

# UNIVERSITAT POLITÈCNICA DE VALÈNCIA

ESCOLA TÈCNICA SUPERIOR D'ENGINYERIA  
AGRONÒMICA I DEL MEDI NATURAL



***Integración de datos proteómicos y  
metabolómicos en modelos metabólicos a  
escala genómica y su aplicación a un  
modelo de la cianobacteria Synechocystis  
sp. PCC 6803.***

TRABAJO FIN DE GRADO/MÁSTER EN BIOTECNOLOGÍA

ALUMNO/A: MARINA PÉREZ NAVEIRA

TUTOR/A: JAVIER FERMÍN URCHUEGUIA SCHOLZEL

COTUTOR/A: MARIA SIURANA PAULA

*Curso Académico: 2014-2015*

VALENCIA, SEPTIEMBRE DE 2015

Creative Commons



**Alumno/a:** D. /Dña. Marina Pérez Naveira

**Tutor Académico:** Prof. D. /Prof. Dña. Javier Fermín Urchueguia Scholzel

**Cotutor/a:** D. /Dña. Maria Siurana Paula

**Valencia, Septiembre de 2015**

**Título:** Integración de datos proteómicos y metabolómicos en modelos metabólicos a escala genómica y su aplicación a un modelo de la cianobacteria *Synechocystis* sp. PCC 6803

**Resumen:** Un modelo metabólico a escala genómica es una reconstrucción informática del metabolismo del organismo de interés que, junto con la utilización de un método analítico, permiten simular parcialmente el comportamiento metabólico de dicho organismo. Como resultado de estas simulaciones se obtienen distribuciones de flujos en las reacciones del organismo.

Las cianobacterias son importantes al poder utilizarse como biofactorías debido a su capacidad para realizar la fotosíntesis. Un conocimiento detallado del funcionamiento de las mismas permite poder optimizar su funcionamiento en términos de aumento de la producción de compuestos de alto valor añadido.

A partir del modelo de la cianobacteria *Synechocystis* sp. PCC 6803 es posible analizar el comportamiento general de las cianobacterias y optimizar su rendimiento de producción de compuestos de interés.

El objetivo de este proyecto es aplicar un algoritmo desarrollado por Yizhak et al. (2010), que permite la utilización de datos proteómicos y metabolómicos para precisar las distribuciones de flujos resultantes, en la cianobacteria *Synechocystis*. El IOMA ha sido el primer algoritmo que permite la incorporación de este tipo de datos 'ómicos', permitiendo obtener resultados más realistas al disponer de mayor información sobre su comportamiento.

En su artículo, Yizhak et al. (2010) ponen a prueba su algoritmo con organismos de menor complejidad y ampliamente estudiados, por lo que este proyecto resulta muy útil para conocer las ventajas de este algoritmo y sus limitaciones en comparación con un FBA. Además se analizará la sensibilidad respecto de los datos iniciales.

**Palabras clave:**

- Modelado metabólico
- Biología de sistemas
- Análisis de balance de flujos (FBA)
- Reacción de biomasa
- Reacciones de intercambio
- Asociación Gen - proteína - reacción (GPR)
- Modelo metabólico a escala genómica.
- Genome-scale network reconstruction (GENRE).
- Programación lineal (LP)
- Programación cuadrática (QP)

**Abstract:** A genome-scale metabolic model is an informatic assembly of the whole set of reactions that are present in the metabolism of an organism that, used in combination with an analytic model, enables to simulate partially its metabolic response. The result obtained from the simulations is a flux distribution through the reactions of the organism.

Cyanobacteria are important due to its ability to perform photosynthesis and be used as biofactories. A detailed knowledge of their behavior allows the optimization of its performance on the production of high-valuable compounds.

From the cyanobacteria model of *Synechocystis* sp.PCC 6803 is possible to analyze the general behavior of cyanobacteria and to optimize the yield of production of the compound of interest.

The main objective of this project is to apply an algorithm created by Yizhak et al. (2010) that allows the utilization of metabolomic and proteomic data, in *Synechocystis*. This algorithm is the first algorithm that allows the usage of this type of 'omic' data together in order to obtain more realistic results due to the availability of more information about its behavior.

In their article, Yizhak et al. (2010) use simpler and more studied organisms to prove the efficiency of their algorithm, and this project is useful to verify the efficiency of the algorithm in more complicated organisms. This project allows to know the advantages and drawbacks of this method and to study its accuracy.

**Keywords:**

- Metabolic modeling
- Systems biology
- Exchange reactions
- Flux-balance analysis (FBA)
- Gene–protein-reaction (GPR) association:
- Genome-scale model (GEM).
- Genome-scale network reconstruction (GENRE).
- Linear programming (LP)
- Quadratic programming(QP)

## **Agradecimientos**

Me gustaría agradecer en primer lugar a mis tutores por la oportunidad que me han dado de aprender una disciplina relativamente nueva, tan interesante y con tanto futuro, y por la confianza que me han dado durante el proyecto. En segundo lugar agradecer a los autores de los artículos utilizados para la obtención de los datos tanto metabolómicos como proteómicos, en especial al doctor Takashi Osanai, por su gran ayuda y gran disposición. También me gustaría agradecer a mi familia por su apoyo incondicional, por darme fuerzas cuando he flaqueado y animarme a seguir adelante. Por último, agradecer a mis amigos, a mis compañeros en este aprendizaje, no podría pensar en mejor compañía que vosotros para pasar estos 4 maravillosos años.

## Índice

1.	INTRODUCCION.....	1
1.1.	Presentación del trabajo .....	1
1.2.	Modelado de sistemas biológicos .....	2
1.2.1.	Enfoque 'top-down' .....	4
1.2.2.	Enfoque 'bottom-up' .....	5
1.2.3.	Modelos metabólicos a escala genómica .....	6
1.3.	Construcción de un modelo metabólico .....	7
1.3.	Cianobacterias.....	10
1.3.1.	Generalidades sobre las cianobacterias .....	10
1.3.2.	Phylum cianobacteria .....	10
1.3.4.	Importancia e información sobre <i>Synechocystis</i> sp.PCC 6803 .....	12
2.	Objetivos .....	13
3.	Materiales y métodos .....	14
3.1	Algoritmo.....	14
3.2.	Modelo .....	17
3.4.	Modelo de <i>Synechocystis</i> 6803.....	19
3.3.1.	Datos proteómicos y metabolómicos .....	19
3.3.2.	Datos cinéticos .....	20
3.3.3.	Restricciones introducidas .....	21
3.4.	Simulación FBA.....	21
4.	Resultados.....	22
4.1.	Simulación modelo juguete 2.....	22
4.2.	Simulación con modelo de <i>Synechocystis</i> 6803 .....	23
4.2.1.	Acumulación de glucógeno .....	24
4.2.2.	Ciclo del ácido cítrico .....	27
4.2.3	Ruta de la glicolisis .....	29
4.2.4.	Metabolismo del piruvato.....	32
4.2.5.	Ruta de las pentosas fosfato .....	34
4.2.6.	Síntesis de otros compuestos .....	36
5.	Discusión .....	37
6.	Conclusión.....	41

## Figuras

**Figura 1.** Pirámide que explica la complejidad jerárquica de los organismos, hecha por Oltvai y Barabasi (2010).

**Figura 2.** Explicación de los enfoques 'top-down' y 'bottom-up' en el artículo 'The nature of systems biology' de Bruggeman y Westerhoff (2006).

**Figura 3.** Representación gráfica de cómo funciona el FBA, definiendo primero el espacio de posibles soluciones del sistema y dando la solución más óptima para la función objetivo definida. Imagen extraída del trabajo de Orth et al., (2010).

**Figura 4.** Árbol filogenético creado por Swingley et al. a partir de 341 familias de proteínas.

**Figura 5.** Primera línea del programa en la que define que entradas necesita para funcionar.

**Figura 6.** Representación del modelo juguete 2

**Figura 7.** Resultado utilizando un algoritmo tipo FBA

**Figura 8.** Resultado obtenido utilizando el algoritmo IOMA.

**Figura 9.** Predicción de la salida de glucógeno al exterior utilizando los dos algoritmos en ambas condiciones.

**Figura 10.** Representación gráfica de las reacciones en las que interviene el glucógeno.

**Figura 11.** Flujos que arrojan los diferentes algoritmos para las reacciones en las que interviene el glucógeno.

**Figura 12.** Ciclo del ácido cítrico. En verde aparecen las enzimas descritas en *Synechocystis*. Imagen extraída de la base de datos KEGG.

**Figura 13.** Flujos que arrojan los diferentes algoritmos para las reacciones en las que interviene el ciclo del ácido cítrico.

**Figura 14.** Representación gráfica del ciclo del ácido cítrico.

**Figura 15.** Ruta de la glucólisis. En verde aparecen las enzimas descritas en *Synechocystis*. Imagen extraída de la base de datos KEGG.

**Figura 16.** Distribución de flujos de la ruta de la glucólisis tras 24 horas en condiciones de falta de nitrógeno.

**Figura 17.** Esquema de las reacciones escogidas para evaluar la ruta de la glicólisis.

**Figura 18.** Reacciones metabólicas que relacionan el piruvato, producto de la glicólisis, con otras rutas metabólicas como la síntesis de algunos aminoácidos o el ciclo de Krebs. Imagen extraída de KEGG.

**Figura 19.** Distribución de flujos de algunas de las reacciones relacionadas con el piruvato según el método de análisis utilizado, para las diferentes condiciones.

**Figura 20.** Esquema del metabolismo del piruvato.

**Figura 21.** Ruta completa de las pentosas fosfato.

**Figura 22.** Distribución de flujos en esta ruta en las 3 condiciones estudiadas.

**Figura 23.** Esquema que representa la ruta de las pentosas fosfato.

**Figura 24.** Histograma de la dispersión de los datos de flujo calculados utilizando el algoritmo IOMA.

**Figura 25.** Histograma de la dispersión de datos de flujos calculados utilizando el FBA.

## Tablas

**Tabla 1.** Lista de reacciones del modelo juguete 2. Las 9 primeras reacciones son las nueve reacciones del modelo juguete 1.

**Tabla 2.** Bases de datos utilizadas.

## Ecuaciones

**Ecuación 1.** Variación temporal de la concentración de un metabolito

**Ecuación 2.** Concentración del metabolito X en función del tiempo en un estado estacionario.

**Ecuación 3.** Ecuación que utilizan en el algoritmo desarrollado por Yizhak et al. (2010) para estimar los flujos metabólicos utilizando datos cuantitativos extraídos gracias a las estrategias ómicas.

**Ecuación 4.** Valor de saturación en la dirección 'forward'.

**Ecuación 5.** Valor de saturación en dirección 'backwards'.

**Ecuación 6.** Ecuación 3 sustituyendo los productorios de sustratos y productos por  $a^+$  y  $a^-$ , respectivamente.

**Ecuación 7.** Resultado de dividir la concentración del enzima por la concentración del enzima en el estado de referencia y de multiplicar  $k_{cat}^+$  y  $k_{cat}^-$  por la  $e^{ref}$ .

**Ecuación 8.** Define el estado estacionario, es decir que la suma de todos los flujos es igual a 0, para todas las condiciones j.

**Ecuación 9.** Representa las condiciones de capacidad de flujo y direccionalidad de la reacción.

**Ecuación 10.** Define los flujos cinéticos teóricos y les añade un error para permitirles un poco de libertad a los flujos, ya que si no sería imposible llegar a una solución factible y óptima. Este error representa la falta de concentraciones para alguno de los metabolitos de la reacción, la presencia de ruido en los datos tanto proteómicos como metabolómicos, y las simplificaciones hechas en la ecuación cinética de la velocidad.

**Ecuación 11.** Optimización del programa tratando de que se cumplan todas las condiciones anteriores minimizando la suma total de las varianzas de los errores para cada enzima en todas sus condiciones.

## Abreviaturas

RNA: Ácido Ribonucleico

ADN: Ácido desoxirribonucleico

PCA: Análisis de componentes principales

FBA: Análisis de balance de flujos

GPR: Asociación Gen - proteína - reacción

LP: Programación lineal

QP: Programación cuadrática

RuBisCO: Ribulosa-1,5-bifosfato carboxilasa

*Synechocystis 6803*: *Synechocystis sp. PCC 6803*

*E.Coli*: *Eschericia Coli*

RNA-seq : RNA sequencing

IOMA: Integrative omics metabolomic analysis

Acetyl-coA: acetyl coenzima-A

PRRP: fosforibosil pirofosfato



# 1. INTRODUCCION

## 1.1. Presentación del trabajo

La biología de sistemas pretende alcanzar reconstrucciones informáticas realistas de los organismos vivos sin añadir demasiada complejidad, permitiendo así simular el comportamiento de dicho organismo en cualquier tipo de condiciones. Para lograrlo no sólo es necesario mejorar las reconstrucciones de los organismos de estudio, si no también desarrollar métodos analíticos de las simulaciones. En la actualidad se están desarrollando diversos algoritmos que permiten aproximar el resultado de las simulaciones con la realidad, gracias a la incorporación de diversos datos 'ómicos'. Es tos datos que aportan información acerca de las concentraciones de diversas moléculas, para precisar las distribuciones de flujos resultantes de las simulaciones anteriormente mencionadas.

Actualmente encontramos diversos algoritmos que permiten la utilización de datos transcriptómicos, metabólicomicos y proteómicos individualmente (Agren et al., 2012; Blazier et al., 2012; Schmidt et al., 2013), y comienzan a surgir algoritmos que permiten la integración de dos de estos tipos de datos conjuntamente. A pesar de que los datos transcriptómicos son difícilmente utilizables cuantitativamente, se han desarrollado un mayor número de algoritmos para este tipo de datos. El primer y único algoritmo desarrollado hasta el momento que integra datos metabólicomicos y proteómicos es el algoritmo 'Integrative Omics Metabolomic Analysis' (en adelante llamado IOMA) desarrollado por Yizhak et al (2010).

En este proyecto se analiza la aplicación de este algoritmo en *Synechocystis* sp. PCC 6803 y comprobar su mejora de resultados en comparación con el FBA básico. Esto supone un gran reto ya que este algoritmo solo se ha utilizado en un modelo metabólico relativamente simple como es el de un eritrocito, o en un modelo metabólico a escala genómica de un organismo muy bien caracterizado, *E.coli*. *Synechocystis*, pese a su gran potencial, es un organismo poco caracterizado cinéticamente y de una gran complejidad. Este proyecto analiza los resultados arrojados por el algoritmo IOMA con el fin de conocer sus ventajas y limitaciones y permitir mejorarlo y estandarizarlo para uso habitual. Esto permitiría tratar la cianobacteria *Synechocystis* desde una nueva perspectiva y arrojar aproximaciones más realistas de su comportamiento.

Este proyecto se enmarca dentro de un proyecto concedido por la Comisión Europea en el seno del Séptimo Programa Marco, cuyo objetivo global es el diseño, la construcción y la demostración de la producción de biocombustibles solares mediante el uso de nuevas factorías celulares fotosintéticas: CyanoFactory ([cyanofactory.eu/](http://cyanofactory.eu/), FP7 Energy Contract 308518). Se trata de un proyecto a desarrollar por un consorcio formado por diez grupos de investigación pertenecientes a siete países europeos. Dentro de este proyecto, el grupo de investigación en el que se desarrolla este trabajo final de grado es el encargado de modelizar el comportamiento de las cianobacterias empleadas como plataforma de producción y optimizar las condiciones en que se maximiza dicha producción. Para ello, se hace uso de los datos experimentales que provienen de los demás grupos de investigación que participan en el proyecto, y a su vez, los datos resultantes de las simulaciones pueden guiar las investigaciones de estos grupos. Se han utilizado datos bibliográficos debido a la falta de datos de colaboradores pero en el futuro se pretende aplicar este algoritmo con datos 'ómicos' facilitados por otro de los grupos de este proyecto europeo.

En este capítulo introductorio se explicaran las bases de la biología de sistemas y del modelado metabólico, así como las herramientas más comúnmente utilizadas. Además se hará una breve introducción sobre las cianobacterias, y en concreto *Synechocystis sp. PCC 6803*, organismo sobre el que se trabaja.

## 1.2. Modelado de sistemas biológicos

Los organismos vivos son esencialmente sistemas químicos que están integrados por moléculas inanimadas que cumplen las mismas leyes químicas y físicas que las moléculas que forman la materia inerte y no existe nada en ellos que desobedezca estas leyes (Nelson y Cox, 2005).

Es posible definir un organismo vivo de acuerdo con las reacciones que suceden en su interior, tomando cualquiera de sus distintos niveles de organización (células, orgánulos o rutas metabólicas). En el interior de cada célula, materia y energía extraídas del ambiente se someten a una serie de transformaciones que reciben el nombre de metabolismo celular y que relaciona genotipo y fenotipo.

Este proyecto está basado en varios enfoques simultáneos, utilizando varias ciencias como son la ingeniería metabólica, la biología de sistemas y la bioinformática. Se utiliza el conocimiento sobre el metabolismo de un organismo para obtener una visión global sobre la respuesta del organismo en diversas condiciones.

La ingeniería metabólica se define como la tecnología de manipulación de las rutas metabólicas del organismo, utilizando diversas técnicas para la mejora herramienta del mismo, a partir del conocimiento de su metabolismo (Stephanopoulos, 1994). Para ello se pueden utilizar técnicas como la ingeniería genética, la tecnología recombinante o la alteración racional y dirigida de las rutas metabólicas, el objetivo suele ser redirigir el consumo de ciertos metabolitos hacia la producción de moléculas de interés. En este caso se hace uso de las ciencias computacionales para prever el funcionamiento metabólico de una cianobacteria, con la aparición de las plataformas 'ómicas' que daban grandes sets de datos sobre la composición y la respuesta de un organismo a estímulos, surgió la posibilidad de utilizar toda esta información para la comprensión del mismo como un sistema en su conjunto, en lugar de como partes independientes (Westerhoff y Palsson, 2004). Y en esto se basa la biología de sistemas (o biología sistémica), en el estudio de un organismo o sistema biológico, visto como un sistema integrado e interrelacionado de genes, proteínas y reacciones bioquímicas, que da lugar a procesos biológicos.

La biología sistémica trata de reconstruir y modelizar matemáticamente (Selinger et al, 2003) diferentes redes de interacción molecular que se producen dentro de un proceso o sistema biológico, con el fin de buscar propiedades emergentes de interés.

Para definir un modelo matemático de cualquier sistema, primero deben establecerse las entradas de las que se dispone, ej. las propiedades del sistema, y las salidas que se desea obtener. El propio modelo impone una serie de restricciones que relacionan las entradas del sistema con las salidas. También debe definirse el rango de valores que pueden tomar dichas entradas y el nivel de precisión mínimo deseado para las predicciones.

Un enfoque completo tomando el organismo como un sistema biológico requiere una completa caracterización de los constituyentes, de las interacciones entre moléculas, y de las funciones celulares; una caracterización espacio-temporal del organismo, es decir conocimiento sobre la dinámica de los componentes, su transporte y compartimentalización; y

por último un análisis minucioso sobre las respuestas moleculares a los cambios tanto externos como internos. La información obtenida de la caracterización de los componentes y su dinámica se utiliza para construir modelos matemáticos que permiten comprobar las respuestas moleculares obtenidas, hacer predicciones teóricas y comprobarlas, descubrir nuevos mecanismos biológicos no descritos con anterioridad y como meta ideal, servirían para desarrollar estrategias racionales para controlar y manipular las células (Bruggeman y Westerhoff,2007).

La bioinformática aplicada a la biología de sistemas trata de simplificar el funcionamiento de un organismo lo suficiente para poder utilizarlo y predecir su comportamiento, manteniendo a su vez la complejidad intrínseca de los organismos vivos. Una de las formas en que la informática es capaz de utilizar el metabolismo celular para una mejor comprensión de los mecanismos celulares e intentar predecir cómo va a reaccionar el organismo a ciertos cambios es a través un modelo metabólico del organismo que comprenda todas las reacciones del metabolismo del mismo, todas las proteínas y metabolitos que forman parte de ellas, así como toda reacción de intercambio de energía y materia con el medio y una función de biomasa.

Un modelo metabólico puede ser construido siguiendo dos aproximaciones principales: un enfoque ‘top-down’ o ‘bottom up’. El enfoque ‘top-down’ se basa en datos provenientes de plataformas ómicas diversas como el RNA-seq para dilucidar los componentes y el funcionamiento del organismo, mientras que el enfoque ‘bottom-up’ se basa en información genómica y bibliográfica (Shahzad and Loor, 2012).

Oltvai y Barabasi (2002) representaron los diferentes niveles de organización jerárquica de un organismo de forma piramidal y describieron las aproximaciones descritas anteriormente como direcciones opuestas para abordar el conocimiento de la estructura interna de un organismo. La aproxima partiendo de las moléculas que forman el organismo hasta llegar a la construcción de módulos o redes (bottom-up) se conoce como universal, mientras que la aproximación partiendo de una red universal hasta los módulos propios del organismo (top-down) recibió la nomenclatura de especificidad de organismo. Esta terminología se basa en que la composición de los organismos a nivel molecular es única para cada organismo y que conforme se sube en el nivel de organización jerárquica del organismo se encuentra una mayor similitud con otros organismos, es decir comparten más rutas metabólicas importantes, motivos o módulos mientras que las mayores diferencias se encuentran en la presencia de componentes únicos (Oltvai y Barabasi, 2002).

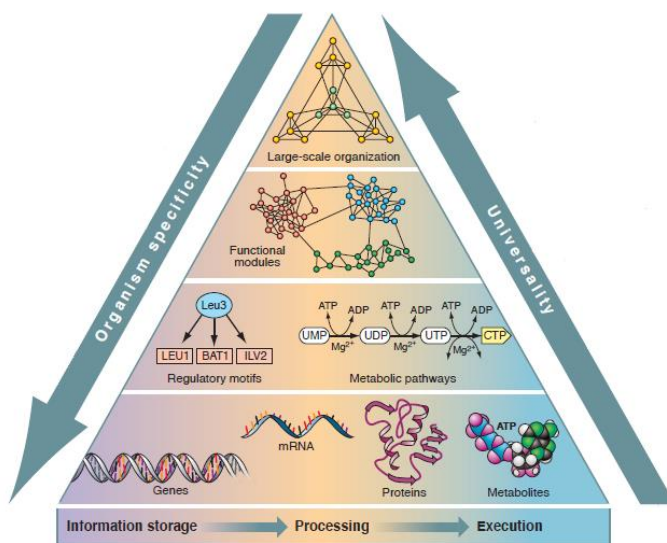


Figura 1. Pirámide que explica la complejidad jerárquica de los organismos. Oltvai y Barabasi (2010).

Es importante tener en cuenta que ambas estrategias son, no solo correctas, si no esenciales para obtener el máximo conocimiento del funcionamiento sistémico de un organismo. Es decir que todos los datos posibles se emplean para acercarse al conocimiento del organismo como sistema biológico, pero difieren en el material de partida y en la finalidad para la que se utilizan.

### 1.2.1. Enfoque 'top-down'

El enfoque 'top-down' comienza con una vista panorámica del comportamiento del sistema midiendo datos experimentales del genoma y pretende caracterizar mecanismos biológicos más específicos, es decir las partes y sus interacciones. El objetivo al usar este tipo de aproximación es descubrir nuevos mecanismos moleculares usando ciclos iterativos. Estos ciclos iterativos cuentan con **5 pasos principales**. Comienzan con la realización de experimentos de laboratorio, utilización de técnicas 'ómicas' y recogida de datos y una vez se comprueba la significación estadística de los datos, se utilizan herramientas bioinformáticas para la anotación de los genes o la predicción de la función que realizan, y la categorización de esta función en rutas metabólicas, en su proceso biológico o en su localización celular. Posteriormente se pueden utilizar herramientas bioinformáticas más complejas para analizar la influencia de estos datos en sus rutas metabólicas o para dilucidar interacciones biológicas. Estas herramientas bioinformáticas fueron diseñadas con la intención de proporcionar una visualización de los genes y sus interacciones, generar redes de interacciones proteína-proteína o evaluar los cambios en los flujos metabólicos. Por último es necesario evaluar manualmente los datos que arrojan estas herramientas bioinformáticas, interpretarlos, contrastarlos con artículos publicados y con las bases de datos específicas del organismo que estén disponibles; y ,en caso de que sea posible, formular hipótesis relativas a la regulación e interacción de los componentes del organismo (Shahzad y Loor,2012).

Estas hipótesis predicen nuevas correlaciones, que pueden ser comprobadas en nuevos experimentos o en posteriores análisis bioquímicos. Los mayores puntos fuertes de la biología de sistemas 'top-down' son que es potencialmente completa y hace uso de los datos que arrojan los análisis del metaboloma, el fluxoma, el transcriptoma y el proteoma (Westerhoff y Palsson, 2004).

Cualquier estudio de biología de sistemas de este tipo requiere de un set de datos experimentales muy grandes de un solo organismo sometido a unas pocas perturbaciones que pueden ser genéticas, ambientales o inducidas ya sea por RNA de interferencia, modificación de su organización dinámica o administración de medicamentos. Estos estudios persiguen ampliar la visión del comportamiento del sistema para descubrir patrones de comportamiento que sean lo suficientemente genéricos como para tener potencial predictivo sobre los mecanismos biológicos presentes en el sistema y para descubrir procesos relacionados funcionalmente que puedan estar bajo el control de un set común de factores de transcripción.

Los modelos utilizados en la biología de sistemas 'top-down' son fenomenológicos, es decir no están basados en los mecanismos del sistema y a menudo no emplean el conocimiento pre-existente sobre las relaciones específicas entre los componentes moleculares. Aun así estos modelos suelen describir bien las correlaciones pero están muy lejos de la realidad. Normalmente cuando se usa este enfoque se utilizan diferentes modelos virtuales generales en diferentes fases de la caracterización del sistema. En estos una vez confirmada la presencia de los componentes, se utilizan modelos matemáticos con grupos de variables lineales dependientes o independientes como el análisis de componentes principales (PCA) que identifica las moléculas predominantes y diferenciadoras de cierta condición, o modelos Bayesianos que producen arboles de regulación.

Los resultados que arrojan estos modelos pueden propiciar nuevos experimentos o dar lugar al descubrimiento de nuevas interacciones y mecanismos en posteriores iteraciones. Es un proceso iterativo. Este tipo de enfoque de biología sistemática se utiliza generalmente con aquellos sistemas o subsistemas celulares que no están caracterizados en detalle y en los que falta mucho por descubrir. Sin embargo, al utilizar este enfoque se debe ser consciente de sus puntos débiles, entre los que se encuentra la imprecisión de algunas de las reglas matemáticas que se aplican al obviar procesos importantes como son el feedback, las modificaciones post-traduccionales o la degradación de proteínas.

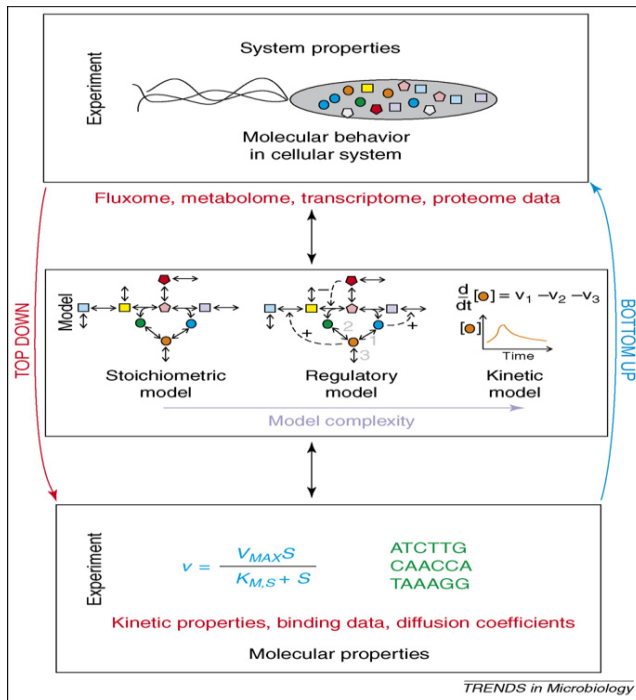


Figura 2. Explicación de los enfoques 'top-down' y 'bottom-up' en el artículo escrito por Bruggeman y Westerhoff (2006).

### 1.2.2. Enfoque 'bottom-up'

Mientras la biología de sistemas 'top-down' esclarece el comportamiento celular a través de la inducción, la 'bottom-up' deduce las propiedades funcionales a partir de un sistema o subsistema muy bien caracterizado usando mecanismos moleculares. Recibe su nombre porque comienza desde abajo, desde las partes constitutivas del sistema, formulando las ecuaciones que describen el

comportamiento de cada componente, e integrándola hasta conseguir predecir el comportamiento del sistema. La meta a conseguir con este enfoque es combinar todas las rutas e interacciones de cada componente hasta conseguir un modelo potencialmente predictivo para el sistema en su conjunto.

Este tipo de enfoque depende principalmente de estudios experimentales que determinen las propiedades de cada componente, datos sobre las respuestas del sistema o los subsistemas a las perturbaciones a las que se pueda ver expuesto, la construcción de modelos detallados basados en los mecanismos internos del sistema, y el desarrollo de herramientas que permitan analizar el modelo y representarlo.

No todos los modelos usados en este tipo de aproximación recogen información de las mismas fuentes para la construcción del modelo, ni todos se basan en el metabolismo celular para predecir el comportamiento, por ejemplo un modelo de la respiración celular podría hacerse a nivel atómico. Sin embargo, los modelos metabólicos son los más utilizados pues el metabolismo actúa como puente para relacionar el genotipo de un organismo y su fenotipo. En este tipo de modelos metabólicos normalmente se pretende que la estructura del modelo refleje la estructura estequiométrica de las reacciones del sistema y la de la regulación alostérica. Un modelo bien construido debe poder ser simulado bajo diferentes condiciones fisiológicas.

### 1.2.3. Modelos metabólicos a escala genómica

Uno de los modelos metabólicos más utilizados en la actualidad es el llamado ‘modelo-metabólico a escala genómica’, el cual se considera que se construye siguiendo una aproximación intermedia o ‘middle-out’ (Stephanopoulos, 2004). Se considera que es intermedio porque aplica ambos métodos, se reconstruye un modelo mecanicista basado en las partes y en las reacciones metabólicas siguiendo un enfoque ‘bottom-up’ pero se pueden utilizar datos ómicos para establecer ciertos mecanismos regulatorios.

Existen diferentes tipos de modelos y se pueden clasificar diferentemente atendiendo al grado de profundidad con el que se estudian. Se pueden encontrar modelos teóricos, modelos causales o modelos cinéticos. Los modelos teóricos son modelos que estudian el sistema como un todo sin entrar en detalles, y mecanismos particulares aunque siguiendo una serie de reglas globales, modelos causales basados en algunas reglas particulares de interacción y evolución de las partes y los modelos cinéticos que tienen en cuenta los cambios cinéticos en velocidades y constantes de las partes. En un modelo cinético la concentración de un metabolito en el tiempo  $x(t)$  dependerá de las velocidades de las reacciones que lo producen ( $v_p(t)$ ) y de las velocidades que lo consumen ( $v_c(t)$ ), y por tanto se puede definir matemáticamente así:

$$\frac{\delta X}{\delta t} = v_p - v_c$$

Ecuación 1. Variación temporal de la concentración de un metabolito,  $x(t)$ .

Las velocidades de producción y consumo del metabolito dependen de la concentración de los sustratos, de la concentración de los productos y de los parámetros cinéticos. Este tipo de modelos son muy útiles para estudiar la dinámica de un sistema y los mecanismos regulatorios de los sistemas pequeños. Sin embargo, estos modelos tienen una utilidad limitada en sistemas grandes, pues no hay gran disponibilidad de parámetros cinéticos, ya que estos parámetros están medidos generalmente *in vitro*, donde las condiciones suelen variar de las condiciones *in vivo*.

Por el contrario los modelos metabólicos a escala-genómica son modelos que reflejan la totalidad del metabolismo del organismo. El precio que se paga por incluir todas las transformaciones bioquímicas es que se hace necesario obviar los procesos cinéticos asociados. La aproximación que se utiliza para estudiar estos modelos se basa en que la mayoría de rutas acaban alcanzando el estado estacionario en el que el consumo y la producción de todos los metabolitos internos se equilibran y no se produce acumulación de los mismos.

$$\frac{\delta X}{\delta t} = v_p - v_c = 0$$

Ecuación 2. Concentración del metabolito X en función del tiempo en un estado estacionario.

La ecuación 2 refleja que la concentración de un metabolito X en el tiempo, una vez alcanzado el estado estacionario, es constante. Es decir, la velocidad a la que se crea el metabolito es igual a la velocidad con la que se destruye. A estas velocidades de las reacciones en estado estacionario se las denomina flujos.

### 1.3. Construcción de un modelo metabólico

Para construir un modelo metabólico a escala genómica se combina toda la información del organismo disponible que aporte detalles de las interacciones biológicas que ocurren en su interior. La construcción de un modelo metabólico se basa en cuatro etapas básicas con múltiples pasos.

1. En la primera etapa se reconstruye **un esquema base de las reacciones metabólicas que ocurren en el sistema** basado en la anotación más actual del genoma del organismo que se desea modelar y en la información de bases de datos bioquímicas. El primer paso consiste en relacionar la anotación del genoma con las funciones metabólicas de estos genes, y si fuera posible con las proteínas que codifican. Esto se puede hacer automáticamente con ciertas herramientas bioinformáticas como Pathway tools (Stein, 2001). El segundo paso, que también se puede hacer de forma automática es la asociación de estas proteínas con reacciones metabólicas utilizando su número EC y bases de datos como BRENDA o KEGG (M Kanehisa & Goto, 2000; Chang et al., 2009). El problema es que algunas proteínas que no catalizan reacciones metabólicas también tienen números EC, como las proteínas relacionadas con la metilación del DNA. Por esta y otras razones, como la inespecificidad en el uso de cofactores o en la denominación de algunos compuestos o la falta de información en la anotación del genoma, es necesaria la segunda etapa, el refinamiento manual de la reconstrucción.
2. En esta etapa se trata de **comprobar el contenido del borrador anteriormente creado para eliminar aquellas reacciones que no formen parte del metabolismo o añadir aquellas que sean necesarias para el funcionamiento del organismo y no estén**. Al finalizar esta etapa se debe tener una reconstrucción del metabolismo del organismo lo más precisa posible para convertirlo en un modelo matemático. Para ello se evalúa individualmente cada reacción y se compara con la información disponible en la literatura, pues no todas las anotaciones del genoma son fiables y las bases de datos con las que se trabaja no suelen ser específicas de cada organismo por lo que se puede estar añadiendo reacciones no presentes en él mismo, u obviando otras que son necesarias. Esta etapa es la más complicada y costosa de todo el proceso pues consta de muchos pasos y es la piedra angular del modelo metabólico.

En primer lugar se verifican todas las reacciones del esquema inicial, normalmente se recomienda seguir un orden por rutas metabólicas porque esto permite analizar simultáneamente la reacción individual y la ruta al completo, identificando reacciones no presentes en el borrador pero necesarias para conectar la ruta metabólica en la que se incluye (Thiele y Palsson, 2010). En ese caso si no se encuentra indicada explícitamente en la literatura se puede recurrir a organismos relacionados. También es necesario si existe más de una reacción conocida y anotada en las bases de datos para una misma enzima comprobar qué sustratos y co-factores utiliza el organismo de interés para elegir la reacción o reacciones que es capaz de catalizar. Además, se puede comprobar si una reacción tiene términos genéricos (p.e. aceptor de electrones) y corregirlo o eliminar la reacción. También se comprueban las cargas de los metabolitos de cada fórmula en el pH celular, la estequiometría y la direccionalidad de la reacción. Durante todo este proceso se estudia la existencia de isoenzimas o la posibilidad de que la enzima actúe como complejo proteico. Por último, hay que

asegurarse de que todos los metabolitos tengan un identificador reconocido por la comunidad científica.

Una vez hecho el refinamiento manual de la reconstrucción, es el momento de anotar aquellas reacciones que no estaban anotadas en el genoma y no se han encontrado descritas para el organismo de interés en las bases de datos pero son necesarias para el correcto funcionamiento del modelo. Por eso es necesario contar con un sistema de puntuación de la confianza que se tiene en la existencia de las reacciones para agilizar el proceso de corrección del modelo en posteriores revisiones. En el protocolo escrito por Thiele y Palsson (2010) para la construcción de este tipo de modelos se aconseja un sistema de puntuación del 0 al 4 donde 0 es la confianza más baja o nula, y 4 una confianza muy alta en que este presente esta reacción. Una vez hecho esto se puede comenzar a añadir aquellas reacciones no presentes en el genoma pero necesarias, como: 1.Reacciones espontaneas, 2.Reacciones de transporte extra-celular e intra-celular, 3.Reacciones que corresponden a funciones necesarias pero no anotadas, 4.Reacciones que permiten la acumulación de componentes que de otra forma no se acumularían en una simulación en estado estacionario, 5.Reacciones asociadas con el mantenimiento y la hidrólisis del ATP necesario como fuente de energía para el crecimiento y la replicación. Además de todas las reacciones anteriormente comentadas, es muy importante añadir una reacción para la composición de la biomasa. Esta última reacción es crucial pues define todos los componentes necesarios para producir la biomasa y la fracción de estos que es necesaria para producir una unidad de biomasa que corresponde a 1 gramo de peso seco de las células. Esta reacción se puede o bien calcular experimentalmente, o bien puede ser estimada.

3. Una vez se ha asumido que tenemos una reconstrucción que reproduce la realidad lo más fielmente posible con la información disponible se procede a convertir esta reconstrucción en un modelo matemático, que correspondería a la tercera etapa del proceso. En esta etapa se transforma el modelo en una matriz estequiométrica que relaciona los diferentes metabolitos del sistema con las reacciones de las que forman parte y, consecuentemente, con las proteínas que catalizan estas reacciones. Además, en esta etapa se deben imponer una serie de restricciones al modelo para limitar el número de soluciones posibles, fijando unos límites coherentes para las entradas y las salidas.

La elección del método analítico usado depende en gran medida del conocimiento disponible que se incorpore al modelo (Kitano, 2002). En el caso de los modelos metabólicos a escala genómica hablamos de sistemas en estado estacionario en los que no son necesarios parámetros cinéticos de las reacciones, y un análisis en estas condiciones solo necesita la estructura de la red. Como se ha dicho anteriormente también es conveniente establecer una serie de restricciones de entrada y salida. Para este tipo de modelos se han desarrollado métodos analíticos como el FBA (Flux Balance Analysis) y el MOMA (Minimization of metabolic adjustment), Regulatory On/Off Minimization (ROOM), Flux Variability Analysis (FVA), Flux Coupling Analysis (FCA), y muchos otros. En este caso nos centraremos en el FBA que es el método comúnmente utilizado. EL FBA o Análisis de balance de flujos.

El FBA (Orth et al., 2010) es un método de modelización basado en restricciones, es decir que el espacio de posibles soluciones viene definido por las restricciones. El primer paso del FBA es representar matemáticamente las reacciones metabólicas en forma de una matriz numérica de los coeficientes estequiométricos de cada reacción que imponen restricciones en el flujo de estos metabolitos a través de la red. La matriz



estequiométrica impone las restricciones de balance de masas asegurando que la cantidad total que se produce de cualquier componente es igual a la que se consume, en el estado estacionario (Orth et al., 2010). Además cada reacción tiene unos límites inferiores y superiores que definen los valores máximos y mínimos permitidos. Estas restricciones llevan a un sistema compatible indeterminado, sin embargo es posible identificar un espacio limitado de todos los flujos posibles (Kauffman et al., 2003).

El siguiente paso del FBA es definir un fenotipo, en forma de objetivo biológico que sea relevante para el problema que se estudia por ejemplo, en caso de estudiar el crecimiento, la función objetivo es la de producción de biomasa anteriormente descrita.

Si se tienen en cuenta las representaciones matemáticas de las reacciones y del objetivo se observa que un modelo metabólico bajo estas características se puede ver que el FBA no es más que un problema de programación lineal (LP). La LP es una técnica de optimización, en la que una función objetivo, que debe ser lineal, es optimizada (maximizada o minimizada) sujeta a las restricciones de igualdades y desigualdades.

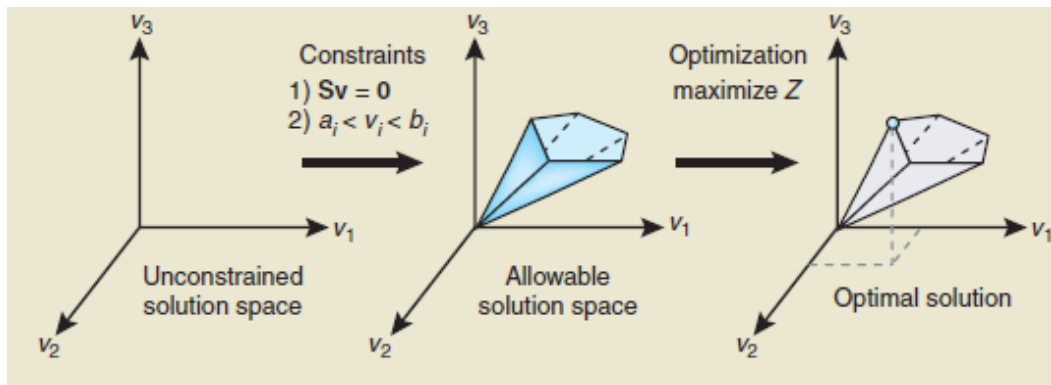


Figura 3. Representación gráfica de cómo funciona el FBA, definiendo primero el espacio de posibles solución del sistema y dando la solución más óptima para la función objetivo definida. Imagen extraída del trabajo de Orth et al., (2010).

El FBA permite hacer todas las simulaciones deseadas, definiendo diferentes funciones objetivo. Con esto ya se tiene construido un modelo metabólico que nos permite calcular los flujos del sistema en diferentes condiciones, pero no sabemos cuánta confianza tenemos en estos resultados. Por esta razón es necesario evaluar el modelo.

4. Y así ocurre en la cuarta etapa de la construcción de un modelo, se evalúa la red construida y se trata de descubrir sus posibles fallos. Para ello se comprueba su habilidad para crear los precursores de la biomasa descubriendo así si falta alguna reacción esencial para la síntesis de los mismos, buscando puntos muertos que den lugar a metabolitos desconectados de la red, verificando la producción no sólo de los precursores, sino también la de los subproductos, así como simulando incapacidades conocidas del organismo.
5. La construcción de un modelo metabólico a escala genómica es un proceso iterativo, por lo que, una vez descubiertos los errores en esta etapa, se vuelve a la etapa 2 y se repite todo el proceso cíclicamente hasta que se llega a un modelo metabólico útil y preciso en la medida de lo posible. Sin embargo, la revisión se sigue haciendo durante todo el uso del modelo. Para mejorar la precisión de las simulaciones se requiere de tanta información experimental como sea posible de un modelo

En este trabajo se utiliza un modelo metabólico a escala genómica de la cianobacteria *Synechocystis sp. PCC 6803* reconstruido por Montagud et al. (Montagud et al., 2010), (Montagud et al., 2011).

## 1.3. Cianobacterias

### 1.3.1. Generalidades sobre las cianobacterias

Las cianobacterias son un grupo de microorganismos foto-autotróficos, que se incluyen en el dominio Bacteria. Pueden ser definidos como organismos que contienen, en el interior de una célula procariota, un aparato fotosintético muy similar en función, estructura y propiedades moleculares, al presente en un cloroplasto eucariota (Giovannoni et al., 1988). Por esta razón se cree que las cianobacterias fueron las inventoras de la fotosíntesis oxigénica y tuvieron un papel importante en la aportación de oxígeno a la atmósfera durante la era precámbrica. Además cada vez cobra más fuerza la teoría de que los cloroplastos fueron originados por endosimbiosis de algún ancestro de las cianobacterias actuales.

La maquinaria fotosintética de este grupo difiere de los otros dos grupos de bacterias fotosintéticas, bacterias moradas y verdes, en el sistema de pigmentos fotosintéticos y en su capacidad de llevar a cabo una fotosíntesis oxigénica, dejándolos por tanto como el único grupo de bacterias aeróbicas fotoautotróficas (Stanier y Cohen-Bazire, 1977).

Las cianobacterias, colonizan sistemas tanto terrestres como acuáticos, aunque es en el agua donde generalmente se agregan y proliferan. Crecen a temperaturas entre 15 y 30 °C y a pH diversos entre 6 y 9. Se reproducen asexualmente por bipartición, fragmentación de filamentos o mediante esporas. La mayoría de ellas son capaces de tomar N<sub>2</sub> y reducirlo a amonio para metabolizarlo. Algunas tienen motilidad, y tienen un metabolismo secundario muy activo que les confiere la capacidad de sintetizar gran cantidad de compuestos. Estos compuestos pueden ser toxinas, antibióticos, antivirales o biocombustibles.

### 1.3.2. Phylum cianobacteria

El phylum cianobacteria comprende organismos con morfología y comportamiento muy diversos. Se encuentran organismos muy pequeños, organismos que se pueden ver a simple vista, organismos unicelulares, organismos filamentosos, organismos capaces de fijar el nitrógeno, organismos incapaces de hacerlo (Knoll, 2008), etc., y sin embargo conservan muy bien el ARN ribosomal y algunas de sus proteínas. Swingley et al. (2008) generaron un árbol filogenético a partir de 340 familias de proteínas presentes en todas las cianobacterias.

En él se puede ver que *Synechocystis sp. PCC6803* se encuentra altamente relacionado con *Crocospaera Watsoni* y bastante alejado de cianobacterias estudiadas como *Synechococcus*.

*Synechocystis sp. PCC 6803* (en adelante puede ser nombrado como *Synechocystis 6803*) es una cianobacteria unicelular no fijadora de nitrógeno que habita el ambiente marino. Es uno de los organismos más populares para estudios genéticos y fisiológicos debido a dos razones principales: se transforma naturalmente con DNA exógeno, (Grigorieva and Shestakov, 1982) y es capaz de crecer en autotrofia, en heterotrofia a partir de glucosa (Rippka et al., 1979), con una mínima aportación de luz (Williams, 1988) y en mixotrofia (Wang et al., 2002).



Figura 4. Árbol filogenético creado por Swingley et al. a partir de 341 familias de proteínas.

Además su genoma fue secuenciado completamente en 1996 (Kaneko et al., 1996). Fue el cuarto genoma en ser secuenciado en su totalidad y el primero de los organismos fototróficos. Otra propiedad interesante de *Synechocystis 6803* es su motilidad fototáctica descrita por Yoshihara e Ikeuchi (2004).

Respecto a su crecimiento, es muy dependiente de la disponibilidad de nutrientes y las condiciones de cultivo. *Synechocystis 6803* es capaz de llevar a cabo tanto la fotosíntesis oxigénica como la fijación de  $\text{CO}_2$ , gracias a la presencia de la enzima Ribulosa-1,5-bisfosfato carboxilasa (RuBisCO) que cataliza el primer paso del ciclo de Calvin. Como esta enzima es muy sensible al oxígeno, *Synechocystis* ha desarrollado un orgánulo llamado carboxisoma que es impermeable al oxígeno, permitiendo así que la fotorespiración se mantenga a niveles bajos durante el crecimiento autotrófico. *Synechocystis 6803* puede crecer tanto fotoautotróficamente como fotoheterotróficamente a partir de glucosa vía la ruta glicolítica y la ruta oxidativa de las pentosas fosfato.

Las cianobacterias no pueden absorber toda la luz del sol debido a la reflexión de la misma, a la disipación, al efecto de sombreado, al espectro de absorción limitado de la antena fotosintética. A esto hay que sumarle su protección a radiaciones demasiado energéticas, que podrían provocar la oxidación, con mecanismos de disipación que también reducen la cantidad de fotones que reciben. Además los fotones absorbidos por las antenas no pueden ser utilizados en su totalidad para su conversión en energía. Actualmente, la limitación de luz que recibe un cultivo es el paso limitante para una producción de biomasa eficiente.

La mayoría de las cianobacterias son capaces de usar nitrato y amonio como fuentes de nitrógeno. *Synechocystis 6803* no puede fijar  $\text{N}_2$  de forma natural pero es capaz de almacenar amonio en el interior de la célula a través de la producción de cianoficina, un polipéptido rico en residuos de arginina y aspartato (Berg et al., 2000). Este polímero puede ser degradado por una enzima hidrolítica, cianoficinas, para usar el nitrógeno en condiciones de limitación de nitrógeno. Además es capaz de acumular fosfato inorgánico para utilizarlo si es necesario.

En general, la falta de nitrógeno y fósforo, conlleva una menor expresión de la fotosíntesis y una estimulación de las respuestas metabólicas necesarias para el mantenimiento de la homeostasis, por lo que el suplemento de estos nutrientes es crucial para el cultivo de *Synechocystis 6803*.

Las cianobacterias son un grupo altamente estudiado por su interés tanto en estudios filogenéticos o de biodiversidad, como su potencial uso en aplicaciones biotecnológicas. Los estudios filogenéticos pretenden relacionar las cianobacterias con el resto de grupos de procariotas, ya que su taxonomía todavía no está clara, y ayudar a dilucidar su papel en la evolución de los plástidos presentes actualmente en las plantas (Kaneko et al., 1996).

#### 1.3.4. Importancia e información sobre *Synechocystis sp. PCC 6803*

Las cianobacterias en general y *Synechocystis 6803* en particular debido a su importancia como organismo modelo y al interés de su metabolismo fotoautotrófico, han sido utilizadas como productores de biocombustibles, químicos, biomateriales y compuestos relacionados con la salud (revisado por Yu et al., 2013). Esto se debe en gran parte a su facilidad para ser transformadas.

Además de los compuestos producidos por *Synechocystis* gracias a la ingeniería genética, existen compuestos de interés industrial que es capaz de sintetizar de forma natural como alcoholes o hidrógeno. El hidrógeno es un biocombustible limpio, pero además es ampliamente utilizado para la síntesis y mejora de los combustibles fósiles, por lo que sería muy conveniente obtener un microorganismo capaz de producirlo en cantidades considerables a partir de fuentes renovables, como la luz solar y el CO<sub>2</sub>. *Synechocystis 6803* es capaz de producirlo en atmósfera anaeróbica y en ausencia de luz gracias a la presencia de una hidrogenasa bidireccional. Por todas estas características únicas se están haciendo muchos esfuerzos y progresos en el estudio y la comprensión de esta cianobacteria.

Por tanto, *Synechocystis 6803* es una potencial factoría celular pero presenta una serie de inconvenientes. En primer lugar su crecimiento es mucho más lento que el de factorías celulares actualmente utilizadas como *E.coli*. Además su rendimiento y la velocidad a la que produce las moléculas de interés es muy bajo, por lo que se requiere de más tiempo para sintetizar cantidades considerables de un producto, lo que aumenta el riesgo de contaminaciones o de degradación del producto (Yu et al., 2013). Por último existen inconvenientes debido al alto coste de producción de los biorreactores adecuados y a la dificultad del escalado.

La ingeniería metabólica trata de solucionar estos problemas mediante knock-outs de aquellas reacciones que supongan un problema o mediante la adición de otras reacciones que permitan solucionarlo. Y es aquí donde entran los modelos metabólicos. Un modelo robusto y preciso cuyos resultados sean fidedignos permite realizar una aproximación computacional previa de los experimentos de ingeniería metabólica que se pretende llevar a cabo para comprobar si estos están bien planteados.

## 2. OBJETIVOS

El objetivo de este proyecto es comprobar la mejora que se obtiene al aplicar el algoritmo IOMA, desarrollado por Yizhak et al. (2010), a la cianobacteria *Synechocystis* y comparar los resultados obtenidos con respecto a la aplicación de un FBA básico. El algoritmo IOMA integra datos proteómicos y metabolómicos cuantitativos a un FBA a diferencia del FBA que únicamente utiliza las restricciones estequiométricas y las restricciones de entradas y salidas establecidas por el usuario.

Más concretamente, el IOMA se aplicará al modelo de *Synechocystis* PCC 6803 detallado en el artículo escrito por Montagud et al. (2011).

Este proyecto supone un reto en sí mismo pues se comprobará el funcionamiento del IOMA en un organismo más complejo que el modelo de *E. Coli K-12 MG1655* (Feist et al., 2007) utilizado por Yizhak et al. al describir el algoritmo. El modelo que utilizaremos supera tanto en número de reacciones como en diversidad de rutas y mecanismos al utilizado de *E. Coli*, ya que trabajamos con un organismo fotosintético capaz de crecer en condiciones de autotrofia, mixotrofia y heterotrofia. Asimismo, existe la dificultad añadida de que trabajaremos con un organismo mucho menos caracterizado, lo que nos permitirá extraer conclusiones adicionales sobre la sensibilidad del funcionamiento del IOMA respecto a la variación de las condiciones iniciales.

En primer lugar se realizará un análisis exhaustivo del funcionamiento del IOMA partir de un modelo simplificado de sólo 20 reacciones que permite visualizar claramente los cambios que se producen en las distribuciones de flujos al incorporar ciertos datos proteómicos y metabolómicos.

Posteriormente se aplicará al modelo *Synechocystis* sp. PCC 6803. Este modelo consta de 956 reacciones (de las que 58 son de intercambio), 811 genes y 911 metabolitos. Además, a este modelo se incorporarán los datos 'ómicos' conocidos de este organismo, así como los datos cinéticos de las reacciones de interés. Para obtener dicha información se realizará una revisión bibliográfica de la literatura científica reciente.

A partir de los resultados obtenidos en el desarrollo de este trabajo, se desarrollará un método de análisis de los datos experimentales proporcionados por colaboradores en el proyecto europeo 'Cyanofactory'.

### 3. MATERIALES Y MÉTODOS

En este apartado se pretende explicar el conocimiento aplicado, las herramientas utilizadas y las estrategias seguidas para el desarrollo de este trabajo. En primer lugar se explica el algoritmo que se pretende evaluar, y más adelante las herramientas utilizadas para aplicarlo, en este caso, el modelo metabólico utilizado, la obtención de los datos tanto metabolómicos y proteómicos como cinéticos, las restricciones impuestas y la herramienta de medida elegidas para comparar resultados.

#### 3.1 Algoritmo

Este proyecto está basado en la implementación en *Synechocystis sp. PCC 6803* del algoritmo desarrollado por Yizhak et al. (2010). Se trata de un método de modelado basado en restricciones que permite la integración de datos proteómicos y metabolómicos cuantitativos en un modelo metabólico a escala genómica para predecir la distribución de flujos metabólicos.

Para asociar las medidas cuantitativas de proteínas y metabolitos con flujos metabólicos el algoritmo utiliza la siguiente ecuación que sigue una cinética de Michaelis-Menten (ecuación 3) donde  $v$  es el flujo de una reacción que transforma un conjunto de sustratos  $S$  a un conjunto de productos  $P$ .  $s_i$  y  $p_i$  son las concentraciones del sustrato  $i$ ésimo y el producto  $i$ ésimo respectivamente,  $e$  es la concentración de enzima, y  $k_{m,s_i}$  y  $k_{m,p_i}$  son las constantes de disociación de cada sustrato y producto, respectivamente.  $k_{cat}^+$  y  $k_{cat}^-$  representan el número de recambio en ambas direcciones.

$$v = k_{cat}^+ \cdot e \prod_{s_i \in S} \left( \frac{s_i}{k_{m,s_i} + s_i} \right) - k_{cat}^- \cdot e \prod_{p_i \in P} \left( \frac{p_i}{k_{m,p_i} + p_i} \right)$$

Ecuación 3. Ecuación que utiliza el algoritmo desarrollado por Yizhak et al.(2010) para estimar los flujos metabólicos utilizando datos cuantitativos extraídos gracias a las estrategias ómicas.

Los productorios de las concentraciones divididas entre la suma de las concentraciones más las constantes de disociación de la enzima para cada sustrato o producto reciben el nombre de valores de saturación, que se abrevian como  $a^+$  (ecuación 4) y  $a^-$  (ecuación 5), para sustratos y productos respectivamente.

$$a^+ = \prod_{s_i \in S} \left( \frac{s_i}{k_{m,s_i} + s_i} \right)$$

Ecuación 4. Valor de saturación en la dirección 'forward'.

$$a^- = \prod_{p_i \in P} \left( \frac{p_i}{k_{m,p_i} + p_i} \right)$$

Ecuación 5. Valor de saturación en dirección 'backwards'.

En la ecuación 6 se muestra el resultado de sustituir las ecuaciones 4 y 5 en la ecuación 3.

$$v = e \cdot (a^+ k_{cat}^- - a^- k_{cat}^+)$$

Ecuación 6. Ecuación 3 obtenida al sustituir los productorios de sustratos y productos por  $a^+$  y  $a^-$ , respectivamente.

Para poder utilizar cuantificaciones relativas de proteínas, la ecuación 6 se puede reescribir añadiendo la concentración del enzima en el estado de referencia,  $e^{ref}$ , dando lugar a la ecuación 7.  $v_{max}^+$  y  $v_{max}^-$  corresponden al número de recambio multiplicado por la  $e^{ref}$ .

$$v = \frac{e}{e^{ref}} \cdot (a^+ v_{max}^+ - a^- v_{max}^-)$$

Ecuación 7. Resultado de dividir la concentración del enzima por la concentración del enzima en el estado de referencia y de multiplicar  $k_{cat}^+$  y  $k_{cat}^-$  por la  $e^{ref}$ .

Por lo tanto, teniendo en cuenta la ecuación 7, se necesitara el ratio de concentración de proteína, las concentraciones de los metabolitos y sus constantes de disociación, y  $v_{max}^+$  y  $v_{max}^-$ , para estimar los flujos teóricos de las reacciones.

Ahora que se ha descrito la ecuación cinética tipo Michaelis-Menten que se utiliza se puede explicar el funcionamiento del programa en su conjunto.

Este algoritmo se utiliza para simulaciones en estado estacionario y trata de encontrar una distribución de flujos posible y óptima que cumpla las condiciones del estado estacionario (ecuación 8), las restricciones estequiométricas, de balance de masas y de la direccionalidad enzimática (ecuación 9); y además, intenta que los flujos a través de las reacciones del ‘core set’ sean tan consistentes como sea posible con las estimaciones teóricas de flujo que se obtienen gracias a la ecuación cinética tipo Michaelis-Menten (ecuación 10). En la ecuación 10 se definen los flujos cinéticos teóricos y se añade un error en la ecuación para permitir una cierta flexibilidad a los flujos, ya que si no sería imposible llegar a una solución factible y óptima. Este error representa la ausencia de datos experimentales de las concentraciones para alguno de los metabolitos de la reacción, la presencia de ruido en los datos tanto proteómicos como metabolómicos, y las simplificaciones hechas en la ecuación cinética de la velocidad. Para obtener una distribución de flujos lo más consistente posible con los datos proteómicos y metabolómicos con los que se cuenta se formula el problema de optimización como un problema de programación cuadrática (Ecuación 11) en el que calcula la distribución de flujos  $v$ , para una serie de condiciones ( $j=1, \dots, k$ ) con las concentraciones relativas de varias enzimas del ‘core set’ (denominado E), las concentraciones absolutas de sustratos y productos, y las constantes de disociación.

$$N\bar{v}^j = 0 \quad j = 1, \dots, k$$

Ecuación 8. Define la condición del estado estacionario, es decir que la suma de todos los flujos de producción y consumo para cada metabolito es igual a 0, para todas las condiciones  $j$ .

$$\bar{v}_{tb} \leq \bar{v} \leq \bar{v}_{ub} \quad j = 1, \dots, k$$

Ecuación 9. Representa las condiciones de capacidad de flujo y direccionalidad de las reacciones.

$$\bar{v}_i^j = \frac{e_i^j}{e_i^{ref}} (a_i^{j+} v_{max,i}^+ - a_i^{j-} v_{max,i}^- + \varepsilon_i^j) \quad j = 1, \dots, k$$

Ecuación 10. Define los flujos cinéticos teóricos y les añade un error para permitirles un poco de libertad a los flujos.

$$\min_{\bar{v}_j, \bar{v}_{max}^+, \bar{v}_{max}^-, \varepsilon_i^j} \sum_{i \in E} var(\bar{\varepsilon}_i)$$

Ecuación 11. Optimización del programa tratando de que se cumplan todas las condiciones anteriores minimizando la suma total de las varianzas de los errores para cada enzima en todas sus condiciones.

La ecuación 7 indica que se requieren varios parámetros cinéticos que no siempre están disponibles, sobretodo en el caso de  $k_{cat}$  o  $v_{max}$ , que son parámetros cinéticos en general menos estudiados que la constante de disociación de Michaelis-Menten. Para solucionar este problema el algoritmo incluye  $v_{max}^+$  y  $V_{mas}^-$  como parámetros de la optimización, y, por tanto le asignará el valor que más se adecue a cada enzima para todas las condiciones.

Este algoritmo se formuló para simular diferentes knock-outs, siendo cada condición un KO. Como en el caso de estudio las diferentes condiciones son ambientales y no se ha realizado ningún knock-out la parte del algoritmo en que se eliminan las reacciones relacionadas con los genes eliminados se suprimió.

El algoritmo además requiere de cierta información experimental adicional, como la velocidad de crecimiento en cada condición o las velocidades de captación y secreción de nutrientes. Cuando esta información es aportada, el programa asigna estos datos a los flujos en las correspondientes reacciones de forma fija. Es decir no añade errores ni permite ciertas fluctuaciones, si no que asume que el sistema introduce exactamente ese flujo de nutrientes. Esto supone la primera limitación de este algoritmo, pues aunque el crecimiento suele ser calculado experimentalmente y, por tanto es relativamente fiable, aun teniendo en cuenta que siempre suele haber un error asociado a la medición experimental; la entrada de nutrientes no suele ser calculada y no suele ser exacta esta medición. Por tanto estas entradas suelen ser poco fiables y deben ser tratadas con cautela, definiendo para ello las reacciones de intercambio, sus límites inferiores y superiores, añadiendo así una posibilidad de error. En este caso, se ha intentado utilizar el número mínimo de entradas de nutrientes fijas que requiere el algoritmo IOMA definiendo el resto de acuerdo a sus límites superiores e inferiores.

El IOMA está basado en operaciones con matrices sobre MATLAB, además IOMA utiliza un optimizador diferente al de MATLAB, ya que el optimizador propio se satura llegado a un cierto nivel y no permite hacer más iteraciones, por esta razón utilizan un optimizador TOMLAB, en nuestro caso el C-PLEX (TOMOPT, 2014).

El programa escrito está formulado como una función de MATLAB cuya primera línea corresponde a la Figura 5. En esta línea el programa define la función y el resultado de la función ([V\_v]) que será una matriz de flujos en las diferentes condiciones, en la que cada columna corresponde a todos los flujos del sistema en una condición determinada, y cada fila a todos los flujos de una reacción en las diferentes condiciones. Define también los inputs necesarios para que el sistema funcione, que en este caso son un modelo metabólico en el formato explicado más adelante, una matriz que contiene los datos proteómicos en forma de ratios (exp\_mat), una matriz con los valores de saturación de los sustratos de cada reacción para cada condición (met\_frw\_mat) donde cada fila corresponde a una condición y cada columna a una reacción del 'core set', una matriz con los valores de saturación de los productos de cada reacción para cada condición (met\_bck\_mat) del mismo formato que la anterior, un vector fila con los índices de las reacciones de los que se tiene datos proteómicos y/o metabólicos (rxns\_idx), y una matriz en la que se especifican los índices de las reacciones en las que se hace knock-out para cada condición (ko\_rxns). Esta última matriz en la simulación que se propone no se utiliza porque no se hacen knock-outs de ninguna reacción, únicamente pretendemos simular el crecimiento de *Synechocystis* en diferentes condiciones de cultivo. Además el programa pide como entradas una matriz con los flujos de entrada de nutrientes o energía al sistema, y de salida de subproductos o deshechos (ex\_flux) donde cada fila tiene una reacción de intercambio y cada columna una condición, y un vector con los índices que corresponden a estas reacciones (ex\_rxns). Por último pide un vector con los flujos de crecimiento para cada condición, es decir con la producción de biomasa, porque supone



que esta es la reacción objetivo, si no lo fuera, habría que introducir los flujos de la reacción objetivo. No es necesario introducir el índice de la reacción objetivo porque el modelo tiene un parámetro que define cuál es esta reacción.

```
function [V_v] = IOMA(model, exp_mat, met_frw_mat, met_bck_mat, rxns_idx, ko_rxns, ex_flux, ex_rxns, GR)
```

Figura 5. Primera línea del programa en la que define qué entradas necesita.

Es importante saber que en la matriz de flujos de las reacciones de intercambio no se exige que se introduzcan todos los flujos de todas las reacciones de entrada y de salida, sólo de aquellas de las que se tengan datos. Aquellas de las que no se tienen datos las trata como una reacción más del sistema y solo tendrán flujo si son necesarias para el crecimiento o la formación de algún subproducto. Para aquellas entradas o salidas del sistema para las que no se pueden obtener las mediciones del experimento, se puede recurrir a la bibliografía para obtener ciertos límites naturales que limitan la capacidad máxima y mínima de flujo de esa reacción de acuerdo a la ecuación 9.

El resultado de cada simulación es una matriz con los flujos de tantas filas como reacciones hay en el modelo y tantas columnas como condiciones. Para modelos grandes resulta imposible trabajar con una matriz tan grande por lo que se ha desarrollado en este trabajo una función en Matlab (Fig. 1, anexo I) que permite simplificar la matriz escogiendo sólo los valores de aquellas reacciones que el usuario desee.

## 3.2. Modelo

El algoritmo requiere ser alimentado con un modelo en formato SBML que contiene la matriz estequiométrica (diferente a todas las anteriormente mencionadas), los límites superiores e inferiores de cada reacción y el nombre del modelo.

En este caso se utilizaron 3 modelos, 2 de ellos modelos ficticios, un modelo simplificado que nos permite explicar el funcionamiento del FBA, y un modelo de juguete (en adelante llamado Toy\_model) para permitirnos analizar mejor el funcionamiento del programa. El modelo simplificado está representado en la Figura 1 del anexo II, y el Toy\_model\_2 en la Figura 6. El modelo simplificado resultó un modelo demasiado simple por lo que se ha añadido en el Anexo I como ejemplo para la explicación del FBA básico, método de análisis comúnmente utilizado con el que se pretende comparar el algoritmo IOMA.

El Toy\_model es un modelo de 20 reacciones, 7 de ellas de intercambio, y 11 metabolitos internos. Los vectores que definen sus restricciones base para los límites superior e inferior de cada reacción son dos vectores 20x1. Es un modelo con unos pocos metabolitos más que el anterior muchas de las reacciones llevan a la creación de D como se puede ver en la Figura 4 y en la tabla 1. D sería equivalente a la unidad de biomasa, ya que en el caso de ambos modelos juguete la reacción objetivo es la reacción que externaliza D.

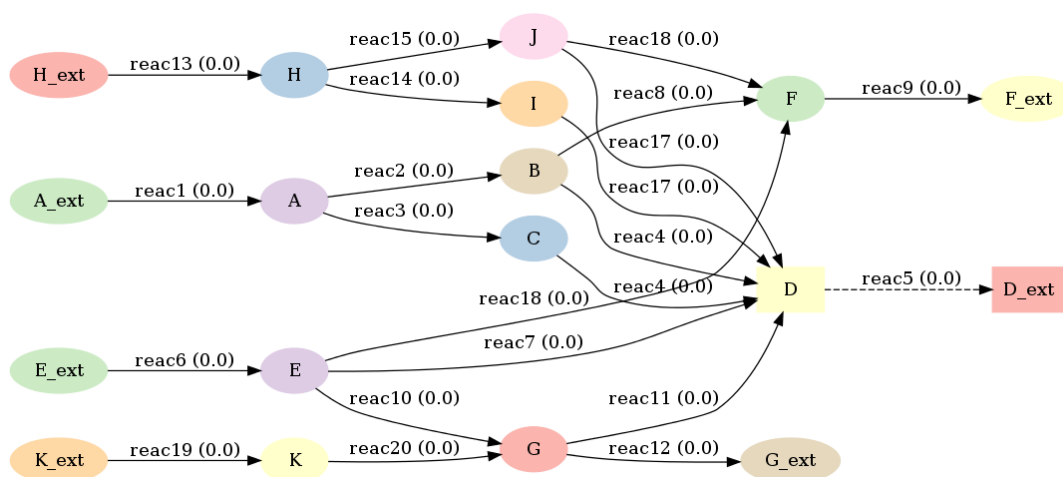


Figura 6. Representación del Toy\_model.

Nombre: reacción
Reac1: 1 A <sub>ext</sub> -> 1 A
Reac2: 1 A -> 1 B
Reac3: 1 A -> 1 C
Reac4: 1 B + 1 C -> 1 D
Reac5: 1 D -> 1 D <sub>ext</sub>
Reac6: 1 E <sub>ext</sub> -> 1 E
Reac7: 1 E -> 1 D
Reac8: 1 B + 1 E -> 2 F
Reac9: 1 F -> 1 F <sub>ext</sub>
Reac10: 1 E -> 1 G
Reac11: 1 G -> 1 D
Reac12: 1 G -> 1 G <sub>ext</sub>
Reac13: 1 H <sub>ext</sub> -> 1 H
Reac14: 1 H -> 3 I
Reac15: 1 H -> 1 J
Reac16: 1 I -> 2 D
Reac17: 1 J + 1 I -> 1 D
Reac18: 1 J + 1 E -> 1 F
Reac19: 1 K <sub>ext</sub> -> 1 K
Reac20: 1 K -> 1 G

Tabla 1. Lista de reacciones del modelo juguete 2. Las 9 primeras reacciones son las nueve reacciones del modelo juguete 1.

El tercer modelo, es el utilizado para la simulación de *Synechocystis 6803* fue reconstruido por Montagud et al. (2011) llamado iSyn811. Este modelo consta de 956 reacciones, que corresponden a 811 genes, 58 de ellas de transporte y 911 metabolitos. La reacción objetivo es la reacción de producción de biomasa. Los vectores para los límites inferiores y superiores son vectores de tamaño 956x1. En este modelo los nombres de las reacciones que están catalizadas por una enzima identificada corresponden al número EC de la proteína que las cataliza, lo que nos permite una rápida asociación reacción-proteína.

### 3.4. Modelo de *Synechocystis* 6803

#### 3.3.1. Datos proteómicos y metabolómicos

Como ya se ha dicho anteriormente, el modelado metabólico requiere de una gran aportación de datos experimentales por parte de los laboratorios. La idea original de este proyecto consistía en lanzar una simulación previa con datos bibliográficos y, si el programa demostraba funcionar y predecir mejor el comportamiento metabólico que el algoritmo FBA básico, lanzar una simulación con datos crudos, no analizados ni publicados y tratar de analizar el comportamiento completo del organismo en respuesta a ciertos estímulos, en lugar de una interpretación básica de la expresión de los componentes individuales o de rutas metabólicas aisladas. Por problemas logísticos en el laboratorio en el que se realizaban los experimentos, los datos no publicados no llegaron a tiempo para la presentación de este trabajo.

Además encontramos dificultades a la hora de encontrar referencias bibliográficas que encajaran con las necesidades del programa. Sólo dos artículos habían realizado en el mismo experimento análisis proteómicos y metabolómicos de los componentes y ninguno de los autores de ambos artículos han podido facilitarnos los datos.

Por lo tanto se ha tenido que recurrir a la opción menos deseable, tomar datos de dos artículos distintos que siguen la misma línea, pero son experimentos independientes. A pesar de no ser la situación ideal, creemos que estos datos pueden combinarse en este tipo de algoritmo, puesto que se trata del mismo tipo de estrés, y la idea es aportar un punto de referencia al que aproximar los flujos (y no condicionarlos totalmente) para obtener una distribución más realista bajo ciertas condiciones, en este caso bajo falta de nitrógeno. Ambos experimentos se hacen para averiguar los cambios que se producen en *Synechocystis* 6803 tras la exposición a falta de nitrógeno. Utilizan el mismo medio de cultivo y el mismo aporte de fotones en ambos experimentos. Sin embargo, en el experimento metabolómico llevado a cabo por Osanai et al. (2014) las células fueron sometidas a cuatro horas de exposición a un medio sin nitrógeno, mientras que en el experimento proteómico llevado a cabo por Huang et al. (2013) se miden las concentraciones de proteínas tras 0, 24 y 48 horas de exposición.

Una vez obtenidos los datos de estos dos experimentos, es necesario identificar las reacciones a las que pertenecen esos datos. En primer lugar, los identificadores de los datos proteómicos son nombres de genes (nomenclatura tipo slr0000 o slr0000) que han de ser convertidos a números EC. Para ello se ha escrito un programa en Python que, recurriendo a una base de datos local construida a partir de datos disponibles de KEGG, asocia el nombre del gen con la proteína y su número EC. Una vez determinada la asociación gen-proteína, es necesario asociar esta proteína con las reacciones del modelo que cataliza. En el caso del modelo de *Synechocystis* esta asociación es relativamente fácil porque las reacciones llevan como nombre el código EC de la proteína que las cataliza; de nuevo se empleó un script de Python que permite identificar los nombres de las reacciones a partir de los códigos EC. Una vez identificadas las reacciones se comprueba de qué metabolitos se dispone de concentraciones conocidas y si corresponden a metabolitos de las reacciones de interés. Lo ideal sería que para cada reacción con datos proteómicos se conocieran las concentraciones de todos los metabolitos que intervienen en la reacción. Pero si esto no es posible, con tener al menos un sustrato en las reacciones irreversibles, y un sustrato y un producto en las reacciones reversibles, es suficiente.

Los datos metabolómicos fueron aportados como áreas de pico del espectrómetro de masas y datos sobre los estándares internos. Posteriormente fueron traducidos a concentraciones. Los datos proteómicos se encuentran como ratios de proteínas tras 24 ó 48 horas de exposición con respecto a las concentraciones de referencia al inicio del experimento. Por tanto los ratios de proteínas al inicio del experimento serán uno en todos los casos y equivalen a la condición de referencia en la que la cianobacteria se encuentra sin estrés.

### 3.3.2. Datos cinéticos

Para hacer el cálculo de las velocidades por Michaelis-Menten es necesario conocer los valores de las constantes de disociación de la enzima. Lo más recomendable sería disponer de estos datos para cada sustrato y, en caso de que la reacción sea reversible, de cada producto. Pero como no se tienen todos esos datos, y tampoco se dispone de todas las concentraciones de metabolitos y de proteínas del modelo, se busca al menos una constante de disociación para un sustrato del que se tengan datos metabolómicos, de cada reacción que tenga datos proteómicos. En caso de ser reversible será necesaria como mínimo una  $k_m$  para uno de los sustratos y una  $k_m$  para uno de los productos. Estos parámetros cinéticos se obtienen de las bases de datos, como BRENDA y Sabio-RK, que aparecen explicadas en la tabla 2.

Base de datos	Descripción
<b>BRENDA</b>	Base de datos bioquímicos de las reacciones, de especial importancia los datos cinéticos que contiene.
<b>SABIO-RK</b>	Base de datos cinéticos de las reacciones bioquímicas.
<b>BIONUMBERS</b>	Base de datos que contiene medidas de todos los tipos para una gran variedad de organismos
<b>CyanoBase</b>	Base de datos de cianobacterias. Permite hacer BLAST con otras cianobacterias
<b>KEGG</b>	Base de datos de reacciones bioquímicas, no contiene parámetros cinéticos pero contiene una buena herramienta de visualización de las rutas metabólicas.

Tabla 2. Bases de datos utilizadas

Los parámetros cinéticos han sido escogidos siguiendo unas reglas. En primer lugar, si se dispone del valor cinético para *Synechocystis 6803* se elige éste. Si no se encuentra, se recurre a las constantes cinéticas de organismos relacionados preferiblemente unicelulares y fotosintéticos como *Chlamydomonas*, que es un alga unicelular relativamente estudiada o bien de alguna otra cianobacteria. Utilizando la base de datos CyanoBase, que permite hacer un BLAST de las secuencias proteicas de *Synechocystis 6803* en comparación con otras cianobacterias, se elige de entre las cianobacterias que tengan  $k_m$  para esa reacción, el valor de aquella con la que comparte mayor similitud proteica. En caso de que no exista ningún valor para *Synechocystis 6803* ni para ningún organismo relacionado, se observan las rutas metabólicas de las que forma parte la reacción, si alguna ruta está relacionada con la fotosíntesis entonces se escogen los valores de un organismo fotosintético aunque sea eucariota y pluricelular. Si la reacción no está relacionada con la fotosíntesis se escogen los parámetros del organismo procarionta unicelular que tenga información para todos los metabolitos de los que se conoce la concentración intracelular, si son muchos metabolitos este organismo suele ser *E.coli*.

En caso de tener más de un valor para elegir se hace un primer sesgo entre proteínas 'wild-type' y proteínas mutantes. Si aun así seguimos teniendo más de un valor se elige el valor que haya sido medido en unas condiciones más similares a las de los experimentos, es decir 30°C y pH 7'8.

### 3.3.3. Restricciones introducidas

El algoritmo IOMA requiere bastante precisión al introducir los flujos para las reacciones de intercambio porque el programa asume que todos los flujos de entrada que el usuario aporta son fijos, es decir que no permite fluctuaciones en los flujos de entrada aportados por el usuario. Por esta razón el cálculo de los flujos de las reacciones de intercambio ha de ser muy preciso o introducir pocos datos fijos sobre las entradas. Aquellas entradas o salidas que no han sido introducidas por el usuario como una variable del algoritmo están definidas por sus límites inferior y superior.

El modelo calcula flujos que se miden en milimoles de un sustrato que se consumen o producen por gramos de peso seco y por hora. La reacción de crecimiento es una reacción cuyas unidades están en horas<sup>-1</sup>. Los crecimientos se obtuvieron de las curvas de crecimiento calculadas por Huang et al.(2013), artículo del que obtuvimos los datos proteómicos. Para pasar los incrementos de OD a gramos de biomasa producida por hora se utilizaron conversiones que aparecen en la base de datos Bionumbers, que especifican que 10<sup>11</sup> células/L equivalen a una OD de 0,25 (Burrows et al., 2008) y 5 unidades de OD730 equivalen a 1 gramo de peso seco.

La cantidad de fotones y CO<sub>2</sub> se obtuvieron de Montagud et al. (2010) que calcularon qué cantidad de estas entradas limitantes del crecimiento se consideran habituales para un crecimiento normal. La cantidad de nitrato utilizado se obtuvo por minimización inversa, es decir, la cantidad de nitrato mínimo para este crecimiento.

Una vez se han definido las entradas se puede crear una matriz a mano, o programar un código que pregunta al usuario cuales son los datos y los ordena en la matriz. Los códigos presentes en las Figuras 2, 3 y 4 del anexo I fueron escritos para cumplir esta función, se encargan de ordenar las restricciones en los flujos de las reacciones de intercambio y de crecimiento y constituyen las matrices de entrada del programa.

## 3.4. Simulación FBA

Se pretende comparar el resultado que arroja el IOMA con las conclusiones extraídas de los datos proteómicos y metabolómicos, pero también con el resultado que arrojaría un FBA clásico al que se le imponen las mismas restricciones de flujos de entrada y de salida, para comprobar si el IOMA es realmente capaz de elegir la solución óptima más cercana a los flujos teóricos y realmente arroja un resultado cualitativa y cuantitativamente diferente al del FBA.

Para la simulación de un FBA simple se utiliza el mismo modelo y las mismas restricciones en las reacciones de intercambio de componentes, y se realiza utilizando una plataforma web llamada Cyanodesign, una herramienta muy intuitiva y fácil de usar que permite la simulación de un FBA utilizando cualquier modelo. Cyanodesign tiene además la ventaja de devolver como resultado un mapa de la distribución de flujos, que además puede personalizarse definiendo qué reacciones se desea visualizar.

## 4. RESULTADOS

En este apartado se procede a comentar detalladamente los resultados obtenidos. En primer lugar, se utilizó un modelo simplificado para explicar el funcionamiento de un FBA básico (explicado en el Anexo II), herramienta con la que vamos a comparar el algoritmo IOMA. En segundo lugar se aplican los algoritmos FBA y IOMA en un modelo básico al que hemos llamado Toy\_model. Por último, se realiza una simulación del comportamiento de *Synechocystis* aplicando y teniendo en cuenta el conocimiento obtenido del Toy\_model.

### 4.1. Simulación modelo juguete 2

Para analizar el comportamiento del IOMA en un modelo más simple, se ha elaborado el Toy\_model, un modelo de 20 reacciones con metabolitos ficticios. Este modelo carece de significación biológica pues la formación del metabolito objetivo no se obtiene de la suma de los metabolitos intermedios. En este caso varias reacciones independientes y que cuentan con metabolitos comunes, pueden dar lugar al metabolito D, cuya externalización está fijada como objetivo. Es decir, la externalización de D sería el equivalente a la externalización de la biomasa (crecimiento) en un sistema biológico. Sin embargo, se ha planteado el Toy\_model de esta manera para evitar añadirle restricciones adicionales y permitir una monitorización más simple.

Con este modelo se han realizado varias simulaciones en función de los datos que se querían obtener para tratar de comprender el funcionamiento del algoritmo y sus errores. Son datos inventados para guiar los resultados de las simulaciones. De igual forma se limitaron las entradas y salidas a voluntad con el fin de obtener conclusiones sobre el comportamiento del algoritmo y del modelo en cada situación. Se realizaron gran cantidad de pruebas con el fin de obtener la mayor cantidad de información posible. Algunos de los resultados se pueden observar en las Figuras 7 y 8 de este documento o en el segundo apartado del Anexo II.

En las Figuras 7 y 8 se puede comparar a entradas y salidas iguales el resultado que aporta un FBA sin datos internos, y el que aporta IOMA incorporando datos internos. En el ejemplo que se encuentra en este apartado las entradas de los metabolitos A y E, y la salida de D son fijas. El resultado obtenido utilizando el FBA (Figura 7) representa el camino más óptimo para obtener con esas entradas la salida de D necesaria. En el resultado obtenido por el IOMA (Figura 8) se habían utilizado unos datos de expresión de las proteínas y concentración de los metabolitos, que pretendían guiar la 3 a tener flujos muy bajos y la reacción 11 a tener flujos medios, con esto se obtiene una distribución en la que A y E son utilizados para la formación de F por medio de la reacción 8, mientras que K es introducido para la formación de D. Las Figuras 7 y 8 se han realizado utilizando los datos presentes en la Tabla 1 del Anexo II y corresponden a la segunda condición. Estos gráficos se han obtenido la plataforma Cyanodesign, elaborada por otros compañeros del proyecto.

De acuerdo a los datos introducidos esta distribución de flujos resultaría más cercana a la realidad que la aportada por el FBA. Por lo tanto una primera conclusión extraída de estas simulaciones es que IOMA aporta resultados más realistas, al tener en cuenta los datos internos. Las soluciones aportadas por el IOMA permiten más pérdidas en el sistema y dan

soluciones menos eficientes, lo cual resulta más similar a la realidad que las soluciones óptimas que aporta el FBA.

Además este análisis previo, realizando diferentes simulaciones con el Toy\_model (presentes en el Anexo II), permitió llegar a otras conclusiones sobre el funcionamiento de este algoritmo.

- **El IOMA es un algoritmo cuyos resultados son consistentes.** Para datos proteómicos y metabolómicos similares y diferentes condiciones de entrada y salida de nutrientes, siendo estas no limitantes, la solución obtenida presenta flujo en las mismas rutas. Esto se observa en la simulación presente en la tabla 1 del Anexo II.
- **El algoritmo IOMA aporta mejores datos cuanto mayor sea el número de condiciones.** Esto se debe a que el parámetro  $v_{max}$  es un parámetro libre de la optimización asignado a cada reacción de forma independiente pero siendo el mismo para todas las condiciones. Por tanto, a mayor número de condiciones, mejor se ajustará este parámetro.
- **Es posible que el algoritmo IOMA asigne un valor nulo a  $v_{max}$ .** Al simular en un mismo análisis varias condiciones y tratar de minimizar el flujo de una reacción en una de las condiciones, es posible que el algoritmo asigne un valor nulo a la  $v_{max}$  de esa reacción, impidiendo que haya flujo en esta reacción en otras condiciones.

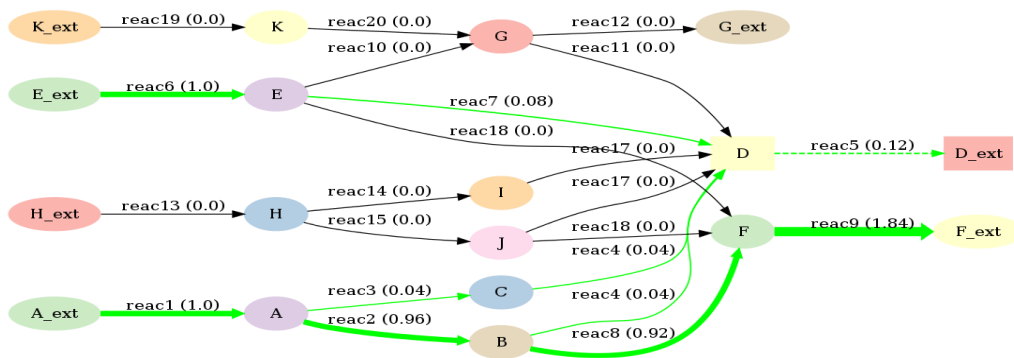


Figura 7. Resultado utilizando un algoritmo tipo FBA

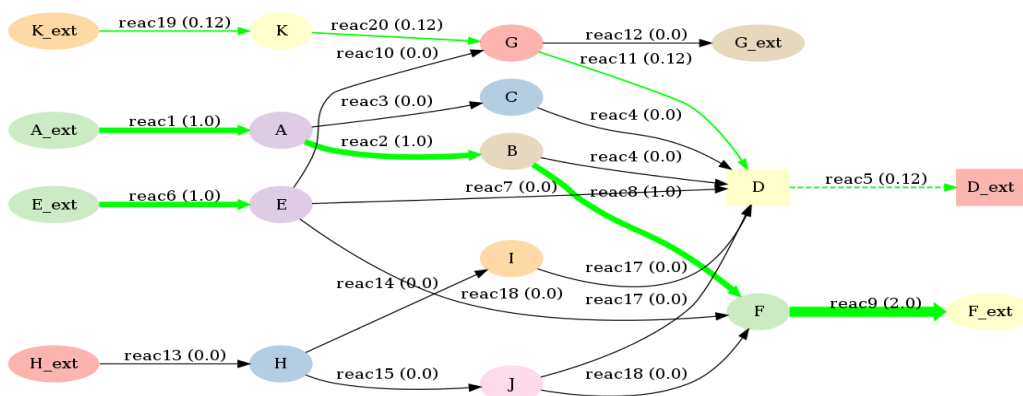


Figura 8. Resultado obtenido utilizando el algoritmo IOMA.

## 4.2. Simulación con modelo de *Synechocystis 6803*

Desde el desarrollo de las técnicas ‘high-throughput’ se ha evolucionado mucho en la comprensión de los mecanismos de supervivencia y respuesta a estrés de muchos organismos,

entre ellos *Synechocystis* 6803. Por ejemplo, se han hecho análisis del transcriptoma, el proteoma y el metaboloma de *Synechocystis* 6803 en diferentes condiciones o bajo diversos tipos de estrés ambiental. Sin embargo esto nos da una visión parcial de lo que está ocurriendo en el organismo sometido a cierto estrés que deriva en unas conclusiones correctas pero parciales de los mecanismos de regulación. Esto se debe a que se analizan reacciones independientes o rutas metabólicas pero no es posible analizar el comportamiento de todo el sistema. Por eso surge una necesidad creciente de contar con herramientas bioinformáticas precisas que integren estos datos en el modelo metabólico acercando la distribución de flujos óptima a una distribución de flujos real del organismo en esas condiciones. Por esta razón se ha decidido evaluar el funcionamiento de este algoritmo en un modelo de *Synechocystis* 6803 utilizando en la medida de lo posible los datos reales.

El comportamiento de *Synechocystis* 6803 como cianobacteria ante la falta de nitrógeno, nutriente esencial para su crecimiento, ha sido estudiada en diversas ocasiones (Osanai et al., 2006; Von Webeser et al., 2011; Krasikov et al., 2012). Utilizando las conclusiones de los mecanismos de respuesta al estrés de previos estudios se pretende comparar la simulación del algoritmo de interés con los resultados aportados por un FBA normal.

La simulación del comportamiento de *Synechocystis* 6803 con el IOMA se hace comparando la condición de referencia, crecimiento con nitrógeno a tiempo 0, con las condiciones de crecimiento en ausencia de nitrógeno, tras 24 y 48 horas. Se realizaron de forma independiente tres simulaciones, una en la que se compara el comportamiento de *Synechocystis* con nitrógeno y tras 24 horas sin nitrógeno, otra en la que se compara las condiciones de presencia de nitrógeno y 48 horas en ausencia de nitrógeno y, por último, una simulación en la que se comparan estas tres condiciones simultáneamente. En este documento se presentan los datos que corresponden a esta última simulación en la que se comparan los 3 sets de datos, y las simulaciones que realiza el FBA para estas mismas condiciones. Los resultados de las simulaciones en las que se comparan dos condiciones están presentes en el Anexo III.

#### 4.2.1. Acumulación de glucógeno

Osanai et al. (2006) aseguran que *Synechocystis* 6803 sufre una acumulación de glucógeno bajo condiciones de falta de nitrógeno. En la simulación realizada, en condiciones con nitrógeno se obtiene una salida neta de glucógeno utilizando el FBA mientras que el algoritmo IOMA indica que no hay externalización del glucógeno en condiciones de presencia de nitrógeno. Tras 24 horas en ausencia de nitrógeno, la salida del mismo utilizando el FBA disminuye desde un flujo de 0,006 (mmoles/gDW·h) a 0,001 (mmoles/gDW·h), mientras que el IOMA nos indica que esta salida se dispara y obtiene un flujo neto de 0,0012 (mmoles/gDW·h) de glucógeno. Tras 48 horas en ausencia de nitrógeno ambos algoritmos predicen una disminución de la salida de glucógeno (Figura 9). Esta disminución del flujo de externalización de glucógeno puede deberse a que la acumulación de glucógeno ocurra en las primeras fases de ausencia de nitrógeno y posteriormente el flujo de producción de glucógeno disminuya. Sin embargo, el glucógeno acumulado total aumentaría aunque a una velocidad menos. Aunque es posible que cuantitativamente estos datos no reflejen con precisión la realidad, es dato indicativo para saber que el algoritmo IOMA es capaz de predecir con más precisión que un FBA básico esta reacción.

Teniendo en cuenta los datos anteriores, se puede tratar de analizar qué ocurre en la ruta de la síntesis de glucógeno y porqué se producen estas diferencias en su salida del sistema. En la



Figura 10 está representada esta ruta de síntesis de glucógeno y en la Figura 11, se observan los flujos obtenidos en cada reacción para la simulación conjunta de las tres condiciones. En la Tabla 1 del Anexo III se pueden verificar las reacciones representadas en la Figura 10, tal y como están escritas en el modelo. Los resultados de las dos simulaciones que constan de dos condiciones cada una también están presentes en el Anexo III, y se puede observar que cualitativamente, los resultados obtenidos son similares, aunque cuantitativamente se encuentran grandes diferencias al compararlos.

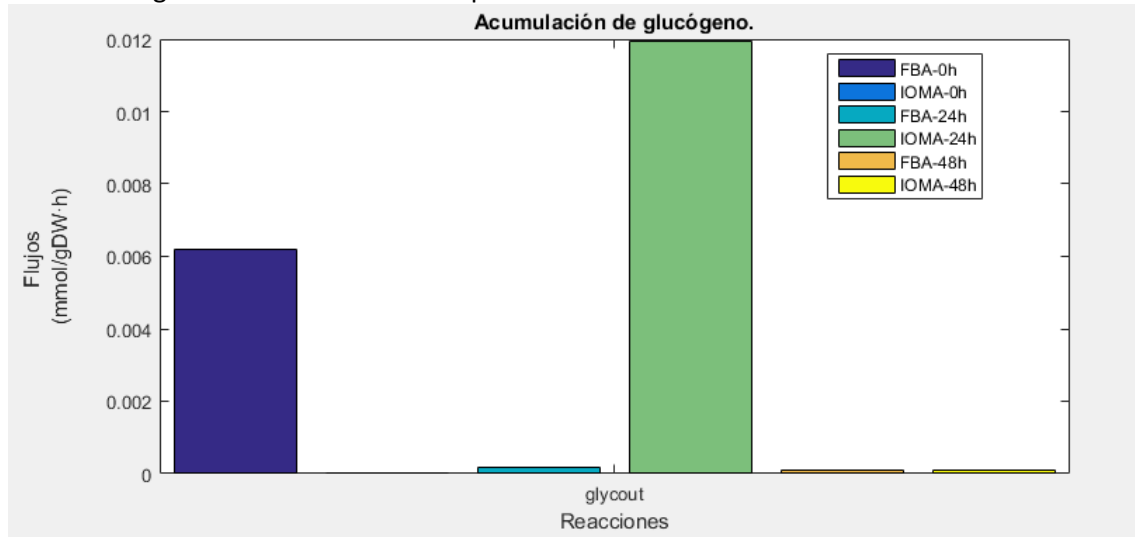


Figura 9. Predicción de la salida de glucógeno al exterior utilizando los dos algoritmos en ambas condiciones.

En la Figura 11 se encuentran representados los flujos para las reacciones en las que el glucógeno está involucrado. En la Figura 1 del Anexo III, se pueden observar estos mismos resultados con la salvedad de que algunos de los resultados de las simulaciones del IOMA en las condiciones de 24 y 48 horas en ausencia de nitrógeno obtienen flujos de 8 y 20 (mmol/gDW·h) esto se debe a que son ciclos fútiles, en los que un metabolito se convierte a un producto, y al final de la ruta un producto vuelve a dar este metabolito inicial. En la Figura 10 están representados los flujos netos de las reacciones.

Glycbm es la reacción que transforma el glucógeno a glucógeno de la reacción de biomasa. En el modelo, esta conversión se ha añadido para poder monitorizar la cantidad de glucógeno necesaria en la función de biomasa. El glucógeno se encuentra en la reacción de biomasa pese a que no es necesario para generar la misma. Esto se debe a que el glucógeno se acumula de forma natural en *Synechocystis 6803*, y dado que en un FBA no se puede acumular ningún componente en el organismo, se le añade a la función de biomasa.

En la reacción de biomasa se especifica la cantidad de glucógeno que acumula *Synechocystis 6803* de forma natural y la reacción de externalización de glucógeno (glycout), representa el flujo extra de glucógeno que se acumula en el interior del sistema.

Como se puede observar en la figura 10, la reacción 2.4.1.1, es la reacción que transforma el polímero de glucógeno en una cadena más corta y unidades de  $\alpha$ -D-glucosa. Esta reacción es la que utilizan los ciclos fútiles para reconvertir el glucógeno a glucosa. Sin embargo, una vez extraídos estos datos erróneos encontramos que el flujo de reconversión a glucosa es nulo.

Estos resultados no sólo nos indican que el algoritmo IOMA es capaz de aportar resultados más cercanos a la realidad, sino que además es capaz de aportar resultados sobre qué ruta se utiliza en cada condición. Esto se observa claramente en los resultados tras 48 horas sin nitrógeno (Figura 11), en los que el FBA predice la formación de glucógeno ('glycogen\_n1') a

través de las reacciones *glycsyn* y 2.4.1.21b, mientras que el IOMA indica que en esta condición el glucógeno se está formando a través de la reacción *glycsyn* y 2.4.1.18, esta última funcionando de forma inversa.

Este resultado indica que el algoritmo IOMA es capaz de darnos una simulación más realista que el FBA. Una de las razones más obvias es la cantidad de glucosa que llega a esta ruta, sobretudo en la primera condición. El FBA aporta la solución más óptima de todo el sistema, por lo que previsiblemente todas las rutas estarán optimizadas y todavía hay una cantidad importante de glucosa que es redirigida a esta ruta. Tras 24 y 48 horas en condiciones de estrés ya no hay tanta glucosa disponible y por tanto la producción de glucógeno disminuye. El IOMA asume un uso de la energía menos eficiente por lo que hay una menor disponibilidad de glucosa, y aun así una producción de glucógeno más realista.

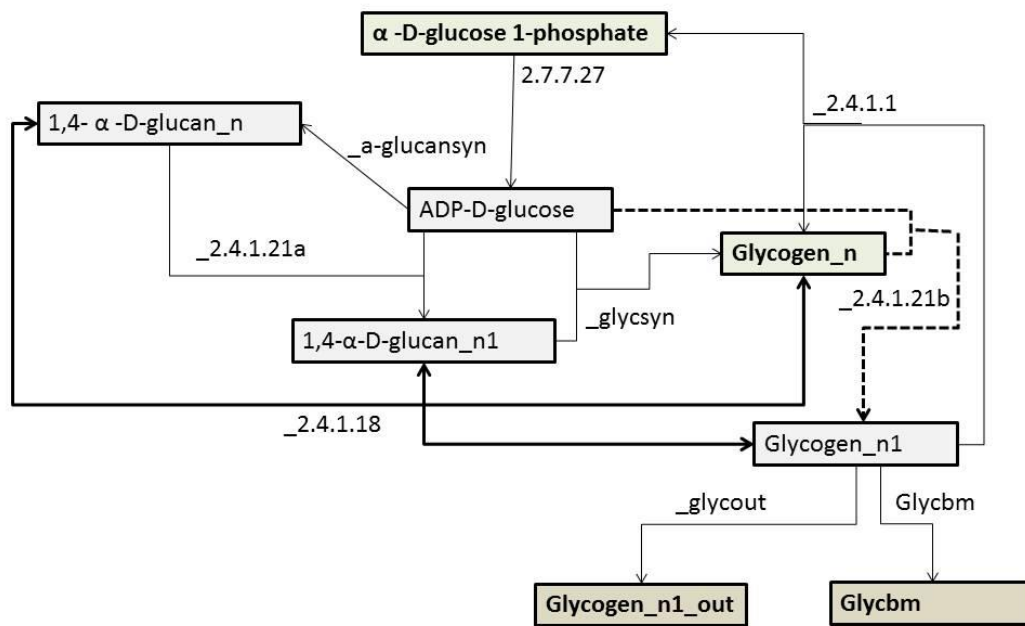


Figura 10. Representación gráfica de las reacciones en las que interviene el glucógeno.

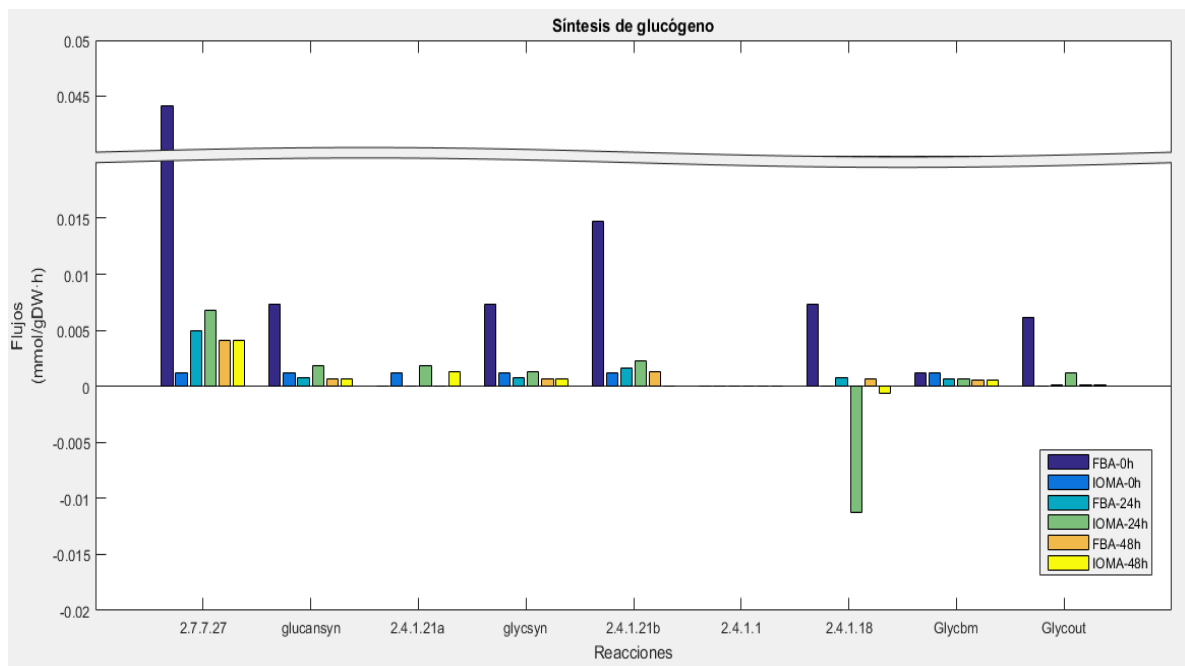


Figura 11. Flujos que arrojan los diferentes algoritmos para las reacciones en las que interviene el glucógeno.

### 4.2.2. Ciclo del ácido cítrico

La ruta del ácido cítrico, comúnmente conocida por ciclo de Krebs y también denominada ruta de los ácidos tricarbóxicos, es una ruta aerobia muy importante ya que forma parte de la respiración. Se suele describir su comienzo por una molécula de oxaloacetato y una molécula de Acetyl-coenzima A (a partir de ahora llamado acetyl-coA) y se producen diversas modificaciones hasta reciclar otra vez el oxaloacetato consiguiendo la producción de energía y CO<sub>2</sub>. Algunos organismos parecen no contar con el ciclo completo de acuerdo con los datos de su genoma, y parecen utilizarlo para sintetizar moléculas intermedias. Este parece ser el caso de *Synechocystis 6803*. En la Figura 12 se puede observar el ciclo del ácido cítrico al completo y, en verde, aparecen señaladas las enzimas de las que se tiene constancia de su presencia en *Synechocystis 6803*. En la Figura 14 las reacciones que no se encuentran en el modelo se encuentran representadas en rojo con un asterisco. En la tabla II del anexo se pueden comprobar las reacciones representadas en la Figura 14.

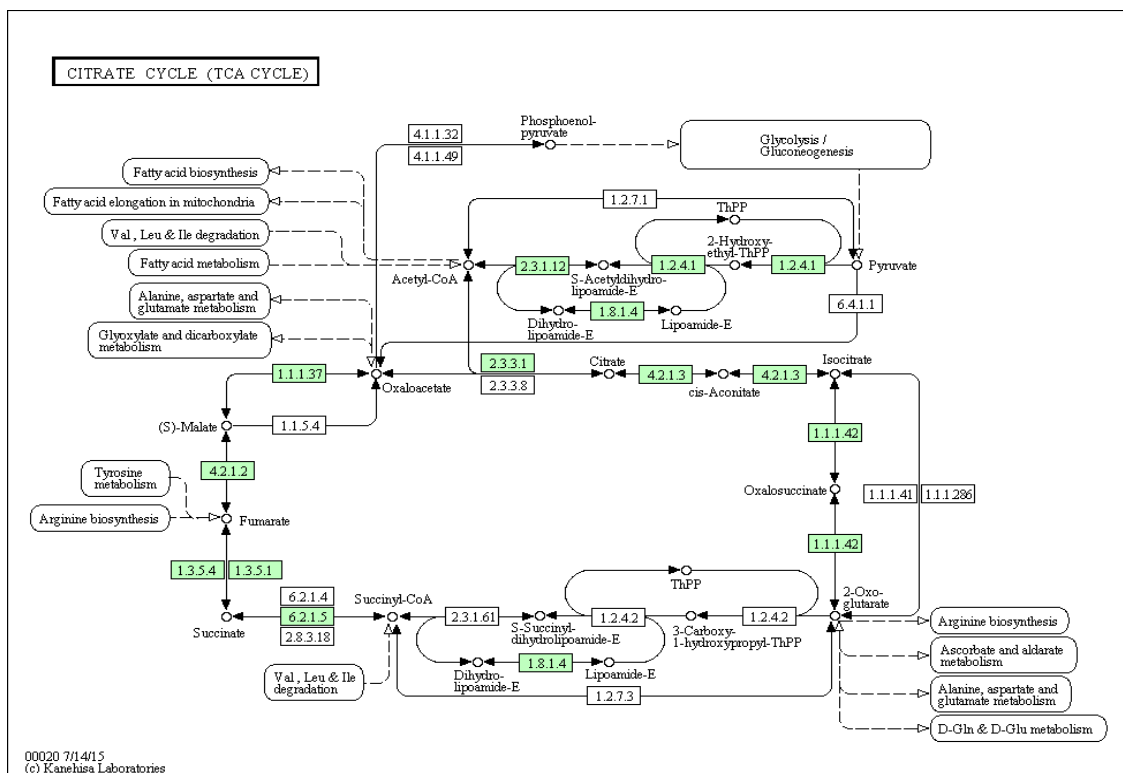


Figura 12. Ciclo del ácido cítrico. En verde aparecen las enzimas descritas en *Synechocystis*. Imagen extraída de la base de datos KEGG.

Estudios previos sobre el ciclo de Krebs sugieren que este ciclo se encuentra sobreexpresado en *Synechocystis 6803* en condiciones de estrés por falta de nitrógeno, y describen un aumento del nivel de 2-oxoglutarato (Muro-Pastor et al., 2001), lo que se confirma en el estudio del metaboloma del que se han obtenido los datos (Osanai et al., 2014). En este experimento de Osanai et al., también se apunta un crecimiento importante de los niveles de succinato, malato y fumarato. Sin embargo, los autores del artículo del que se han cogido los datos proteómicos (Huang et al., 2013) observan una disminución de la concentración de las enzimas que forman parte de este ciclo a las 24 horas.

En las simulaciones realizadas con el algoritmo IOMA se utilizan ambos datos, obteniendo como resultado un aumento de los flujos de este ciclo cuando la cianobacteria se cultiva en medio sin nitrógeno respecto al crecimiento en presencia de nitrógeno (Figura 13). Esto se debe a que las concentraciones de algunos de los metabolitos del ciclo del ácido cítrico como

el succinato o el malato habían aumentado entre 6-8 veces las concentraciones con respecto a sus concentraciones cuando *Synechocystis* se cultiva en un medio con nitrógeno.

En la Figura 13 se pueden observar los flujos calculados por el FBA y el algoritmo IOMA para las 3 condiciones de crecimiento. En todas las figuras de este documento los datos de crecimiento con nitrógeno son anotadas como los resultados a cero horas, esto se debe a que los datos fueron tomados en ambos artículos una vez alcanzada la OD deseada y antes de poner a *Synechocystis* 6803 en ausencia de nitrógeno. Los datos tras 24 y 48 horas en ausencia de nitrógeno se anotas como 'método'-24h y 'método'-48h respectivamente. Además, los asteriscos presentes en algunas de las reacciones de la Figura 13 indican que estas reacciones han sido invertidas en la gráfica con respecto a las reacciones presentes en la tabla 2 del Anexo III, cuyo orden indica que las direcciones positivas de las reacciones son aquellas que van desde el fumarato al 2-oxo-glutarato.

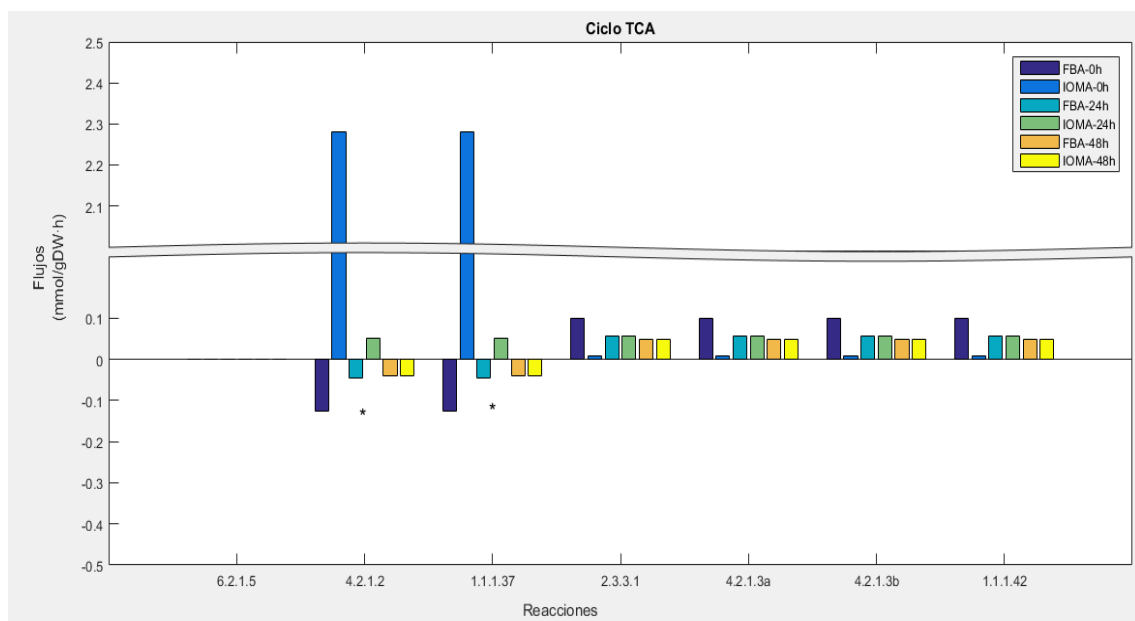


Figura 13. Flujos que arrojan los diferentes algoritmos para las reacciones en las que interviene del ciclo del ácido cítrico.

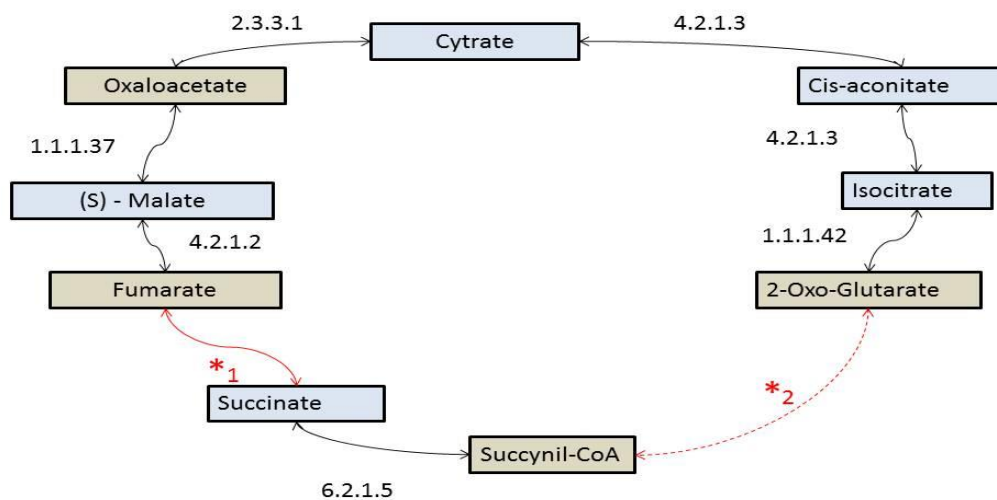


Figura 14. Representación gráfica del ciclo del ácido cítrico.

De esto se puede deducir que el funcionamiento de este algoritmo se ajusta bien a grandes cambios en la concentración de metabolito, pero la simulación realizada no permite saber si realmente simulando las mismas condiciones con concentraciones de metabolitos obtenidas

bajo estrés por falta de nitrógeno durante 24 horas, se observaría un descenso de la actividad de este ciclo o si, por el contrario, la menor concentración de enzima se ve compensada por una mayor actividad de las mismas. Sin embargo, aunque no es posible saber cuál es la respuesta correcta de *Synechocystis 6803* tras 24 horas en ausencia de nitrógeno, sí resulta claro que el algoritmo IOMA ha permitido simular mejor la respuesta metabólica, ya que se observa un aumento en los flujos tras 24 y 48 horas en ausencia de nitrógeno, en comparación al resultado obtenido por el IOMA en presencia de nitrógeno. El FBA por el contrario predice una disminución de esta ruta en ambas condiciones de falta de nitrógeno con respecto a la condición en presencia de nitrógeno.

Tras 48 horas bajo condiciones de falta de nitrógeno se observa una reducción de la actividad del ciclo del ácido cítrico con respecto a la condición de 24 horas en ausencia de nitrógeno, resultado obtenido tanto utilizando el FBA como utilizando el IOMA y consistente con lo observado por Huang et al. (2013).

Una observación interesante a realizar con los resultados de la Figura 13 es que el algoritmo IOMA predice un flujo alto en condiciones con nitrógeno o tras 24 horas sin nitrógeno de conversión de oxaloacetato a fumarato (reacciones 1.1.1.37 y 4.2.1.2). Esto es indicativo de unos requerimientos superiores de fumarato en presencia de nitrógeno o en cortas ausencias del mismo que tras 48 horas sin nitrógeno, momento en el que el fumarato está siendo degradado para contribuir a la formación de 2-oxoglutarato.

En las Figuras 6 y 7 del Anexo III se pueden observar los resultados de las dos simulaciones independientes de dos condiciones.

#### 4.2.3 Ruta de la glicolisis

La glicolisis es la ruta encargada de transformar la glucosa en piruvato para su posterior transformación a otros compuestos como acetyl-coA o etanol. En la Figura 15 se puede observar esta ruta y su completa caracterización en *Synechocystis*. En verde están todas las enzimas de esta ruta presentes en *Synechocystis*, que suponen la gran mayoría de las que forman la ruta. En la Figura 17 se encuentran representadas las reacciones utilizadas en los gráficos para analizar esta ruta metabólica.

Osanai et al.(2014) describieron un aumento de la actividad glucolítica tras las primeras cuatro horas de exposición a estrés por falta de nitrógeno. A su vez, Huang et al. (2013) describieron una reducción de la actividad de esta ruta en condiciones de estrés por falta de nitrógeno. En la Figura 16 se pueden observar la distribución de flujos en las diferentes reacciones y condiciones. Es importante tener en cuenta que las reacciones en las que encontramos flujos muy altos no son realistas, se trata de ciclos fútiles o ciclos de conversión de glucosa en subproductos y vuelta a glucosa, sin embargo, puede suceder que hay formación de subproductos indeseada, lo que altera la distribución de flujos de todas o la mayoría de reacciones de aquellos productos que intervienen en estas reacciones. Un ejemplo de esta formación de subproductos se puede observar en la reacción 6.2.1.1, en la que el acetyl-coA se disocia en acetato y coenzima A formando además una molécula de ATP, a partir de AMP como se puede observar en la Tabla 3 del Anexo III. Esta clase de ciclos supone un problema debido a la formación de subproductos y a la regeneración de moléculas energéticas como el ATP. La razón y solución a este problema se encuentran comentadas en la discusión.

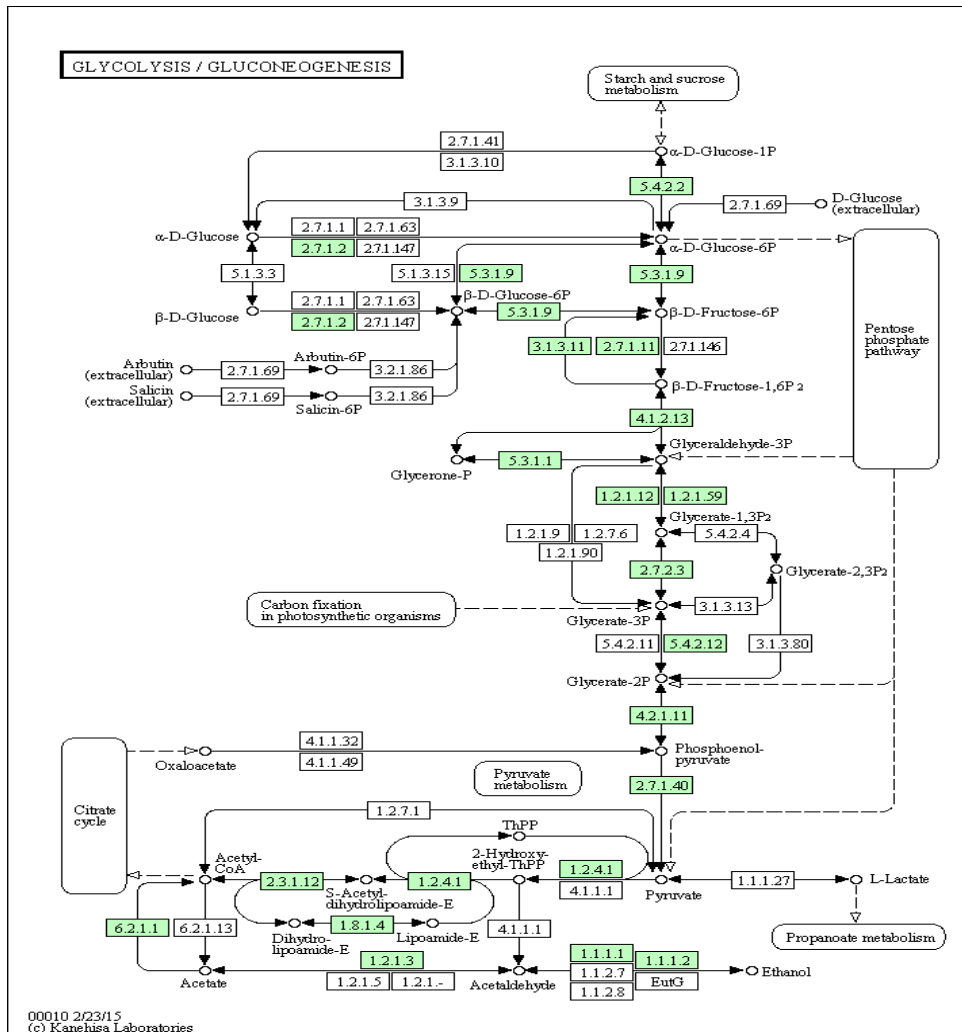


Figura 15. Ruta de la glucólisis. En verde aparecen las enzimas descritas en *Synechocystis*. Imagen extraída de la base de datos KEGG.

Por lo tanto, debido a la carencia de significación biológica y a la reconversión de estos subproductos en varias reacciones independientes, lo que hace imposible saber el flujo neto de cada reacción, se van a obviar estas reacciones en los comentarios de los resultados. La Figura 16 muestra una reducción de la actividad glucolítica en *Synechocystis* 6803 que es predicha tanto por el FBA como por el IOMA correctamente, aunque es más gradual en los datos arrojados por el IOMA (reacciones de la 5.4.2.12 a la 2.3.1.12), mientras que los datos que se obtienen utilizando el FBA son muy similares en ambas condiciones de falta de nitrógeno. Estos resultados sugieren un flujo neto de producción de Acetyl-coA, mientras que las reacciones 2.7.2.3 y 1.2.1.12 obtienen flujos inversos altos, que sugieren que las reacciones están desviadas a la producción de gliceraldehído-3P, que previsiblemente volverán a glucosa. Esta vuelta a la glucosa puede ser observada en la Figura 22 de la ruta de las pentosas fosfato. En esta figura se observa que la reacción 2.2.1.2 tiene un flujo alto. Esta reacción reconvierte el gliceraldehído-3P en glucosa como se puede observar en la Tabla 5 del Anexo III.

Sin embargo, aquellas reacciones que no se ven afectadas por los ciclos anteriormente comentados, son consistentes con las conclusiones de Huang et al., (2013).

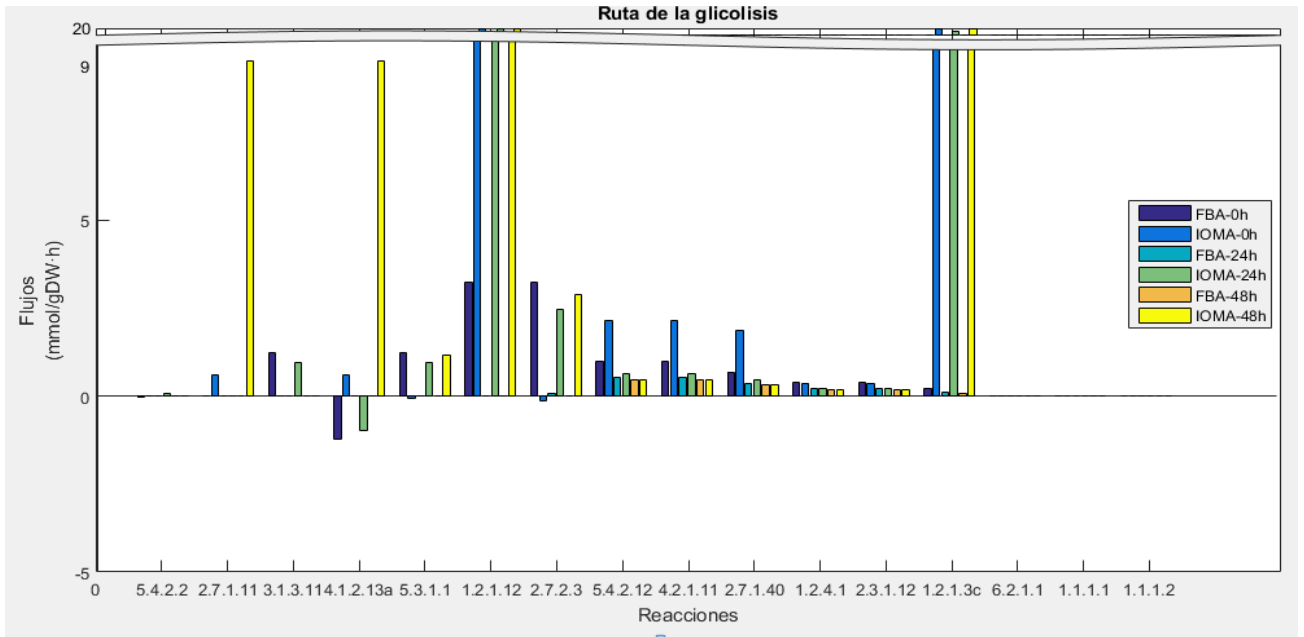


Figura 16. Distribución de flujos de la ruta de la glucólisis tras 24 horas en condiciones de falta de nitrógeno.

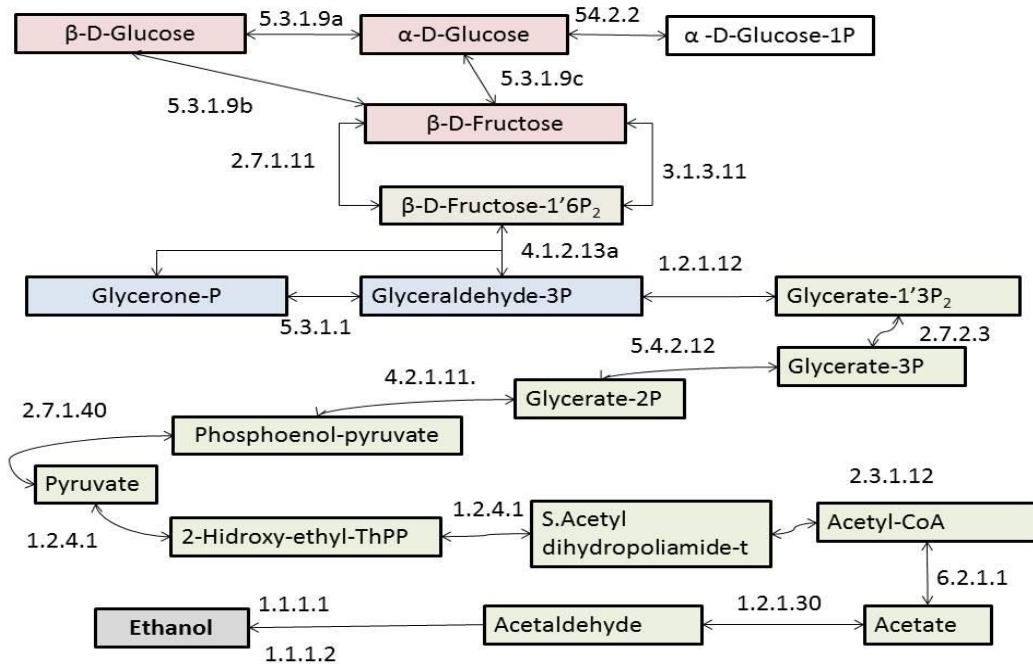


Figura 17. Reacciones escogidas para evaluar la ruta de la glicólisis.

En las Figuras 8, 9 y 10 del Anexo III, se pueden observar los resultados obtenidos de las simulaciones de dos condiciones por separado. Al observar estos datos, resulta evidente que los resultados en ausencia de nitrógeno aportados por el IOMA en las Figuras 9 y 10 reflejan flujos mucho mayores, lo que refuerza la idea de que se obtienen unas simulaciones más correctas y realistas utilizando una mayor cantidad de condiciones.

#### 4.2.4. Metabolismo del piruvato

El metabolismo del piruvato (Figura 18) consiste en las reacciones que tienen lugar para convertir otros sustratos de otros ciclos en piruvato, o viceversa. El piruvato es sustrato de otras rutas como la glicólisis o la síntesis de algunos aminoácidos como la Valina o la Leucina. No se conoce la respuesta de esta ruta a la ausencia de nitrógeno pero el estudio actual permite observar el comportamiento de esta ruta en condiciones de estrés por falta de nitrógeno.

Tal y como se muestra en la Figura 19, algunas de las reacciones que forman el metabolismo del piruvato carecen de flujo, lo que podría ser explicado o bien porque los productos de estas reacciones no sean necesarios para el crecimiento o no intervengan en otras rutas, o bien porque estos metabolitos se están produciendo en otras rutas, siendo esta una ruta prescindible.

Tres de las reacciones representadas en la Figura 19 tienen flujo en estas condiciones. La reacción 2.7.1.40 corresponde a la conversión de fosfoenolpiruvato a piruvato, siendo esta positiva probablemente debido a grandes requerimientos de este metabolito en otras rutas.

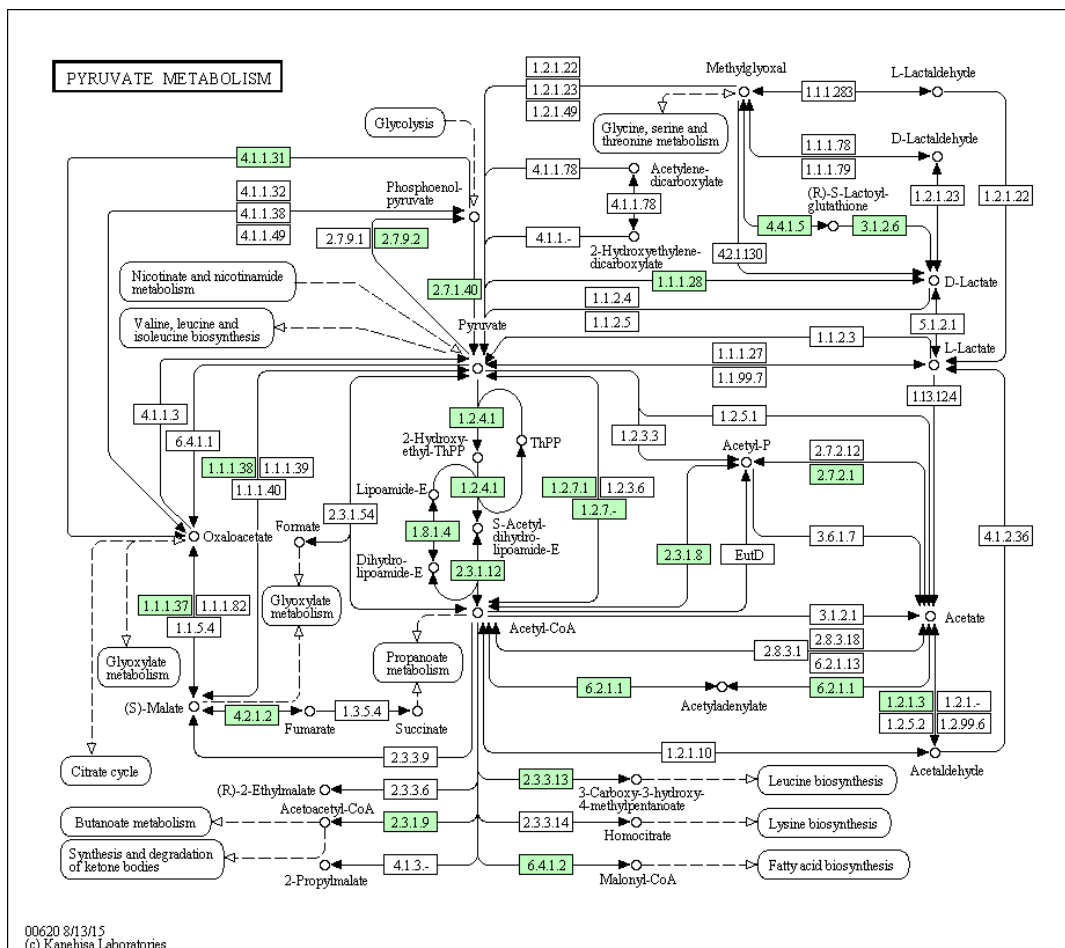


Figura 18. Reacciones metabólicas que relacionan el piruvato, producto de la glicólisis, con otras rutas metabólicas como la síntesis de algunos aminoácidos o el ciclo de Krebs. Imagen extraída de KEGG.

La reacción 4.1.1.31, que supone la conversión de fosfoenolpiruvato a oxaloacetato es positiva en todos los casos, aunque los resultados del IOMA indican una ligera inducción de esta reacción tras 24 horas en ausencia de nitrógeno, mientras que el FBA indica que esta reacción se encuentra reprimida en ausencia de nitrógeno y, en estas condiciones, se mantiene



constante con el tiempo. Por último, la reacción 1.1.1.37 presenta resultados desiguales según el método de análisis utilizado. Utilizando el IOMA el flujo de la reacción en presencia de nitrógeno y tras 24 horas de ausencia de nitrógeno resulta en una conversión de oxaloacetato a malato que, según los resultados reflejados en la Figura 13, se convierte a fumarato. Al utilizar el FBA, los resultados parecen indicar una conversión de malato a oxaloacetato. Von Webeser et al. (2012) han descrito en condiciones de ausencia de nitrógeno una mayor síntesis de algunos aminoácidos, como el aspartato, que utiliza el fumarato para su síntesis. También puede deberse a su utilización en otras rutas.

Los resultados de las simulaciones de dos condiciones se encuentran en las Figuras 11 y 12 del Anexo III. Resultan especialmente interesantes los resultados donde las 4 últimas reacciones si tienen flujo hacia la producción de piruvato en los resultados aportados por el algoritmo IOMA en condiciones de 48 horas en ausencia de nitrógeno.

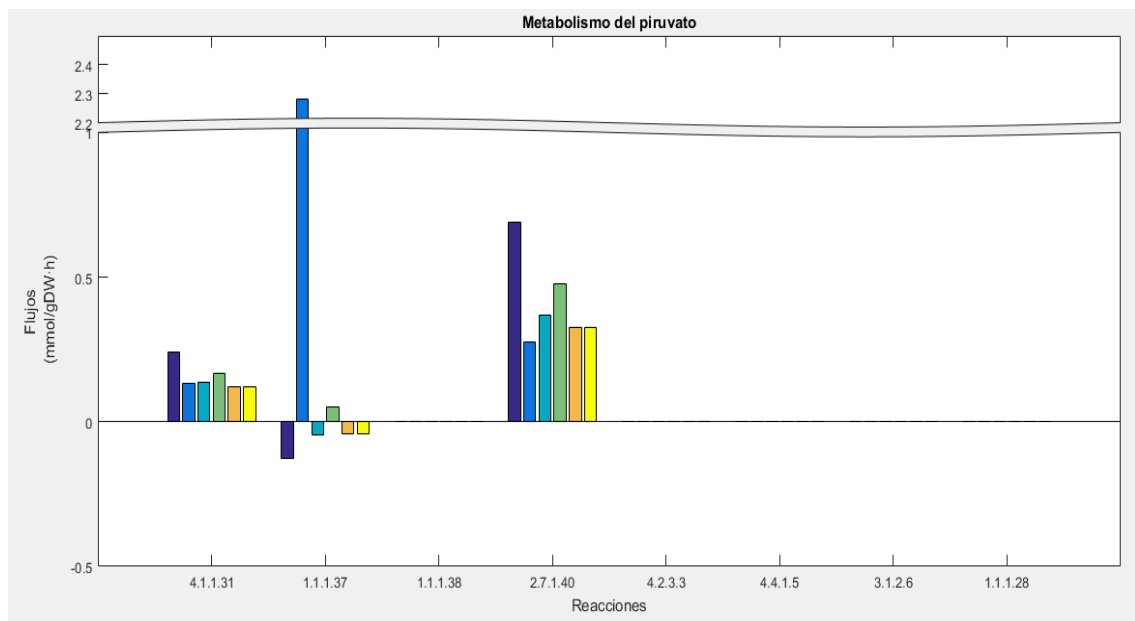


Figura 19. Distribución de flujos de algunas de las reacciones relacionadas con el piruvato según el método de análisis utilizado, para las diferentes condiciones.

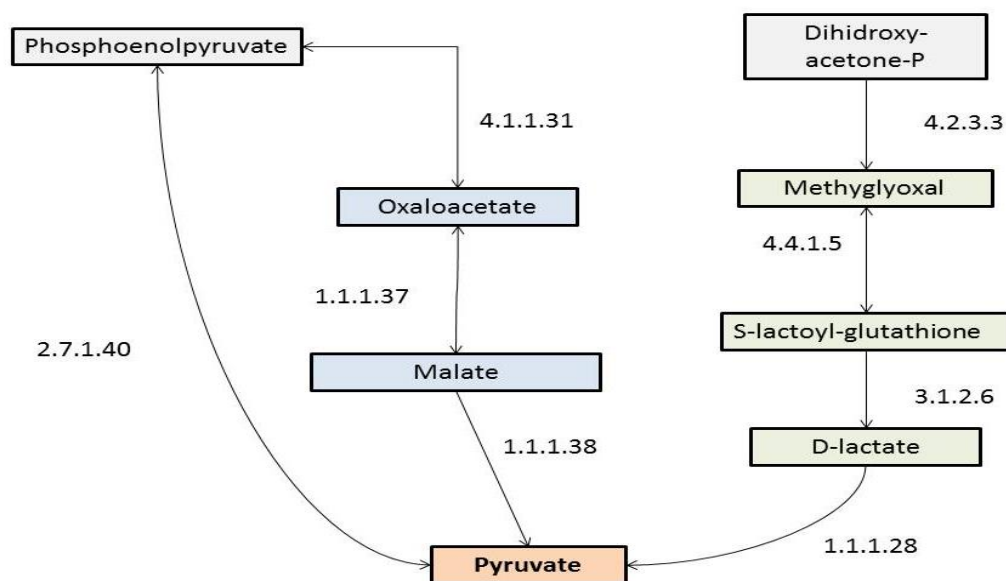


Figura 20. Esquema del metabolismo del piruvato.

#### 4.2.5. Ruta de las pentosas fosfato

La ruta de las pentosas fosfato es una ruta que utiliza como sustrato la glucosa y produce NADPH (o sus equivalentes reductores) y pentosas, pues son esenciales para la formación de nucleótidos. Hay dos fases principales en esta ruta, la primera es una fase oxidativa en la que la glucosa se convierte a ribulosa-5-fosfato generando NADPH, y la otra es una fase reversible no oxidativa en las que los azúcares fosforilados se transforman en pentosas fosfato como la xilulosa-5-fosfato, la eritrosa-5-fosfato y la ribosa-5-fosfato.

Algunos componentes derivados de las pentosas fosfato, como el fosforibosil pirofosfato, se utilizan en la síntesis de aminoácidos, por lo que, previsiblemente el flujo de algunas reacciones de esta ruta, tendrá relación con la necesidad de estos compuestos que tenga la célula. Esta ruta se encuentra representada en las Figuras 21 y 23.

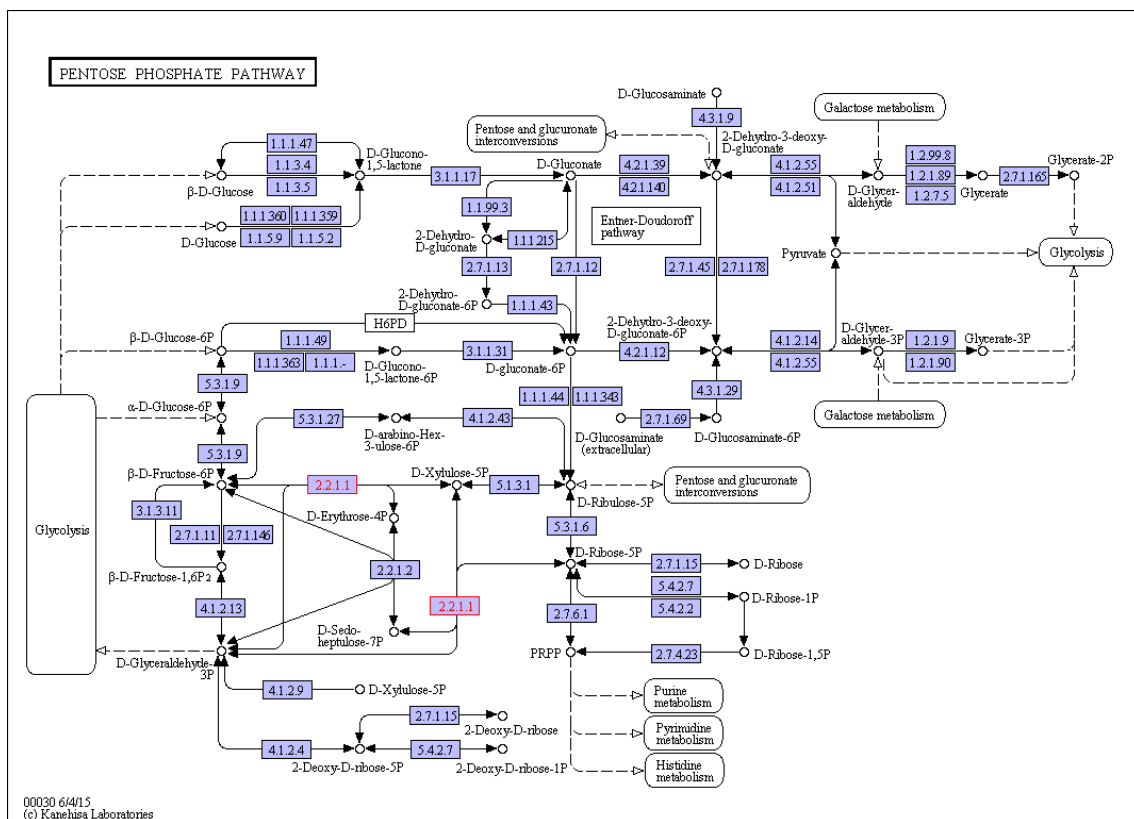


Figura 21. Ruta completa de las pentosas fosfato.

En la Figura 22 se pueden observar los resultados obtenidos al aplicar los dos métodos analíticos de interés en las tres condiciones presentadas. En general, el FBA predice una disminución drástica de los flujos de las reacciones de esta ruta, en condiciones de falta de nitrógeno. Sin embargo, el algoritmo IOMA parece indicar una inducción de esta ruta. Los flujos de las reacciones 2.7.1.11, 4.1.2.13a y 2.2.1.2 en la condición de 48 horas tras la retirada del nitrógeno, forman parte de un ciclo de conversión de glucosa en otros sustratos y reconversión de estos nuevamente a glucosa que ha sido comentado con anterioridad.

El aumento de esta ruta predicho por el IOMA parece responder a necesidades del sistema de pentosas fosfato, ya que la síntesis de fosforibosil pirofosfato (reacción 2.7.6.1) no resulta tan inducida como las reacciones anteriores. El fosforibosil pirofosfato (PRPP) es un sustrato de la síntesis de algunos aminoácidos, por lo que su aumento tras 24 horas sin nitrógeno puede responder a una mayor síntesis de estos aminoácidos. Por otro lado, la ribulosa-5P es un

sustrato muy utilizado tanto en el ciclo de Calvin como en la síntesis de aminoácidos. El aumento de la actividad de las reacciones 5.1.3.1 y 5.3.1.6 (de forma inversa) parecen responder a unos requerimientos mayores de este metabolito en estas rutas.

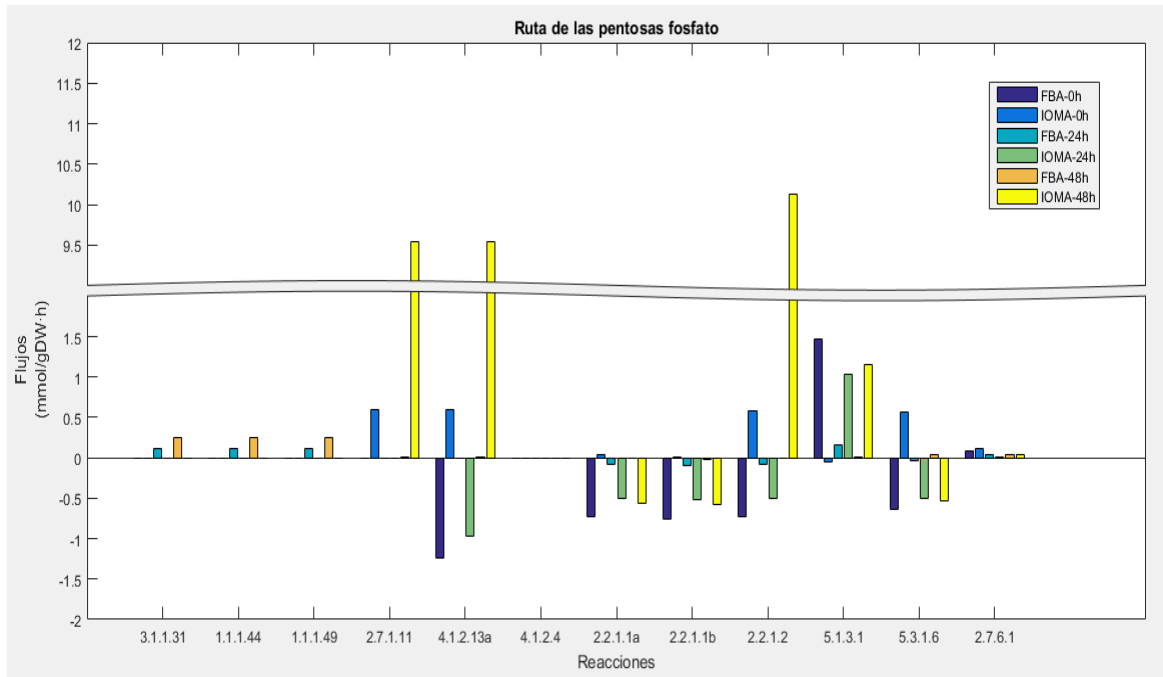


Figura 22. Distribución de flujos en esta ruta en las 3 condiciones estudiadas.

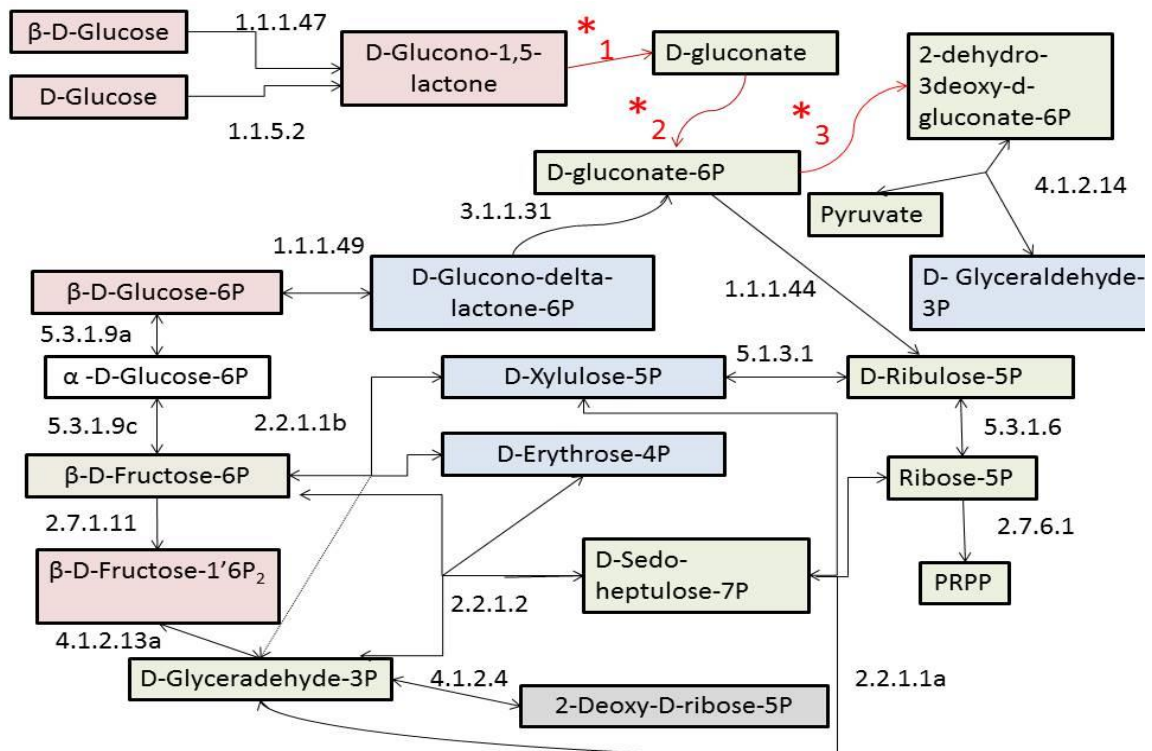


Figura 23. Esquema que representa la ruta de las pentosas fosfato.

En este caso el FBA da como resultado flujos muy pequeños de esta ruta y parece indicar que esta ruta está ligeramente reprimida por la ausencia de nitrógeno. Al no haber datos concluyentes sobre la respuesta de esta ruta en *Synechocystis 6803* a la falta de nitrógeno, no es posible saber cuál predice mejor la realidad.

#### 4.2.6. Síntesis de otros compuestos

Los autores de los artículos observaron una regulación diferencial dependiendo del aminoácido que se estuviera sintetizando. Sin embargo, esto no es comparable en el experimento que se ha realizado debido a la condición impuesta de estado estacionario. Esta condición indica que la cantidad de aminoácidos que se produce es igual a la que se consume y no admite acumulaciones más allá de la composición contemplada en la ecuación de biomasa. El flujo de producción o consumición de un aminoácido depende principalmente del crecimiento y su aportación a la producción de biomasa, por lo que los flujos de síntesis de aminoácidos no registrarán este cambio. Para comprobar esto se podría reformular la reacción de biomasa, teniendo en cuenta los niveles acumulativos de aminoácidos observados experimentalmente y repetir la simulación con esta reacción de biomasa, esperando cambios en los flujos de todo el sistema pero esto requeriría saber exactamente los nutrientes consumidos para producir esta cantidad de aminoácidos. Además, supondría prefijar los cambios esperados, lo que sería un claro caso de overfitting y no aportaría nueva información predictiva.

De igual forma Osanai et al. describen un descenso general en la producción de purinas y pirimidinas, siendo este mayor en las purinas y pirimidinas trifosfato que en las monofosfato donde la reducción es mínima o incluso inexistente. Es imposible monitorizar los flujos de estas moléculas, así como de las moléculas con poder reductor debido a que intervienen en demasiadas reacciones y, en algunos casos, están presentes en las reacciones que forman parte de ciclos fútiles o de ciclos de reconversión, por lo que en algunas reacciones sus flujos pueden ser muy grandes viéndose esto compensado en otras reacciones, no permitiendo la monitorización real de las velocidades a las que se utilizan estas moléculas.

## 5. DISCUSION

Este proyecto trata de analizar el algoritmo desarrollado por Yizhak et al. (2010) que integra datos proteómicos y metabolómicos utilizando datos cinéticos para una mejor predicción de los flujos y comparar su funcionamiento con el FBA.

En primer lugar, el algoritmo IOMA es un algoritmo basado en una cinética tipo Michaelis-Menten y asume que todas las reacciones que se restringen cumplen este tipo de cinética de orden uno. Esto supone añadir un error principalmente a aquellas reacciones que no sean de orden uno y sean tratadas como tal. Sin embargo se elimina el error de aquellas reacciones que no sigan una cinética tipo Michaelis-Menten pues no se encontrará su  $K_m$  lo que impedirá que sean usadas como restricción interna del algoritmo.

En segundo lugar, el algoritmo está formulado de tal forma que los datos en relación a intercambios con el exterior que le aporta el usuario son parámetros fijos, mientras que el FBA permite realizar una simulación indicando límites superiores e inferiores de estas reacciones, no obliga a darles un valor fijo. Esto añade un parámetro variable al FBA lo que puede restarle precisión pero le permite cierta flexibilidad. Esta flexibilidad resulta interesante, ya que normalmente las velocidades de toma de nutrientes de un organismo se obtienen a partir de medidas experimentales (propias o tomadas de la bibliografía), de la composición del medio o por minimización, lo que introduce cierto error. Esta desventaja del IOMA frente al FBA conlleva una pérdida de flexibilidad, por lo que aquellas reacciones de intercambio de las que no se esté relativamente seguro de su velocidad y que, en principio, no sean limitantes para el crecimiento, es conveniente no limitarlas.

Además, esto conlleva otro problema añadido, que ha sido puesto de manifiesto durante la realización de este proyecto. Limitar en exceso la entrada de nutrientes cuando se utiliza el algoritmo IOMA que ya incluye sus restricciones fijas y las condiciones de similitud, puede conllevar que no haya una solución óptima posible y por lo tanto, que el algoritmo IOMA, no devuelva ningún resultado. No limitar la entrada de nutrientes conlleva libertad para el organismo de coger lo que considere oportuno para cada simulación o condición.

En este proyecto se limitó la entrada de  $CO_2$  y de fotones porque son fuentes de carbono y energía, y por lo tanto son limitantes del crecimiento. También se limitó la entrada de nitratos porque se quería simular el comportamiento de *Synechocystis 6803* bajo condiciones de aporte de nitrógeno normales y deficientes. Sin embargo, el resto de entradas no estaban restringidas.

Los niveles de metales y nutrientes que internalizaba el organismo fueron revisados y resultaban coherentes con el crecimiento. Sin embargo uno de los problemas que tienen las simulaciones matemáticas de este tipo de sistema es que tratan de reconstruir un modelo que sea capaz de simular el comportamiento completo de un organismo cualesquiera que sean sus fuentes de materia y energía, es decir estén en el medio que estén. Por tanto reacciones posibles para el organismo pero poco naturales o que se dan de forma natural a flujos muy bajos, en un modelo metabólico pueden llegar a tener flujos muy grandes. Como ya se ha dicho anteriormente esto se resuelve poniendo límites 'naturales' a esa reacción, el problema aparece en aquellas entradas o salidas de las que se carece de cualquier dato orientativo.

Sin embargo, para evitar la entrada de algún componente o la formación de ciclos con flujos muy altos el FBA incluye el principio de parsimonia, que indica que de todas las soluciones

óptimas posibles se tome la que hace que la suma total de flujos sea mínima, dado que se entiende que la célula va a tender a hacer el mínimo esfuerzo posible. Este principio de parsimoneidad no se incluye en el algoritmo IOMA por lo que estas fugas del sistema para obtener materia o energía no se ven penalizadas.

En este caso sucedió con una reacción de internalización de peróxido de hidrógeno para la distribución de flujos en la que se simulaba con un aporte de nitrógeno normal, y con una reacción de internalización de protones en una reacción en la que se simulaba un aporte de nitrógeno bajo. Estas dos reacciones a flujos máximos conllevaron una salida de agua muy superior a la normal, y unos flujos en algunos casos muy por encima de los flujos habituales, aunque manteniendo la relación comparativa entre ambas condiciones ya que en ambos casos estaban afectando a las reacciones por igual. Además como se ha observado en varias ocasiones, las simulaciones realizadas con el IOMA en ocasiones tienen ciclos con flujos muy altos debido a la formación de ciclos fútiles, o ciclos grandes de conversión de metabolitos, en las que finalmente se reconvierte al metabolito inicial. Como se ha comentado anteriormente esto se debe a la falta del principio de parsimoneidad. Inicialmente se realizó una simulación del IOMA con los límites superiores e inferiores en 100 y -100 respectivamente, y después se establecieron en 20 y -20 para tratar de minimizar la aparición de estos flujos. Sin embargo, esto deberá hacerse penalizando las soluciones con flujos muy altos en la etapa de optimización. En la Figura 24 y 25 se muestra la dispersión de los datos de los flujos para el algoritmo IOMA y el FBA, respectivamente.

Como se puede observar en el histograma del algoritmo IOMA (Figura 24) la mayoría de los datos se encuentran alrededor de 0 pero hay un número importante de reacciones que se encuentran alrededor de los límites de flujos que les impone el algoritmo a aquellas reacciones no restringidas. En el histograma del FBA (Figura 25) se puede observar que no existen estos valores de flujos tan altos alrededor de los límites y que por tanto, la dispersión de los datos es mucho menor y hay menos flujos que tomen valores en torno a 20.

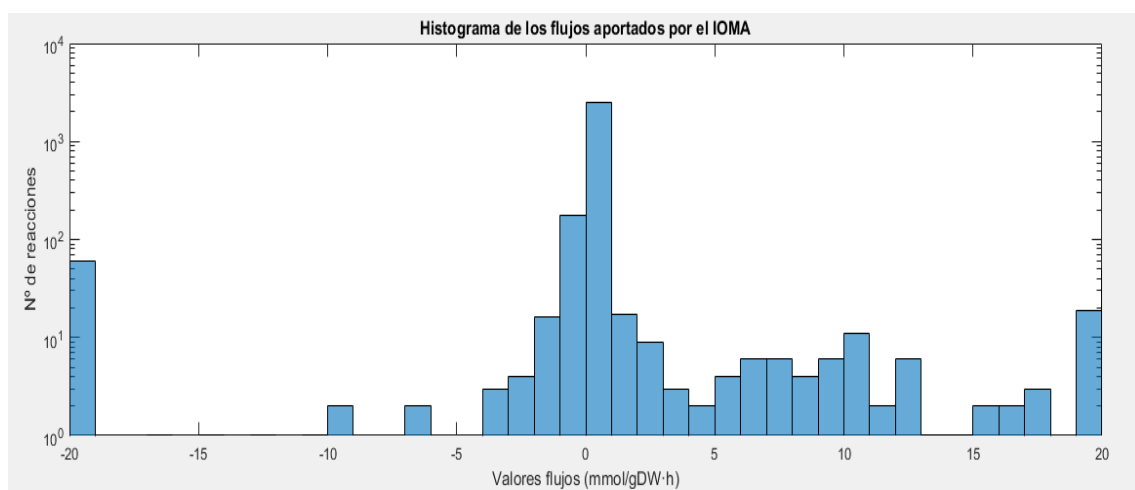


Figura 24. Histograma en escala logarítmica de la dispersión de los datos de flujo calculados utilizando el algoritmo IOMA.

Estos problemas de entrada de flujos grandes no deseados podría resolverse modificando el optimizador para penalizar aquellas soluciones que aun siendo óptimas incluyan flujos muy grandes poco realistas, limitando los flujos máximos y mínimos del programa a opciones más realistas, aunque esto podría conllevar una pérdida de sensibilidad, o realizando una segunda optimización para penalizar aquellas distribuciones con mayor suma de flujos.

Otra fuente de potencial error incluida en este algoritmo es la introducción de la  $v_{max}$  de la reacción en el optimizador del programa, porque el flujo final de esa reacción en cada simulación es distinto, como se ha comprobado, aun estando en las mismas condiciones. Por lo tanto, sería conveniente eliminar este parámetro de la optimización si fuera posible y darle un valor con más confianza cuando se dispusiera de él ( $k_{cat}$ ).

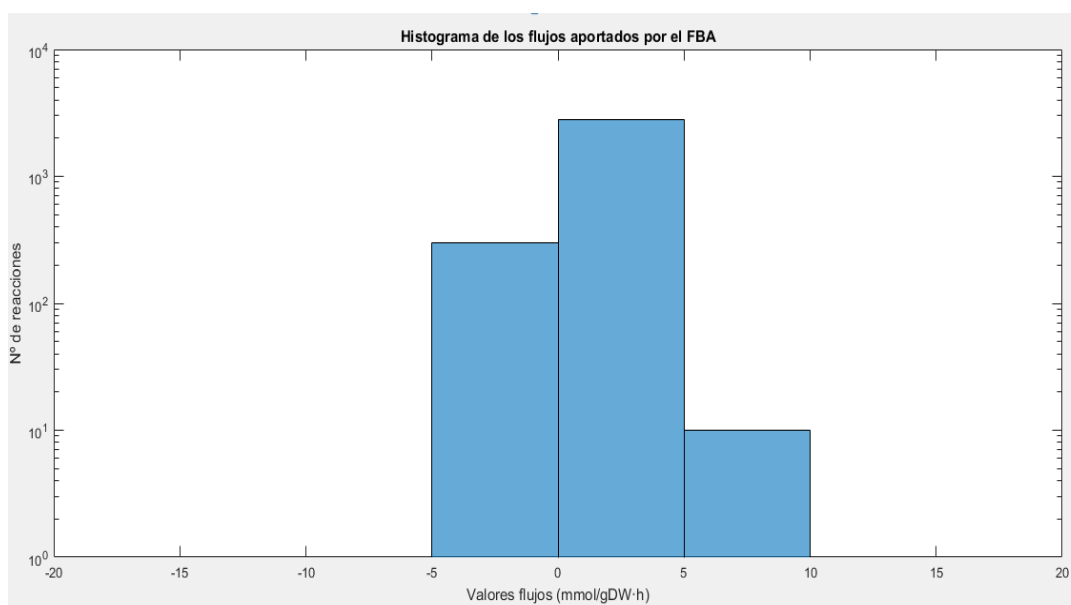


Figura 25. Histograma de la dispersión de datos de flujos calculados utilizando el FBA en escala logarítmica.

Sin embargo, aun teniendo en cuenta todas las limitaciones anteriores el algoritmo IOMA ha aportado datos que resultan más coherentes con las conclusiones bibliográficas revisadas.

1. Se ha observado que en todas las rutas anteriormente comentadas, el FBA daba valores de flujos muchos más pequeños que el IOMA. Esto se debe a que el FBA a partir de unos valores de entrada y de salida da una solución óptima que distribuya los flujos a través del sistema tomando el camino más óptimo por lo que si una reacción es innecesaria tendrá un flujo mínimo y entre dos reacciones equivalentes toma un camino 'al azar' dejando el otro sin flujo. Sólo tiene en cuenta las restricciones anteriormente explicadas. Esto no quiere decir que no sea sensible a cambios en el medio si no que calcula una solución óptima, dadas unas ciertas restricciones. Sin embargo, el algoritmo IOMA proporciona una solución óptima teniendo en cuenta el reparto interno de metabolitos en esa condición.
2. Además resulta evidente que el FBA aporta resultados donde la mayoría de las rutas dan lugar a flujos menores en condiciones de ausencia de nitrógeno, en comparación con la condición de presencia de nitrógeno. Esto puede deberse a que la concentración de nitrógeno aportada como entrada a ambos algoritmos se ha conseguido por minimización inversa, es decir, se ha calculado la concentración mínima de nitrógeno necesaria para que el algoritmo crezca lo que se le ha especificado. Esto implica que las condiciones de nitrógeno son limitantes pero quizás no sean tan bajas como lo serían en la realidad, y, por tanto, el FBA no está dando resultados que correspondan a la falta de nitrógeno si no simplemente el flujo de las rutas anteriormente explicadas a un crecimiento fijo. Por lo tanto si el crecimiento es menor, el flujo de las reacciones de síntesis de los metabolitos necesarios para el crecimiento es menor. Por el

contrario, el algoritmo IOMA es capaz de simular más fielmente la realidad al tener en cuenta los datos proteómicos y metabolómicos aportados, y puede plasmar cambios más allá de los directamente relacionados con el crecimiento.

Por lo tanto, el IOMA es una herramienta útil porque proporciona unas distribuciones de flujos en las que se observan cambios muy definidos en la regulación de las rutas metabólicas en función de las condiciones, pero requiere trabajo para conseguir que la distribución de flujos que aporte sea lo más realista posible para modelos grandes, complejos y cuyo conocimiento es limitado.

En futuros meses se pretende utilizar el conocimiento adquirido para adaptar este programa y tratar de minimizar los errores que se han planteado. Una vez realizadas las mejoras pertinentes, y comprobada la utilidad de este algoritmo, se pretende utilizar para analizar datos proteómicos y metabolómicos no publicados y así poder comprender la respuesta completa del sistema a cada condición.



## 6. CONCLUSION

El algoritmo IOMA desarrollado por Yizhak et al. (2010), es aplicable a modelos metabólicos grandes y poco estudiados pero requiere de una metodología para el cribado de datos y de conocimiento sobre la confianza que se tiene en los parámetros que se está introduciendo. Este algoritmo resulta útil para la comprobación de los cambios en los flujos metabólicos entre condiciones pero requiere de más trabajo para dar distribuciones de flujos comparables con la realidad.

De los resultados expuestos anteriormente también se concluye que la simulación realizada con el algoritmo IOMA resulta más realista en reacciones en las que los flujos no se ven afectados por altos flujos de entrada de nutrientes o de formación de ciclos con flujos muy altos. El FBA es un algoritmo que pese a que cuantitativamente sus resultados resultan más cercanos a la realidad, no refleja correctamente la respuesta interna de *Synechocystis 6803* a la ausencia de nitrógeno. La incorporación de datos proteómicos y metabolómicos que plantea el algoritmo IOMA le permite reflejar los cambios producidos en el metabolismo interno de *Synechocystis 6803* en condiciones de falta de nitrógeno, por lo que su aplicación resulta muy conveniente.

Teniendo en cuenta las limitaciones del algoritmo y de los datos utilizados los resultados obtenidos en este proyecto resultan esperanzadores para su futura utilización con datos originales. Además, si los datos obtenidos tras la mejora de este algoritmo resultan fiables cuantitativamente, esto permitiría obtener información sobre el comportamiento del organismo en su totalidad en las diferentes condiciones. Esta información podría ser utilizada en el planteamiento de experimentos o en la optimización de la producción de moléculas de interés.

Por lo tanto, el algoritmo IOMA resulta una herramienta muy útil para el análisis cualitativo del comportamiento de un organismo en respuesta a ciertos estímulos. Este algoritmo también aporta soluciones útiles en organismos de mayor tamaño y complejidad y cuya información disponible es menor. Además, en el futuro se pretende conseguir la mejora de este algoritmo para obtener una herramienta que permita un análisis cuantitativo fiable del funcionamiento de *Synechocystis 6803* bajo condiciones de interés.

## 7. REFERENCIAS

- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). *Molecular Biology of the Cell*. 4th edition.
- Agren, R., Bordel, S., Mardinoglu, A., Pornputtapong, N., Nookaew, I., & Nielsen, J. (2012). Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Computational Biology*, 8(5), e1002518. doi:10.1371/journal.pcbi.1002518
- Berg, H., Ziegler, K., Piotukh, K., Baier, K., Lockau, W., & Volkmer-Engert, R. (2000). Biosynthesis of the cyanobacterial reserve polymer multi-L-arginyl-poly-L-aspartic acid (cyanophycin): mechanism of the cyanophycin synthetase reaction studied with synthetic primers. *European Journal of Biochemistry / FEBS*, 267(17), 5561–70. doi:10.1046/j.1432-1327.2000.01622.x
- Blazier, A. S., & Papin, J. A. (2012). Integration of expression data in genome-scale metabolic network reconstructions. *Frontiers in Physiology*, 3(August). doi:10.3389/fphys.2012.00299
- Bruggeman, F. J., & Westerhoff, H. V. (2007). The nature of systems biology. *Trends in Microbiology*, 15(1), 45–50. doi:10.1016/j.tim.2006.11.003
- Chang, A., Scheer, M., Grote, A., Schomburg, I., & Schomburg, D. (2009). BRENDA, AMENDA and FRENDA the enzyme information system: New content and tools in 2009. *Nucleic Acids Research*, 37(SUPPL. 1), 588–592. doi:10.1093/nar/gkn820
- Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., ... Palsson, B. Ø. (2007). A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, 3(121), 1–18. doi:10.1038/msb4100155
- Giovannoni, S. J., Turner, S., Olsen, G. J., Barns, S., Lane, D. J., & Pace, N. R. (1988). Evolutionary relationships among cyanobacteria and green chloroplasts. *Journal of Bacteriology*, 170(8), 3584–3592.
- Görl, M., Sauer, J., Baier, T., & Forchhammer, K. (1998). Nitrogen-starvation-induced chlorosis in Synechococcus PCC 7942: Adaptation to long-term survival. *Microbiology*, 144(9), 2449–2458. doi:10.1099/00221287-144-9-2449
- Grigorieva, G., & Shestakov, S. (1982). Transformation in the cyanobacterium Synechocystis sp. PCC 6803. *FEMS Microbiology Letters*, 13, 367–370. doi:10.1111/j.1574-6968.1982.tb08289.x
- Huang, S., Chen, L., Te, R., Qiao, J., Wang, J., & Zhang, W. (2013). Complementary iTRAQ proteomics and RNA-seq transcriptomics reveal multiple levels of regulation in response to nitrogen starvation in Synechocystis sp. PCC 6803. *Molecular bioSystems*, 9(10), 2565–74. doi:10.1039/c3mb70188c
- Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Research*, 42(D1), D199–D205. doi:10.1093/nar/gkt1076
- Kaneko, T., Sato, S., Kotani, H., Tanaka, a, Asamizu, E., Nakamura, Y., ... Tabata, S. (1996). Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Research : An International Journal for Rapid Publication of Reports on Genes and Genomes*, 3(3), 109–136. doi:10.1093/dnares/3.3.109
- Kauffman, K. J., Prakash, P., & Edwards, J. S. (2003). Advances in flux balance analysis. *Current Opinion in Biotechnology*, 14(5), 491–496. doi:10.1016/j.copbio.2003.08.001
- Kitano, H. (2002). Systems Biology: A Brief Overview. *Science*, 295(5560), 1662–1664. doi:10.1126/science.1069492

- Knoll, A. H. (2008). Cyanobacteria and Earth History. *The Cyanobacteria: Molecular Biology, Genomics and Evolution*. Chapter 1.
- Krasikov, V., Aguirre von Wobeser, E., Dekker, H. L., Huisman, J., & Matthijs, H. C. P. (2012). Time-series resolution of gradual nitrogen starvation and its impact on photosynthesis in the cyanobacterium *Synechocystis* PCC 6803. *Physiologia Plantarum*, *145*(3), 426–439. doi:10.1111/j.1399-3054.2012.01585.x
- Montagud, A., Navarro, E., Fernández de Córdoba, P., Urchueguía, J. F., & Patil, K. (2010). Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium. *BMC Systems Biology*, *4*(1), 156. doi:10.1186/1752-0509-4-156
- Montagud, A., Zelezniak, A., Navarro, E., de Córdoba, P. F., Urchueguía, J. F., & Patil, K. R. (2011). Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium *Synechocystis* sp. PCC6803. *Biotechnology Journal*, *6*(3), 330–342. doi:10.1002/biot.201000109
- Muro-Pastor, M. I., Reyes, J. C., & Florencio, F. J. (2001). Cyanobacteria perceive nitrogen status by sensing intracellular 2-oxoglutarate levels. *The Journal of Biological Chemistry*, *276*(41), 38320–8. doi:10.1074/jbc.M105297200
- Nakao, M., Okamoto, S., Kohara, M., Fujishiro, T., Fujisawa, T., Sato, S., ... Nakamura, Y. (2010). CyanoBase: the cyanobacteria genome database update 2010. *Nucleic Acids Research*, *38*(Database), D379–D381. doi:10.1093/nar/gkp915
- Nelson, D. L., & Cox, M. M. (2005). *Lehninger principios de bioquímica*.
- Oltvai, Z. N. (2002). SYSTEMS BIOLOGY: Life's Complexity Pyramid. *Science*, *298*(5594), 763–764. doi:10.1126/science.1078563
- Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis? *Nature Biotechnology*, *28*(3), 245–248. doi:10.1038/nbt.1614
- Osanai, T., Azuma, M., & Tanaka, K. (2007). Sugar catabolism regulated by light- and nitrogen-status in the cyanobacterium *Synechocystis* sp. PCC 6803. *Photochemical & Photobiological Sciences : Official Journal of the European Photochemistry Association and the European Society for Photobiology*, *6*(5), 508–514. doi:10.1039/b616219n
- Osanai, T., Imamura, S., Asayama, M., Shirai, M., Suzuki, I., Murata, N., & Tanaka, K. (2006). Nitrogen Induction of Sugar Catabolic Gene Expression in *Synechocystis* sp. PCC 6803. *DNA Research*, *13*(5), 185–195. doi:10.1093/dnares/dsl010
- Osanai, T., Oikawa, A., Shirai, T., Kuwahara, A., Iijima, H., Tanaka, K., ... Hirai, M. Y. (2014). Capillary electrophoresis-mass spectrometry reveals the distribution of carbon metabolites during nitrogen starvation in *Synechocystis* sp. PCC 6803. *Environmental Microbiology*, *16*, 512–524. doi:10.1111/1462-2920.12170
- Rippka, R., Deruelles, J., Waterbury, J. B., Herdman, M., & Stanier, R. Y. (1979). Generic assignments, strain histories and properties of pure cultures of cyano- bacteria. *Journal of General Microbiology* *11*: 1-61. *Journal of General Microbiology*, *11*(2), 1–61.
- Santos, F., Boele, J., & Teusink, B. (2011). *A Practical Guide to Genome-Scale Metabolic Models and Their Analysis. Methods in Systems Biology* (1st ed., Vol. 500). Elsevier Inc. doi:10.1016/B978-0-12-385118-5.00024-4
- Schmidt, B. J., Ebrahim, A., Metz, T. O., Adkins, J. N., Palsson Bernhard Ø and Hyduke, D. R., Palsson, B. Ø., & Hyduke, D. R. (2013). GIM3E: condition-specific models of cellular metabolism developed from metabolomics and expression data. *Bioinformatics (Oxford, England)*, *29*(22), 2900–2908. doi:10.1093/bioinformatics/btt493
- Selinger, D. W., Wright, M. a, & Church, G. M. (2003). On the complete determination of biological systems. *Trends in Biotechnology*, *21*(6), 251–4. doi:10.1016/S0167-7799(03)00113-6
- Shahzad, K., & Loor, J. J. (2012). Application of Top-Down and Bottom-up Systems Approaches in Ruminant Physiology and Metabolism. *Current Genomics*, *13*(5), 379–394. doi:10.2174/138920212801619269

- Stanier, R. Y., & Cohen-Bazire, G. (1977). Phototrophic Prokaryotes: The Cyanobacteria. *Annual Review of Microbiology*, 31, 225–274. doi:10.1146/annurev.mi.31.100177.001301
- Stein, L. (2001). Genome annotation: from sequence to biology. *Nature Reviews. Genetics*, 2(7), 493–503. doi:10.1038/35080529
- Stephanopoulos, G. (1994). Metabolic engineering. *Current Opinion in Biotechnology*, 5(2), 196–200. doi:10.1016/S0958-1669(05)80036-9
- Stephanopoulos, G. (1999). Metabolic fluxes and metabolic engineering. *Metabolic Engineering*, 1(1), 1–11. doi:10.1006/mben.1998.0101
- Swingle, W. D., Blankenship, R. E., & Raymond, J. (2008). Insights into Cyanobacterial Evolution from Comparative Genomics. *The Cyanobacteria: Molecular Biology, Genomics and Evolution*. Chapter 2.
- Thiele, I., & Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, 5(1), 93–121. doi:10.1038/nprot.2009.203
- Von Wobeser, E. A., Ibelings, B. W., Bok, J., Krasikov, V., Huisman, J., & Matthijs, H. C. P. (2011). Concerted changes in gene expression and cell physiology of the cyanobacterium *Synechocystis* sp. strain PCC 6803 during transitions between nitrogen and light-limited growth. *Plant Physiology*, 155(3), 1445–57. doi:10.1104/pp.110.165837
- Wang, Y., Li, Y., Shi, D., Shen, G., Ru, B., & Zhang, S. (2002). Characteristics of mixotrophic growth of *Synechocystis* sp. in an enclosed photobioreactor. *Biotechnology Letters*, 24, 1593–1597. doi:10.1023/A:1020384029168
- Westerhoff, H. V., & Palsson, B. O. (2004). The evolution of molecular biology into systems biology. *Nature Biotechnology*, 22(10), 1249–1252. doi:10.1038/nbt1020
- Williams JGK (1988) Construction of specific mutations in photosystem II photosynthetic reaction center by genetic engineering methods in *Synechocystis* 6803. *Methods Enzymol* 167: 766-778
- Yizhak, K., Benyamini, T., Liebermeister, W., Ruppin, E., & Shlomi, T. (2010). Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics*, 26(12), i255–i260. doi:10.1093/bioinformatics/btq183
- Yoshihara, S., & Ikeuchi, M. (2004). Phototactic motility in the unicellular cyanobacterium *Synechocystis* sp PCC 6803. *Photochemical and Photobiological Sciences*, 3(6), 512–518.
- Yu, Y., You, L., Liu, D., Hollinshead, W., Tang, Y., & Zhang, F. (2013). Development of *Synechocystis* sp. PCC 6803 as a Phototrophic Cell Factory. *Marine Drugs*, 11(8), 2894–2916. doi:10.3390/md11082894