el comportamiento de respuesta con distintas amplitudes de pico a las inducciones con distintas concentraciones de AHL (Figura 4.2). Sin embargo, con los parámetros obtenidos después de la estimación, el modelo es capaz de responder diferencialmente a las distintas concentraciones de AHL (Figura 4.6). Finalmente, como validación de los parámetros obtenidos, se comparan los resultados de nuevas simulaciones utilizando inducciones distintas a las anteriores (en este caso las otras tres; es decir: 2, 10 y 20 nM) con los resultados experimentales correspondientes. Debido a que dichos datos experimentales no fueron utilizados en la estimación de parámetros, se trata de una validación independiente de los resultados de la optimización. Esta comparación se muestra también en la Figura 4.6.

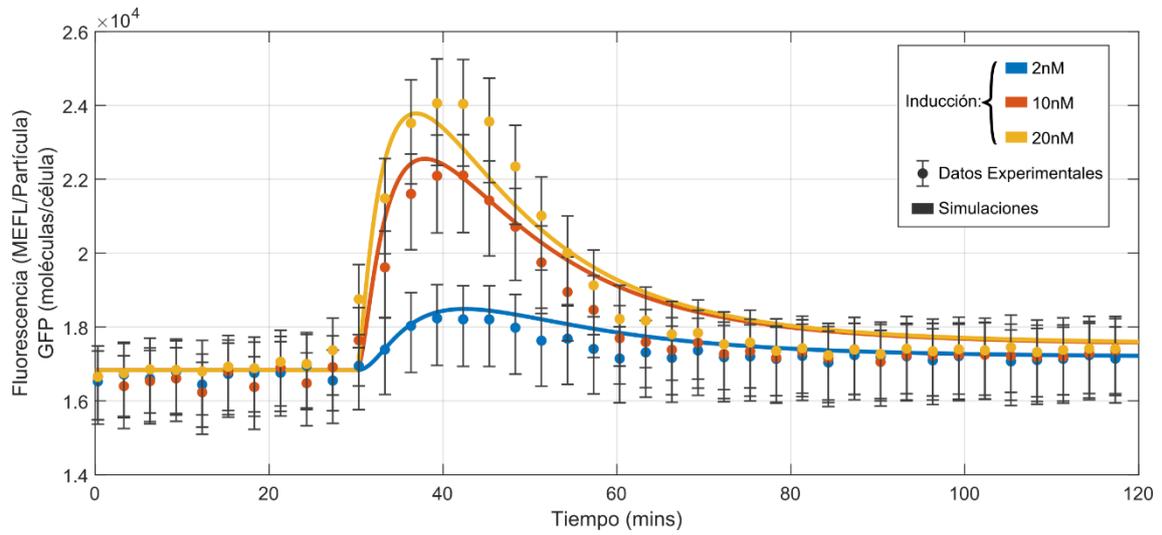


Figura 4.6. Validación de la estimación de parámetros. Las simulaciones computacionales realizadas con el modelo (líneas continuas) se ajustan bien a sus respectivos datos experimentales (puntos). La línea amarilla, que representa la máxima inducción con AHL (20 nM) produce más cantidad de GFP y, por tanto, la altura del pico es mayor. A medida que baja la concentración de AHL inducido, la cantidad de GFP desciende (línea roja (10 nM) y azul (2 nM)).

5. Conclusiones

En este trabajo, siguiendo el ciclo DBTL de la Biología Sintética, se ha conseguido diseñar un circuito I1-FFL con comportamiento adaptativo, modelando matemáticamente las ecuaciones ODE y ejecutando las correspondientes simulaciones *in silico*. Además, se ha planteado la metodología para la construcción de los plásmidos que componen el circuito y se han estimado los valores de los parámetros que generan el comportamiento esperado, utilizando para ello la MOO.

En particular, las conclusiones derivadas de este trabajo son:

- El modelo ODE diseñado a partir del circuito I1-FFL de 3 genes (*luxR*, *cl* y *gfp*), incluyendo la reducción del modelo mediante la Aproximación Quasi- Estacionaria (entre otras asunciones) y el número de plásmidos en la célula (C_N), resultó ser más preciso y detallado que los encontrados en la literatura a la hora de simular el comportamiento real del circuito.
- La simulación computacional demuestra que el modelo diseñado genera el comportamiento adaptativo deseado.
- El sistema de ensamblaje GoldenBraid es eficaz para construir el circuito I1-FFL utilizando las partes biológicas del repositorio del laboratorio SB2CL.
- La calibración y estandarización de los datos son adecuadas para poder compararlos con los resultados obtenidos al simular el modelo a diferentes concentraciones de inductor AHL.
- El enfoque de optimización multiobjetivo (MOO) fue exitoso para estimar *in situ* los valores de los parámetros que generan un comportamiento del modelo coherente con datos experimentales.

Así pues, se demuestra que el ciclo DBTL es un método efectivo en la creación de circuitos genéticos sintéticos con comportamientos concretos. De hecho, actualmente muchas industrias utilizan ya los principios de la Biología Sintética en la producción a nivel comercial como la industria alimentaria, en la que se comercializa ingredientes de sabores cítricos basados en levaduras sintéticas productoras de terpeno (Beekwilder et al., 2014). Los biocombustibles y biopolímeros son otro ámbito importante de aplicación. La compañía Solazyme, por ejemplo, se encarga de producir biodiesel basado en microalgas sintéticas.

Aunque todavía hay que hacer frente a problemas y retos, como los planteados en este trabajo, es innegable que la Biología Sintética es una disciplina con gran potencial la industria biotecnológica.

6. Referencias bibliográficas

- ALARCON-RUIZ, I.; ANDREU, R.; VIGNONI, A.; BOADA, Y.; PICÓ, J. (2020). *Part characterization workflow using calibrated measurements, guaranteed simulations and a multiobjective optimization approach*. IWDBA.
- ALON, U.; SURETTE, M. G.; BARKAI, N. y LEIBLER, S. (1999). *Robustness in bacterial chemotaxis*. *Nature*, 397(6715), 168-171.
- ALON, U. (2007). *An Introduction to Systems Biology. Design Principles of Biological Circuits*. Champan and Hall/CRC.
- AMYRIS. (s. f.). *Amyris*. Visto el 24 de junio de 2020. <https://amyris.com/>
- BEAL, J.; HADDOCK-ANGELLI, T.; BALDWIN, G.; GERSHATER, M.; DWIJAYANTI, A.; STORCH, M.; DE MORA, K.; LIZARAZO, M.; RETTBERG, R. et al. (2018). *Quantification of bacterial fluorescence using independent calibrants*. *PLoS one* 13, 6
- BEAL, J.; OVERNEY, C.; ADLER, A.; YAMAN, F.; TIBERIO, L. y SAMINENI, M. (2019). *Tasbe flow analytics: A package for calibrated flow cytometry analysis*. *ACS synthetic biology* 8(7), 1524–1529
- BEEKWILDER, J.; VAN HOUWELINGEN, A.; CANKAR, K.; VAN DIJK, A. D. J.; DE JONG, R. M.; STOOPEN, G.; BOUWMEESTER, H.; ACHKAR, J.; SONKE, T. y BOSCH, D. (2014). Valencene synthase from the heartwood of Nootka cypress (*Callitropsis nootkatensis*) for biotechnological production of valencene. *Plant Biotechnology Journal*, 12(2), 174-182.
- BIOBRICK FOUNDATION. (2006). *Part registry [online]*. Visto el 20 de junio de 2020. <http://partsregistry.org/>
- BOADA ACOSTA, Y. F. (2018). *A systems engineering approach to model, tune and test synthetic gene circuits*. PhD thesis.
- BOADA, Y., VIGNONI, A., REYNOSO-MEZA, G., y PICÓ, J. (2016). *Parameter identification in synthetic biological circuits using multi-objective optimization*. *IFAC-PapersOnLine*, 49(26), 77-82.
- BOADA, Y.; VIGNONI, A.; ALARCON-RUIZ, I.; ANDREU-VILARROIG, C.; MONFORTLLORENS, R.; REQUENA, A. y PICÓ, J. (2019). *Characterization of gene circuits parts based on multi-objective optimization with standard calibrated measurement*. *ChemBioChem*. Accepted author manuscript
- BOADA, Y., VIGNONI, A., y PICÓ, J. (2020). *Multiobjective identification of a feedback synthetic gene circuit*. *IEEE Transactions on Control Systems Technology*, 28(1), 208-223.
- BOYER, LA.; LEE, TI.; COLE, MF.; JOHNSTONE, SE.; LEVINE, SS.; ZUCKER, JP.; GUENTHER, MG.; KUMAR, RM.; MURRAY, HL.; JENNER, RG. et al. (2005). *Core transcriptional regulatory circuitry in human embryonic stem cells*. *Cell*;122:947–956.
- BRIAT, C.; GUPTA, A. and KHAMMASH, M. (2018). *Antithetic proportional-integral feedback for reduced variance and improved control performance of stochastic reaction networks*. *Journal of The Royal Society Interface*, 15(143), 20180079.

- BURK, M. J.; VAN DIEN, S. J.; BURGARD, A. P. y NIU, W. (2011). *Compositions and methods for the biosynthesis of 1,4-butanediol and its precursors*. Estados Unidos.
- CHAO, R.; YUAN, Y. y ZHAO, H. (2014). *Recent advances in DNA assembly technologies*. *FEMS Yeast Research* 15.1, pp. 1–9.
- CHELLABOINA, V.; BHAT, S.J.; HADDAD, W.M. y BERNSTEIN, D.S. (2009). *Modeling and analysis of mass-action kinetics*. *IEEE Control Systems Magazine* 29.4, pp. 60–78.
- CHURCH, G. M.; ELOWITZ, M. B.; SMOLKE, C. D.; VOIGT, C. A. y WEISS, R. (2014). *Realizing the potential of synthetic biology*. *Nature Reviews Molecular Cell Biology*, 15(4), 289-294.
- COX, R. S.; MADSEN, C.; MCLAUGHLIN, J. A.; NGUYEN, et al. (2018). *Synthetic biology open language (SBOL) version 2. 2. 0*. *Journal of Integrative Bioinformatics*, 15(1).
- EICHENBERGER, P.; FUJITA, M.; JENSEN, ST.; CONLON, EM.; RUDNER, DZ.; WANG, ST.; FERGUSON, C.; HAGA, K.; SATO, T.; LIU, JS.; LOSICK, R. (2004). *The program of gene transcription for a single differentiating cell type during sporulation in Bacillus subtilis*. *PLoS Biol*;2:e328.
- EICHENBERGER, M.; LEHKA, B. J.; FOLLY, C.; FISCHER, D.; MARTENS, S.; SIMÓN, E. y NAESBY, M. (2017). *Metabolic engineering of Saccharomyces cerevisiae for de novo production of dihydrochalcones with known antioxidant, antidiabetic, and sweet tasting properties*. *Metabolic Engineering*, 39, 80-89.
- ELLIS, T.; ADIE, T. y BALDWIN G.S. (2011). *DNA assembly for synthetic biology: from parts to pathways and beyond*. *Integrative Biology* 3.2, pp. 109–118.
- ERASynBio. (2014) *Next steps for european synthetic biology: a strategic vision from erasynbio*. Report, ERASynBio.
- GENOMATICA PRODUCTS. (s. f.). *Genomatica*. Visto el 24 de junio de 2020. <https://www.genomatica.com/products/>
- GOENTORO, L.; SHOVAL, O.; KIRSCHNER, M. W. y ALON, U. (2009). *The incoherent feedforward loop can provide fold-change detection in gene regulation*. *Molecular Cell*, 36(5), 894-899.
- GONZÁLEZ BOSCA, A. (2014). *Design and Characterization of a Robust Incoherent Feedforward Synthetic Genetic Circuit*. Tesis final de grado.
- ISOBIONICS PRODUCTS (s. f.). *Isobionics*. Visto el 24 de junio de 2020. <http://www.isobionics.com/index-Products.html>
- JULLESSON, D.; DAVID, F.; PFLEGER, B. y NIELSEN, J. (2015). *Impact of Synthetic Biology and Metabolic Engineering on Industrial Production of Fine Chemicals*. *Biotechnology Advances*, 33(7), 1395-1402.
- KEASLING, J. D. (2010). *Manufacturing molecules through metabolic engineering*. *Science*, 330:1355-1358.
- KOJIMA, R. y FUSSENEGGER, M. (2019). *Synthetic biology: Engineering mammalian cells to control cell-to-cell communication at will*. *ChemBioChem* 20(8), 994–1002.
- LEE, TI.; RINALDI, NJ.; ROBERT, F.; ODOM, DT.; BAR-JOSEPH, Z.; GERBER, GK.; HANNETT, NM.; HARBISON, CT.; THOMPSON, CM.; SIMON, I. et al. (2002). *Transcriptional regulatory networks in Saccharomyces cerevisiae*. *Science*;298:799–804

- LILACCI, G. y KHAMMASH, M. (2010). *Parameter Estimation and Model Selection in Computational Biology*. In: PLOS Computational Biology 6.3, pp. 1–17.
- MA, W.; TRUSINA, A.; EL-SAMAD, H.; LIM, WA.; TANG, C. (2009). *Defining network topologies that can achieve biochemical adaptation*. Cell; 138:760–73.
- MACÍA, J. y SOLÉ, R. (2011). *Presente y futuro de la Biología Sintética*. LYCHNOS, Cuadernos de la Fundación General CSIC, Nº 5.
- MARTIN, Y.; PAGE, M.; BLANCHET, C. y DE JONG, H. *Wellinverter: a web application for the analysis of fluorescent reporter gene data*. BMC bioinformatics 20(1), 309.
- MANGAN, S.; ITZKOVITZ, S.; ZASLAVER, A.; ALON, U. (2006). *The incoherent feed-forward loop accelerates the response-time of the gal system of E. coli*. J Mol Biol;356:1073–1081.
- MARLER, R. T. y ARORA, J. S. (2004). *Survey of multiobjective optimization methods for engineering*. Structural and multidisciplinary optimization, 26:369 - 395.
- MILO, R.; SHEN-ORR, S.; ITZKOVITZ, S.; KASHTAN, N.; CHKLOVSKII, D. y ALON, U. (2002). *Network motifs: simple building blocks of complex networks*. Science;298:824–827.
- MILO, R., JORGENSEN, P., MORAN, U., WEBER, G., y SPRINGER, M. (2010). *BioNumbers—The database of key numbers in molecular and cell biology*. Nucleic Acids Research, 38(suppl_1), D750-D753.
- MORADALI, M. F. y REHM, B. H. A. (2020). *Bacterial biopolymers: From pathogenesis to advanced materials*. Nature Reviews Microbiology, 18(4), 195-210.
- ONEWORLD HEALTH. (s. f.). *OneWorld Health*. Visto el 24 de junio de 2020. <https://www.oneworldhealth.com/>
- PADDON, C. J., & KEASLING, J. D. (2014). *Semi-synthetic artemisinin: A model for the use of synthetic biology in pharmaceutical development*. Nature Reviews Microbiology, 12(5), 355-367.
- PICÓ, J.; VIGNONI, A.; PICÓ-MARCO, E. y BOADA, Y. (2015). *Modelado de sistemas bioquímicos: De la Ley de Acción de Masas a la Aproximación Lineal del Ruido*. Revista Iberoamericana de Automática e Informática Industrial RIAI 12.3, pp. 241–252.
- REEVES, G. T. (2019). *The engineering principles of combining a transcriptional incoherent feedforward loop with negative feedback*. Journal of Biological Engineering, 13(1), 62.
- REQUENA, A. (2019). *Design and optimization of a genetic circuit for the internal control of the population density of an E. coli culture*. Tesis final de grado.
- REYNOSO-MEZA, G.; GARCÍA-NIETO, S; SANCHÍS, J. y BLASCO, F.X. (2013). *Controller Tuning by Means of Multi-Objective Optimization Algorithms: A Global Tuning Framework*. IEEE Transactions on Control Systems Technology 21.2, pp. 445–458.
- ROBINSON, C. M.; FEDOREC, A. J. H.; YAN WEN, K. y BARNES, C. P. *Flopr: An open source software package for calibration and normalization of plate reader and flow cytometry data*. bioRxiv (2020).
- SARRION-PERDIGONES, A.; FALCONI, E.E.; ZANDALINAS, S.I.; JUAREZ, P.; FERNÁNDEZDEL-CARMEN, A.; GRANDSELL, A. y ORZAEZ, D. (2011). *GoldenBraid: An Iterative Cloning System for Standardized Assembly of Reusable Genetic Modules*. PLOS ONE 6.7, pp. 1–11.

SCHWILLE, P.; SPATZ, J.; LANDFESTER, K.; BODENSCHATZ, E.; HERMINGHAUS, S.; SOURJIK, V.; ERB, T. J.; BASTIAENS, P.; LIPOWSKY, R.; HYMAN, A.; DABROCK, P.; BARET, J.-C.; VIDAKOVIC-KOCH, T.; BIELING, P.; DIMOVA, R.; MUTSCHLER, H.; ROBINSON, T.; TANG, T.-Y. D.; WEGNER, S. y SUNDMACHER, K. (2018). *Maxsynbio: Avenues towards creating cells from the bottom up*. *Angewandte Chemie International Edition*, 57(41), 13382-13392.

SEGALL-SHAPIRO, T. H.; SONTAG, E. D. y VOIGT, C. A. (2018). *Engineered promoters enable constant gene expression at any copy number in bacteria*. *Nature Biotechnology*, 36(4), 352–358.

SHEN-ORR, S. S.; MILO, R.; MANGAN, S. y ALON, U. (2002). *Network motifs in the transcriptional regulation network of E. coli*. *Nature Genetics*, 31(1), 64-68.

SOLAZYME. (s. f.). *Solazyme*. Visto el 24 de junio de 2020. <https://www.solazyme.com/>

SOURJIK, V. y BERG, H. C. (2002). *Receptor sensitivity in bacterial chemotaxis*. *Proceedings of the National Academy of Sciences*, 99(1), 123-127.

TOMAZOU, M. y STAN, G.B. (2018). *Portable gene expression guaranteed*. *Nature Biotechnology*, 36(4), 313-314.

TRABELSI, H., KOCH, M., y FAULON, J.-L. (2018). *Building a minimal and generalizable model of transcription factor-based biosensors: Showcasing flavonoids*. *Biotechnology and Bioengineering*, 115, 2292–2304.

TYSON, J. J.; CHEN, K. C. y NOVAK, B. (2003). *Sniffers, buzzers, toggles and blinkers: Dynamics of regulatory and signaling pathways in the cell*. *Current Opinion in Cell Biology*, 15(2), 221-231. VALERO, D. (2018). *El proyecto europeo IBISBA 1.0 hace que la biotecnología industrial abra sus puertas a la investigación*. *Inspira Biotech*. Visto el 29 de junio de 2020. <https://inspirabiotech.com/2018/06/07/el-proyecto-europeo-ibisba-1-0-hace-que-la-biotecnologia-industrial-abra-sus-puertas-a-la-investigacion/>

VILLAVERDE, A.F. y BANGA, J.R. (2014). *Reverse engineering and identification in systems biology: strategies, perspectives and challenges*. *Journal of The Royal Society Interface* 11.91, p. 20130505.

WAY, J. C.; COLLINS, J. J.; KEASLING, J. D. y SILVER, P. A. (2014). *Integrating biological redesign: Where synthetic biology came from and where it needs to go*. *Cell*, 157(1), 151-161.

WENINGER, A.; KILLINGER, M. y VOGL, T. (2016). *Key Methods for Synthetic Biology: Genome Engineering and DNA Assembly*. In: *Synthetic Biology*. Ed. by Anton Glieder et al. Cham: Springer International Publishing, pp. 101–141.

ZAGARIS, A.; KAPER, H. G. y KAPER, T. J. (2004). *Analysis of the computational singular perturbation reduction method for chemical kinetics*. *Journal of Nonlinear Science*, 14(1):59–91, 1.

7. Anexo I: Artículos de revista y de congresos derivados de este trabajo

Part characterization workflow using calibrated measurements, guaranteed simulations and a multiobjective optimization approach

Iván Alarcon-Ruiz, Raquel Andreu, Alejandro Vignoni, Yadira Boada, Jesús Picó
ivalrui1@upv.es, raanvi@upv.es, alvig2@upv.es, yaboa@upv.es, jpico@upv.es
Synthetic Biology and Biosystems Control Lab
Instituto de Automática e Informática Industrial
Universitat Politècnica de València
València, Spain

1 BACKGROUND

Synthetic biology aims at the targeted design or redesign and construction of new biological and bio-based parts, devices, and systems to perform desired functions [11, 13]. Not only is going up in this hierarchy (DNA, part, device, and system) the final objective of synthetic biology but also its main challenge [3]. To successfully accomplish it, engineering principles and methodologies are to be used. The design-build-test-learn (DBTL) cycle is the common paradigm used in any engineering discipline where the design is made from the bottom by combining basic biological parts into devices and these into systems [8]. Essential for the success of this inherently modular approach of bottom-up synthetic biology is the need of starting from well-characterized parts [9]. Currently, there is still a gap between the possibility of designing a system and its real implementation in the lab. This gap can be partially attributed to both the lack of repeatable and standardized measurements, and the absence of well-characterized biological parts [17]. On the one hand, calibration of equipment and standardization of units is a challenge. We cannot compare two systems in terms of measurements if they have not been created with the same methodology or they can not be taken into the same domain [2]. Luckily, there are many works from this community dealing with this issue [4, 6, 14, 16].

On the other hand, the characterization of biological parts understood as estimation of its model parameters is still a bottleneck in Synthetic Biology. Though the isolated characterization of the steady-state response of basic parts (e.g. promoters) can be performed by means of relatively simple experiments, their behavior when integrated within more complex circuits will be affected by the circuit context [15]. Therefore, *in situ* estimation of the parameters of each basic part in the circuit is carried out to account for context. However, estimating the parameters associated to biological parts embedded in a nonlinear dynamic model of a synthetic gene circuit remains a challenging inverse problem. Non-convexity, ill-conditioning caused by over-parametrization,

experimental measurement errors, data scarcity, and uncertainty are the main difficulties [12]. The multiobjective optimization design (MOOD) framework to perform model parameter estimation has been successfully used for the optimal design of gene networks [5], or closed-loop genetic circuit identification [7], as it allows to address problems often found in synthetic gene networks that are difficult to tackle using single or weighted-objective optimization approaches.

2 METHODS

In [6] we propose a methodology combining measurement calibration with MOOD that is used to characterize the RBS strength of several genetic constructs. In this work, we propose an extended approach (see Figure 1) to solve the problem of biological parts characterization and fill the gap existing in the DBTL cycle based on multiobjective objective optimization that allows us to include uncertainty in the estimation and also to simultaneously include different scenarios and a set or library of genetic circuits or devices (which may have common parts and different parts).

The starting point (Figure 1.-1) is comprised of both the mathematical model and the set of biological parts to characterize (the model has to be one of those parts). From there we need to perform experiments to measure the desired magnitude (fluorescence of the reporter) and to calibrate the measurements (Figure 1.0) to obtain MEFL/Particle [1, 6]. Experimental data in this unit can be compared with mathematical models in molecules/cell. Also, it is necessary to define initial intervals for the model parameters to be estimated (and values for the parameters we don't want to estimate).

Once the prerequisites and the data preprocessing is performed, the actual methodology can be applied to characterize the selected parts. In the first step (Figure 1.1), we need to define the cost function for the optimization problem. In this case, we use a model for guaranteed simulations [10]

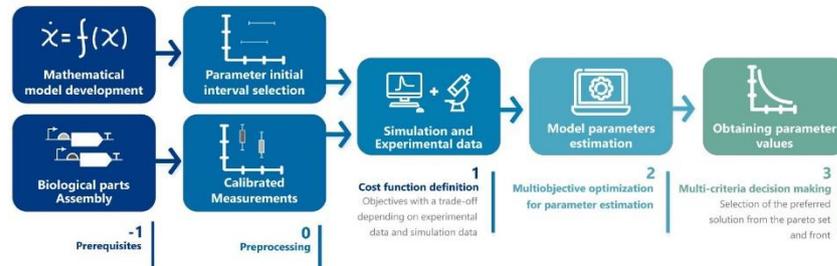


Figure 1: Schematic of the methodology. Prerequisites, data preprocessing, cost function definition, multiobjective optimization and multi-criteria decision making steps.

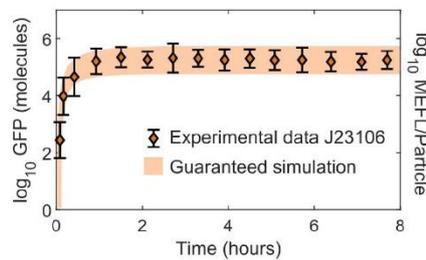


Figure 2: Example of simulation and experimental data to be used with the proposed methodology.

to obtain an interval solution (Figure 2) and in the objective function, we compare this interval solution with the calibrated measurements (including the measurement uncertainty). Then, we perform the optimization in the second step (Figure 1.2) to minimize the discrepancy between the experimental data and the simulation. The solution of the optimization is a pareto front (with the values of the objectives functions for the solutions) and the pareto set with the values of the parameters corresponding to those solutions. The last step (Figure 1.3) is to select a desired solution from the pareto front and set corresponding to the part we wanted to characterize.

REFERENCES

- [1] BEAL, J., FARNY, N. G., HADDOCK-ANGELLI, T., SELVARAJAH, V., BALDWIN, G. S., BUCKLEY-TAYLOR, R., GERSHATER, M., KIGA, D., MARKEN, J., SANCHANIA, V., SISON, A., WORKMAN, C. T., AND . Robust estimation of bacterial cell count from optical density. *bioRxiv* (2019).
- [2] BEAL, J., HADDOCK-ANGELLI, T., BALDWIN, G., GERSHATER, M., DWIJAYANTI, A., STORCH, M., DE MORA, K., LIZARAZO, M., RETTBERG, R., ET AL. Quantification of bacterial fluorescence using independent calibrants. *PLoS one* 13, 6 (2018), e0199432.
- [3] BEAL, J., HADDOCK-ANGELLI, T., FARNY, N., AND RETTBERG, R. Time to get serious about measurement in synthetic biology. *Trends in biotechnology* 36, 9 (2018), 869–871.
- [4] BEAL, J., OVERNEY, C., ADLER, A., YAMAN, F., TIBERIO, L., AND SAMINENI, M. Tasbe flow analytics: A package for calibrated flow cytometry analysis. *ACS synthetic biology* 8, 7 (2019), 1524–1529.
- [5] BOADA, Y., REYNOSO-MEZA, G., VIGNONI, A., AND PICÓ, J. Multi-objective optimization framework to obtain model-based guidelines for tuning biological synthetic devices: an adaptive network case. *BMC Systems Biology* 10 (2016), 27.
- [6] BOADA, Y., VIGNONI, A., ALARCON-RUIZ, I., ANDREU-VILARROIG, C., MONFORT-LLORENS, R., REQUENA, A., AND PICÓ, J. Characterization of gene circuit parts based on multiobjective optimization by using standard calibrated measurements. *ChemBioChem* 20, 20 (2019), 2653–2665.
- [7] BOADA, Y., VIGNONI, A., AND PICÓ, J. Multiobjective identification of a feedback synthetic gene circuit. *IEEE Transactions on Control Systems Technology*, (early access), (2019).
- [8] CANTON, B., LABNO, A., AND ENDY, D. Refinement and standardization of synthetic biological parts and devices. *Nature biotechnology* 26, 7 (2008), 787–793.
- [9] CHURCH, G. M., ELOWITZ, M. B., SMOLKE, C. D., VOIGT, C. A., AND WEISS, R. Realizing the potential of synthetic biology. *Nature Reviews. Molecular Cell Biology* 15, 4 (2014), 289–294.
- [10] DE PEREDA, D., ROMERO-VIVO, S., RICARTE, B., AND BONDIA, J. Guaranteed computation methods for compartmental in-series models under uncertainty. *Computers & Mathematics with Applications* 66, 9 (2013), 1595–1605.
- [11] ERASYNBIO. Next steps for european synthetic biology: a strategic vision from erasynbio. Report, ERASynBio, 2014.
- [12] GÁBOR, A., AND BANGA, J. R. Robust and efficient parameter estimation in dynamic models of biological systems. *BMC systems biology* 9, 1 (2015), 74.
- [13] KOJIMA, R., AND FUSSENEGGER, M. Synthetic biology: Engineering mammalian cells to control cell-to-cell communication at will. *ChemBioChem* 20, 8 (2019), 994–1002.
- [14] MARTIN, Y., PAGE, M., BLANCHET, C., AND DE JONG, H. Wellinverter: a web application for the analysis of fluorescent reporter gene data. *BMC bioinformatics* 20, 1 (2019), 309.
- [15] MCBRIDE, C., SHAH, R., AND DEL VECCHIO, D. The effect of loads in molecular communications. *Proceedings of the IEEE* (2019), 1–18.
- [16] ROBINSON, C. M., FEDOREC, A. J. H., YAN WEN, K., AND BARNES, C. P. Flopr: An open source software package for calibration and normalization of plate reader and flow cytometry data. *bioRxiv* (2020).
- [17] WAY, J. C., COLLINS, J. J., KEASLING, J. D., AND SILVER, P. A. Integrating biological redesign: where synthetic biology came from and where it needs to go. *Cell* 157, 1 (2014), 151–161.