

**ESTIMACIÓN DE FUNCIONES OBJETIVO DE  
REDES METABÓLICAS BASADA EN  
COMPARTIMENTALIZACIÓN Y ANÁLISIS DE  
BALANCE DE FLUJOS**

CÉSAR AUGUSTO VARGAS GARCÍA

UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS  
ESCUELA DE INGENIERÍA DE SISTEMAS  
MAESTRÍA EN INGENIERÍA DE SISTEMAS E INFORMÁTICA  
BUCARAMANGA  
2012

# **ESTIMACIÓN DE FUNCIONES OBJETIVO DE REDES METABÓLICAS BASADA EN COMPARTIMENTALIZACIÓN Y ANÁLISIS DE BALANCE DE FLUJOS**

CÉSAR AUGUSTO VARGA GARCÍA

Monografía de grado para optar al título de:  
Master en Ingeniería de Sistemas

Director:

HENRY ARGUELLO FUENTES, *PhD(c), Mpe.*

Codirector:

RODRIGO TORRES SAEZ, *PhD.*

UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FISICOMECAÑICAS  
ESCUELA DE INGENIERÍA DE SISTEMAS  
MAESTRÍA EN INGENIERÍA DE SISTEMAS E INFORMÁTICA  
BUCARAMANGA  
2012

# Índice general

<b>Introducción</b>	<b>11</b>
<b>1. Análisis de Balance de Flujos - FBA</b>	<b>16</b>
1.1. Modelo Estequiométrico . . . . .	16
1.2. Análisis de Balance de Flujos - FBA . . . . .	19
1.3. Soluciones del FBA . . . . .	20
1.4. Análisis de Variabilidad de Flujos - FVA . . . . .	25
<b>2. Funciones Objetivo Basadas en Compartimentos de Microorganismos</b>	<b>26</b>
2.1. Funciones Objetivo Basadas en Compartimentos . . . . .	26
<b>3. Resultados</b>	<b>29</b>
3.1. FBA Tradicional usando la Reconstrucción iMM904 . . . . .	29
3.2. FBA usando Funciones Objetivo basadas en Compartimentos . . . . .	36
<b>4. Conclusiones</b>	<b>46</b>
<b>5. Apéndice A: Código usado en la producción de los resultados</b>	<b>48</b>

# Índice de figuras

1.1. Ejemplo de modelo estequiométrico. (a) grafo del sistema, (b) reacciones químicas, (c) matriz de incidencia, (d) matriz estequiométrica y (e) vector de velocidades de reacción. . . . .	18
1.2. Principales elementos de un problema FBA.(a) grafo del sistema, (b) matriz estequiométrica $\mathbf{S}$ , (c) vector de velocidades máximas $\beta$ y mínimas $\alpha$ , vector de función objetivo $\mathbf{c}$ y vector solución $\mathbf{v}$ encontrado al resolver el FBA. En (d) se representa gráficamente este resultado. El valor máximo de la función objetivo $f$ es 5. . . . .	21
1.3. valores de función objetivo para cada punto extremo del problema presentado en la Figura 1.2. En color negro están los puntos extremos $\mathbf{u}_k$ que tienen el valor máximo de función objetivo. . . .	23
1.4. ilustración del conjunto solución de la ecuación (1.6) del problema de la Figura 1.2. Izquierda: combinación lineal convexa de los puntos extremos que son solución de (1.6). Derecha: puntos extremos y valores de la combinación lineal convexa. Abajo: solución problema de la Figura 1.2. Las funciones objetivo evaluadas en los puntos extremos mostrados son iguales al máximo valor posible. . . . .	24

---

2.1. Sistema compuesto por 3 compartimentos: $P_1$ , $P_2$ y $P_3$ . Arriba, diagrama que representa el sistema. Abajo, funciones objetivo $\mathbf{C}_1^T \mathbf{v}$ , $\mathbf{C}_2^T \mathbf{v}$ y $\mathbf{C}_3^T \mathbf{v}$ de dichos compartimentos. La función objetivo del compartimento $P_1$ está compuesta por las velocidades de reacción $v_{10}$ y $v_{13}$ , resaltadas en el diagrama y en la función objetivo (cuadros negros del vector $\mathbf{C}_1^T$ ). . . . .	27
3.1. resultados de FBA y FVA bajo las condiciones experimentales de la Tabla 3.1. (a) FBA del crecimiento del sistema. FBA y FVA de (b) etanol, (c) glicerol, (d) succinato, (e) acetato, (f) piruvato y (g) $\text{CO}_2$ . . . . .	35
3.2. FVA bajo las condiciones experimentales de la Tabla 3.1 y valor de función objetivo igual al valor experimental de crecimiento. FVA de (a) etanol, (b) glicerol, (c) succinato, (d) acetato, (e) piruvato y (f) $\text{CO}_2$ . . . . .	43
3.3. resultados obtenidos mediante función objetivo maximización de biomasa, la matriz $\mathbf{C}$ encontrada y condiciones experimentales anaeróbicas. a) Comportamiento del crecimiento del sistema y b) comportamiento del etanol vs consumo de glucosa. . . . .	44
3.4. resultados obtenidos mediante función objetivo maximización de biomasa, la matriz $\mathbf{C}$ encontrada y condiciones experimentales aeróbicas tipo batch. a) Comportamiento del crecimiento del sistema y b) comportamiento del etanol vs consumo de glucosa. . . . .	45

# Resumen

TITULO: ESTIMACIÓN DE FUNCIONES OBJETIVO DE REDES METABÓLICAS BASADA EN COMPARTIMENTALIZACIÓN Y ANÁLISIS DE BALANCE DE FLUJOS.<sup>1</sup>

AUTOR: CÉSAR AUGUSTO VARGAS GARCÍA <sup>2</sup>

PALABRAS CLAVE: Análisis de Balance de Flujos, Reconstrucción a Escala Genómica, iMM904, *S. cerevisiae*, Función Objetivo basada en Compartimentos.

El microorganismo *Saccharomyces cerevisiae* cuenta con gran número de modelos biológicos conocidos como reconstrucciones, las cuales pueden ser a escala genómica. Una de las técnicas más usadas para estudiar estos modelos es el Análisis de Balance de Flujos (FBA). El propósito del FBA es predecir el crecimiento del microorganismo bajo estudio, y la producción y consumo de componentes como el etanol, CO<sub>2</sub>, glicerol, succinato, acetato y piruvato. Para determinar si las predicciones obtenidas mediante FBA son únicas se utiliza la técnica de Análisis de Variabilidad Flujos (FVA). El presente trabajo muestra los resultados de aplicar el FBA a la reconstrucción reciente del microorganismo *S. cerevisiae*, la denominada iMM904 y los compara con un conjunto de datos experimentales presente en la literatura. Este trabajo también estudia la existencia de múltiples predicciones FBA utilizando la técnica FVA. Los resultados ilustran que es posible predecir el crecimiento del microorganismo *S. cerevisiae*, con errores entre el 11 % y 28 %; la producción de CO<sub>2</sub>, con errores entre el 0.3 % y 4.5 % y la producción de etanol, con errores entre el 11 % y 13 %.

Una de las técnicas más utilizadas para la predicción de producción de bioproductos de microorganismos es el Análisis de Balance de Flujos - FBA. El FBA requiere de una función objetivo que represente el objetivo biológico. Se propone un nuevo tipo de funciones objetivo basada en la combinación de objetivos de compartimentos físicos presentes en el microorganismo estudiado. La función objetivo propuesta mejora las predicciones de crecimiento en un 10 % y las predicciones de producción de etanol en un 75 % respecto a las obtenidas por la función objetivo de maximización de biomasa, en condiciones anaeróbicas. En condiciones aeróbicas tipo batch la función objetivo propuesta mejora en un 98 % las predicciones de crecimiento y en un 70 % las predicciones de etanol con respecto a la función objetivo de biomasa.

---

<sup>1</sup>Trabajo de Grado

<sup>2</sup>Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de Sistemas. Director, Henry Arguello. Codirector, Rodrigo Gonzalo Torres Saez.

# Abstract

TITLE: ESTIMATION OF OBJECTIVE FUNCTIONS BASED ON MICROORGANISM COMPARTMENTS AND FLUX BALANCE ANALYSIS <sup>1</sup>

AUTHOR: CÉSAR AUGUSTO VARGAS GARCÍA <sup>2</sup>

KEYWORDS: Flux Balance Analysis, Genome Scale Reconstructions, iMM904, *S. cerevisiae*, Compartment based Objective Function

*S. cerevisiae* microorganism has a large number of biological models named reconstructions. These reconstructions can be genome scale. Mathematical models come from these reconstructions, these models are called Stoichiometric. One of the most used tools for analysis of these models is Flux Balance Analysis (FBA). The objective of FBA is to predict growth and consumption/production of components like glucose, ethanol, glycerol, succinate, acetate and pyruvate. To determine if the predictions made by FBA are unique there is a technique named Flux Variability Analysis (FVA). This work shows results of applying FBA using the reconstruction iMM904 of *S. cerevisiae* microorganism and compare them with experimental data from literature. The results of the present work show that it is possible to predict growth with errors between 11 % and 28 % and production of CO<sub>2</sub> with errors between 0.3 % and 4.5 %; and ethanol production with errors between 11 % and 13 %, using FBA and iMM904 model.

Flux Balance Analysis - FBA - is one of the most used in microorganism bio-products predictions. It requires and objective function that represents biological objective of the studied microorganism. This paper presents a new kind of objective functions based on individual physical compartment objectives in the studied microorganism. These kind of functions was tested with a stoichiometric model extracted from iMM904 reconstruction of *S. cerevisiae* and its performance is compared with the most used objective function in literature, growth maximization, in anaerobic and aerobic batch conditions. The presented objective function outperform growth predictions in 10 % and ethanol predictions in 75 % compared with obtained by maximization of growth objective function, in anaerobic conditions. In aerobic batch conditions the presented objective function outperform in 98 % growth predictions and 70 % ethanol predictions compared with growth maximization.

---

<sup>1</sup>Research Work

<sup>2</sup>Faculty of Physical-Mechanical Engineering. Advisor, Henry Arguello. Co-advisor Rodrigo Gonzalo Torres Saez.

# Introducción

El microorganismo *Saccharomyces cerevisiae* (conocido como levadura de cerveza) se utiliza para la producción de alimentos, químicos y combustibles. Es pieza clave en el área de la biología eucariota debido a que conserva ciertas propiedades de organismos más complejos como el humano [1]. El estudio de reconstrucciones a escala genómica de *S. cerevisiae* ha servido como guía para la ingeniería metabólica en la mejora de producción de alimentos, químicos y combustibles, y como herramienta de interpretación biológica [1].

Una forma de almacenar la información estructural de *S. cerevisiae* es a través de los modelos estequiométricos [2], extraídos de reconstrucciones a escala genómica. Estas reconstrucciones han venido mejorando el nivel de detalle en la descripción de este microorganismo. Este nivel de detalle está representado por la cantidad de genes, reacciones y componentes incluidos en la reconstrucción. La lista de reconstrucciones disponibles en la literatura se muestra en la Tabla 1.

Las reconstrucciones base de la literatura son la iFF708 [3], iND750 [4] e iMM904 [5]. La más reciente de estas reconstrucciones base, la iMM904 se basa en el modelo iND750. Incluye mayor detalle en la descripción de *S. cerevisiae* comparado con las reconstrucciones iFF708 e iND750, producto de nueva información genética, bioquímica y fisiológica. Además, cuenta con una reacción de crecimiento modificada que mejora las predicciones de letalidad con respecto a sus antecesoras [5]. La reconstrucción iMM904 es hasta ahora la más detallada y confiable en la reproducción de condiciones experimentales de *S. cerevisiae* [5]. A partir de la iFF708, iND750 e iMM904 se han realizado

---

<b>Modelo</b>	<b>Trabajo</b>
<b>iFF708</b>	Förster et al., 2003
<b>iND750</b>	Duarte et al., 2004
iLL672	Kuepfer et al., 2005
Yeast 4.0	Herrgard et al., 2008
iIN800	Nookaew et al., 2008
<b>iMM904</b>	Mo et al., 2009
iAZ900	Zomorodi and Maranas, 2010

**Cuadro 1:** Reconstrucciones a escala genómica de *S. cerevisiae* presentes en la literatura. De arriba a abajo se muestran las reconstrucciones de la más antigua a la más reciente. Los modelos iFF708, iND750 e iMM904 son base de nuevos modelos a escala genómica de *S. cerevisiae*. Fuente: autores

nuevas reconstrucciones de *S. cerevisiae* tales como la iLL672 [6], la iIN800 [7], la iZ900 [8] y la Yeast 4.0 [9].

La técnica más utilizada para el estudio de los modelos estequiométricos extraídos de estas reconstrucciones es el Análisis de Balance de Flujos - FBA [10], [11], [12], [13], [14]. En los últimos años el FBA ha permitido estudiar la producción de bioproductos y la distribución intracelular de flujos, esta última a su vez ha facilitado la comprensión de la estructura biológica a escala genómica de microorganismos. El estudio de la distribución intracelular tienen aplicaciones que van desde la ingeniería metabólica hasta el descubrimiento de medicamentos [14]. El FBA ha cobrado interés debido a que no requiere información de la dinámica del microorganismo estudiado [10]. En vez de información dinámica, requiere de una función que represente el objetivo biológico del sistema estudiado.

El trabajo realizado en [15] fue uno de los primeros sobre predicción de consumo y producción de componentes, como el etanol, del organismo *S. cerevisiae* utilizando FBA y un modelo estequiométrico extraído de la reconstrucción a escala genómica iFF708. Las capacidades de los modelos extraídos de iFF708 e iND750 han sido estudiadas en los trabajos [15] y [16] y se han

---

realizado estudios posteriores como en [17] y [18] utilizando estos modelos. En contraste el modelo iMM904 carece de estudios como el de [15] y [16]. A la fecha no existen trabajos de FBA que utilicen el modelo iMM904.

Es posible que el Análisis de Balance de Flujos produzca más de una solución [19]. La enumeración de estas soluciones es computacionalmente costosa en modelos a escala genómica. Una alternativa computacionalmente viable para el estudio de estas soluciones es el Análisis de Variabilidad de Flujos - FVA. El primer trabajo que realiza un Análisis de Variabilidad de Flujos es el de [20].

Uno de los desafíos actuales presentes en el FBA es la determinación de una función objetivo que permita reproducir con mayor exactitud resultados experimentales de microorganismos estudiados [12]. El microorganismo *Escherichia coli* cuenta con numerosos estudios sobre la mejor función objetivo que describe su comportamiento bajo diferentes condiciones experimentales, desde anaeróbica hasta aeróbica, y con diferentes niveles de detalle en sus modelos estequiométricos, desde un conjunto reducido de reacciones hasta modelos extraídos de reconstrucciones a escala genómica [21]. En contraste el microorganismo *S. cerevisiae* tiene solo un estudio que busca nuevas funciones objetivo y ninguno usando modelos a escala genómica [21].

La búsqueda de la mejor función objetivo del FBA se realiza de dos formas: a través de programación binivel [22] y mediante evaluación sistemática de funciones [21]. En la primera se eligen un conjunto de posibles reacciones cuya maximización o minimización puedan explicar una condición experimental y se determinan coeficientes que acompañan a dichas reacciones a través de las soluciones obtenidas de un problema de programación binivel. El grupo de reacciones y los coeficientes que las acompañan se conocen como función objetivo. Los problemas de optimización binivel se vuelven intratables computacionalmente en modelos a escala genómica [23] debido a que su complejidad es NP-dura [24].

---

En la evaluación sistemática de funciones se elige un conjunto de funciones objetivo ya conocidas y probadas y mediante un criterio de selección se determina cual de estas funciones obtiene la predicción más exacta para el modelo y condiciones experimentales estudiadas. Esta solución es computacionalmente menos costosa que la solución de problemas binivel debido a que el número de funciones objetivo probadas en la literatura es reducido [21] y a que la prueba de una función objetivo consiste en resolver un FBA que a su vez consiste en resolver un problema de programación lineal.

Hasta la fecha, la función objetivo que predice con la mayor exactitud la producción de biomasa (o crecimiento del microorganismo) y etanol de *S. cerevisiae* en condiciones experimentales aeróbicas tipo batch es la maximización de producción de biomasa [23]. Sin embargo el estudio no se realizó sobre un modelo estequiométrico extraído de una reconstrucción a escala genómica.

La literatura actual carece de un trabajo que estudie la producción y consumo de los componentes etanol, glicerol, CO<sub>2</sub> y acetato de *S. cerevisiae* a través de FBA y FVA utilizando el modelo extraído de la reconstrucción iMM904. Aunque existe un estudio sobre este modelo [5], son necesarios trabajos que exploren otros componentes además de la glucosa, tales como el etanol el cual es importante en la producción de biocombustibles.

Dentro de los trabajos de evaluación sistemática de funciones no hay registro de alguno que estudie la combinación de funciones objetivo conocidas con el fin de mejorar predicciones de producción de bioproductos como biomasa (o crecimiento) o etanol. Tampoco se ha estudiado que propiedades aporta la división espacial encontrada en microorganismos como *S. cerevisiae* a la predicción de bioproductos.

El primer objetivo de este trabajo es comparar los resultados obtenidos al aplicar la técnica de FBA utilizando el modelo estequiométrico iMM904, contra resultados experimentales de este organismo. Además se busca comparar los resultados de aplicar FVA contra mediciones experimentales. En este trabajo

---

se describen y utilizan las técnicas FBA y FVA sobre el modelo estequiométrico de la iMM904. La capacidad de predicción es representada por los errores entre las predicciones de crecimiento y producción de los componentes CO<sub>2</sub> y etanol del organismo *S. cerevisiae*, y los correspondientes valores experimentales.

El presente trabajo también propone una nueva forma de crear funciones objetivo para el FBA basada en la división espacial del microorganismo *S. cerevisiae*. Esta división es conocida como compartimentalización. A cada compartimento se le asigna una función objetivo conocida. La función objetivo a utilizar en el FBA es la combinación de las funciones individuales de cada compartimento. Este nuevo tipo de funciones es probada en el FBA utilizando un modelo estequiométrico extraído de la reconstrucción a escala genómica iMM904. Las condiciones experimentales a estudiar son anaeróbicas en cultivos continuos<sup>1</sup> y aeróbicas tipo batch<sup>2</sup> en fase exponencial<sup>3</sup>. Los resultados obtenidos con la mejor de estas funciones son comparados con los obtenidos usando la función objetivo maximización de biomasa, que es la sugerida por [23] y [21] y usada en el FBA tradicional.

El documento se encuentra organizado de la siguiente forma: primero se explica el modelo estequiométrico y se describe las técnicas FBA y FVA, y el por qué de la existencia de múltiples soluciones de un FBA. Más adelante se explica una nueva propuesta de funciones objetivo para el FBA, por último se describen los resultados obtenidos al aplicar el FBA tradicional, el FVA y el FBA utilizando la función objetivo propuesta en este trabajo.

---

<sup>1</sup>En cultivos continuos el microorganismo estudiado está en estado estacionario. La tasa de crecimiento es constante así como el consumo de nutrientes y la producción de bioproductos.

<sup>2</sup>El crecimiento del cultivo, la producción de bioproductos y la utilización de sustratos no se mantiene constante y termina después de un periodo de tiempo.

<sup>3</sup>En fase exponencial las tasas de crecimiento, consumo de sustratos y formación de bioproductos es constante y está al máximo de su capacidad

# Capítulo 1

## Análisis de Balance de Flujos - FBA

### 1.1. Modelo Estequiométrico

En el modelo estequiométrico, la variación de la cantidad de un componente a través del tiempo  $\frac{dn_i}{dt}$ , se define como una combinación lineal

$$\frac{dn_i}{dt} = \sum_{j=1}^N S_{i,j} v_j, \quad (1.1)$$

donde  $v_j$  son las velocidades de reacción que lo afectan,  $S_{i,j}$  es el coeficiente estequiométrico que indica la cantidad requerida o producida del componente  $n_i$  en la reacción con velocidad  $v_j$ , y  $N$  es el número de reacciones del sistema. De la misma manera, el conjunto de variaciones de todos los componentes  $n_i$  donde  $i = 1, \dots, M$  se representan por medio del sistema de ecuaciones lineales

$$\frac{d\mathbf{n}}{dt} = \mathbf{S}\mathbf{v}, \quad (1.2)$$

donde  $\mathbf{n} = [n_1, \dots, n_M]^T$  es el vector de cantidades de los componentes del sistema estudiado,  $\mathbf{S}$  es una matriz de dimensiones  $M \times N$ , la cual se conoce como la matriz estequiométrica, y  $\mathbf{v} = [v_1, \dots, v_N]^T$  es el vector de velocidades de reacción (vector de distribución de flujos o vector de flujos). En todo

---

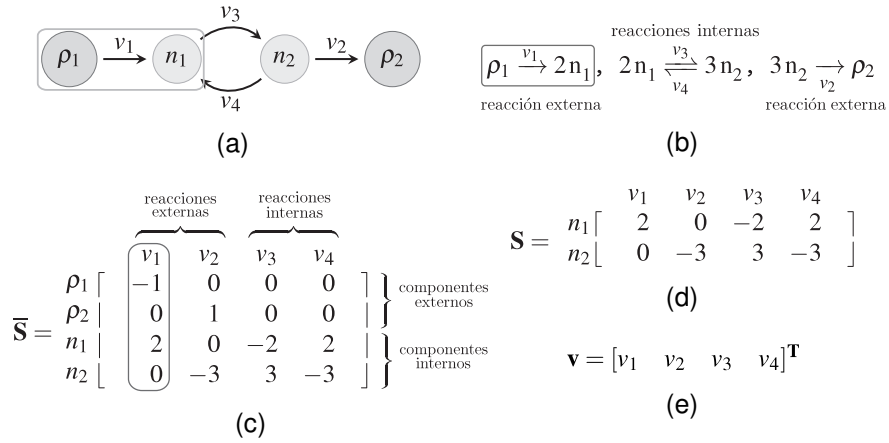
modelo estequiométrico el número de componentes es menor que el número de reacciones, es decir  $M < N$ .

La Figura 1.1 muestra un ejemplo de un modelo estequiométrico. El sistema está compuesto por cuatro componentes cuyas cantidades son:  $\rho_1$ ,  $\rho_2$ ,  $n_1$ ,  $n_2$  y cuatro reacciones cuyas velocidades son:  $v_1$ ,  $v_2$ ,  $v_3$  y  $v_4$ .  $\rho_1$  y  $\rho_2$  son respectivamente la entrada y la salida del sistema y se conocen como componentes externos.  $n_1$  y  $n_2$  representan componentes internos. Este sistema puede representarse mediante un grafo cuyos nodos corresponden a los componentes  $n_i$  y  $\rho_w$   $w = 1, \dots, W$ , y cuyas aristas corresponden a las reacciones  $v_j$  (Figura 1.1(a)).

El conjunto de reacciones que describe el sistema se muestra en la Figura 1.1(b). En este sistema una reacción es un proceso que afecta un conjunto de componentes, así como una arista es una conexión entre un conjunto de nodos en un grafo. Las reacciones se clasifican en internas y externas. Las reacciones internas incluyen solo componentes internos y las externas afectan tanto componentes internos como externos. La cantidad de cada uno de los componentes que involucra cada reacción se muestra en la Figura 1.1(b); por ejemplo, la reacción que ocurre a la velocidad  $v_1$  toma una unidad del componente  $\rho_1$  y produce dos unidades de  $n_1$ .

Para construir la matriz  $\mathbf{S}$  se usa la matriz de incidencia  $\bar{\mathbf{S}}$  de dimensiones  $(W + M) \times N$ , mostrada en la Figura 1.1(c). Las primeras  $W$  filas de  $\bar{\mathbf{S}}$  corresponden a los componentes  $\rho_w$ , y las restantes corresponden a los componentes  $n_i$ . Las columnas de  $\bar{\mathbf{S}}$  corresponden a las reacciones con velocidades  $v_j$ .

En la Figura 1.1(c) la columna 1 de la matriz  $\bar{\mathbf{S}}$  corresponde a la reacción con velocidad  $v_1$ . Esta columna fue construida tomando los coeficientes que aparecen en la Figura 1.1(b). En esta reacción no participan los componentes  $\rho_2$  y  $n_2$ , por tanto  $\bar{S}_{2,1} = \bar{S}_{4,1} = 0$ . El componente  $\rho_1$  se encuentra en la parte inicial de la flecha que representa la reacción  $v_1$  por lo tanto el valor de  $\bar{S}_{1,1}$  es igual al coeficiente que acompaña a  $\rho_1$  con signo negativo, es decir  $\bar{S}_{1,1} = -1$ .



**Figura 1.1:** Ejemplo de modelo estequiométrico. (a) grafo del sistema, (b) reacciones químicas, (c) matriz de incidencia, (d) matriz estequiométrica y (e) vector de velocidades de reacción.

El componente  $n_1$  se encuentra en la parte final de la flecha que representa a  $v_1$ , por lo tanto  $\bar{S}_{3,1}$  es igual al coeficiente que acompaña a  $n_1$  con signo positivo, es decir  $\bar{S}_{3,1} = 2$ .

En general, si  $\bar{S}_{W+i,j} < 0$  entonces  $v_j$  consume  $|\bar{S}_{W+i,j}|$  unidades del componente  $n_i$ , si  $\bar{S}_{W+i,j} > 0$  entonces la reacción  $v_j$  produce el componente  $n_i$   $\bar{S}_{W+i,j}$  unidades, y si  $\bar{S}_{W+i,j} = 0$  el componente  $n_i$  no participa en la reacción  $v_j$ . De forma análoga,  $\bar{S}_{w,j}$  representa la participación del componente  $\rho_w$  en la reacción  $v_j$ .

La matriz  $\mathbf{S}$  se obtiene eliminando las primeras  $W$  filas de  $\bar{\mathbf{S}}$  que corresponden a componentes externos. La Figura 1.1(d) muestra la matriz  $\mathbf{S}$  resultado de eliminar las filas  $w = 1, 2$  de  $\bar{\mathbf{S}}$ , que corresponden a los componentes  $\rho_1$  y  $\rho_2$ . En  $\mathbf{S}$  se cumple que  $M < N$  ( $M = 2$  y  $N = 4$ ). Finalmente, la Figura 1.1(e) muestra el vector de velocidades de reacción  $\mathbf{v}$ .

---

## 1.2. Análisis de Balance de Flujos - FBA

Para estimar el comportamiento de un sistema usando el modelo estequiométrico, se asume estado estacionario del sistema el cual se representa por medio de la ecuación

$$\frac{dn}{dt} = \mathbf{S}\mathbf{v} = \mathbf{0}_M, \quad (1.3)$$

donde  $\mathbf{0}_M$  es un vector de longitud M cuyos elementos son todos cero. La ecuación matricial (1.3) define un sistema de ecuaciones lineales subdeterminado con infinitas soluciones. Para delimitar el espacio solución de la ecuación (1.3) se definen los límites de cada una de las velocidades de reacción  $v_j$  del sistema mediante

$$\alpha_j \leq v_j \leq \beta_j, \quad j = 1, \dots, N, \quad (1.4)$$

donde  $\alpha_j$  y  $\beta_j$  son los límites inferior y superior de  $v_j$ . Por último se establece una función objetivo

$$f = \mathbf{c}^T \mathbf{v}, \quad (1.5)$$

donde  $\mathbf{c} = [c_1, \dots, c_N]$  es el vector que define el objetivo del sistema. Si la reacción  $v_j$  está en la función objetivo  $f$  entonces  $c_j \neq 0$ . En caso contrario  $c_j = 0$ .

Dados la matriz  $\mathbf{S}$ , los vectores  $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_N]$ ,  $\boldsymbol{\beta} = [\beta_1, \dots, \beta_N]$  y  $\mathbf{c}$ , se busca un vector  $\mathbf{v}$  que sea solución del problema de programación lineal

$$\begin{aligned} \underset{\mathbf{v}}{\text{máx}} \quad & f = \mathbf{c}^T \mathbf{v} \\ \text{s.a.} \quad & \\ & \mathbf{S}\mathbf{v} = \mathbf{0}_M, \\ & v_j \geq \alpha_j \\ & v_j \leq \beta_j \end{aligned} \quad (1.6)$$

La búsqueda de soluciones al problema (1.6) se conoce como FBA.

---

La Figura 1.2 muestra un sistema biológico y las definiciones correspondientes a la formulación FBA. Los elementos externos son  $\rho_1, \rho_2, \rho_3, \rho_4$  y los internos son  $n_1, n_2, n_3, n_4$  y  $n_5$ . El grafo que describe el sistema se muestra en la Figura 1.2(a). La matriz estequiométrica  $\mathbf{S}$  se define en la Figura 1.2(b).  $\alpha, \beta$  y  $\mathbf{c}$  se muestran en la Figura 1.2(c). El valor negativo de  $\alpha_1$  indica que la dirección de la reacción con velocidad  $v_1$  puede ser en cualquiera de dos sentidos: consumo del elemento  $n_1$  con un valor máximo  $\beta_1 = 5$ , o producción de éste con valor mínimo  $\alpha_1 = -2$ . Un valor negativo de  $v_j$  indica que lo que se produce en una reacción se consume y lo que se consume se produce. En general si  $\alpha_j < 0$  entonces la reacción con velocidad  $v_j$  se denomina reversible. Si por el contrario  $\alpha_j \geq 0$  entonces  $v_j$  es irreversible, debido a que conserva la dirección con la que aparece el grafo [12].

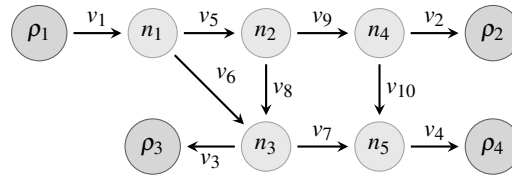
El vector de coeficientes  $\mathbf{c}$  define la función objetivo del sistema de la Figura 1.2. En este ejemplo el vector contiene valores diferentes de 0 en las posiciones correspondientes a las reacciones externas  $v_2$  y  $v_4$ . Esto indica que el objetivo del sistema es la maximización de las velocidades de las reacciones que producen los elementos  $\rho_2$  y  $\rho_4$ . En el caso de la Figura 2, la función objetivo se puede expresar como  $f = \mathbf{c}^T \mathbf{v} = v_2 + v_4$ . La solución  $\mathbf{v}$  mostrada en la Figura 1.2(c) se obtuvo a través de programación lineal utilizando el paquete matemático MatLab<sup>®</sup>. La Figura 1.2(d) muestra de forma gráfica la interpretación de este resultado. Por último la Figura 1.2(e) muestra el código utilizado en MatLab para obtener esta solución.

### 1.3. Soluciones del FBA

El conjunto de ecuaciones lineales  $\mathbf{Sv} = \mathbf{0}_M$  y de inecuaciones lineales  $\alpha_j \leq v_j \leq \beta_j$  definen el poliedro convexo, cerrado y acotado

$$P = \{ \mathbf{v} \mid \mathbf{Sv} = \mathbf{0}_M, \alpha_j \leq v_j \leq \beta_j \}. \quad (1.7)$$

Como lo demuestra [25], todo poliedro convexo, cerrado y acotado tiene la



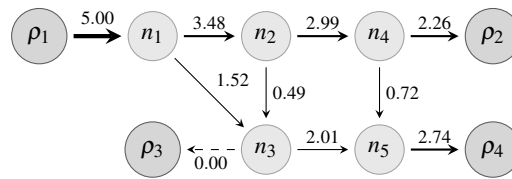
(a)

$$\mathbf{S} = \begin{array}{c} \begin{array}{cc} \text{reacciones} & \text{reacciones} \\ \text{externas} & \text{internas} \end{array} \\ \begin{array}{cccccccccccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 & v_9 & v_{10} \end{array} \\ \begin{array}{l} n_1 \\ n_2 \\ n_3 \\ n_4 \\ n_5 \end{array} \left[ \begin{array}{cccccccccccc} 1 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 & -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right] \end{array}$$

(b)

$$\begin{array}{l} \beta^T = [ \quad v_1 \quad v_2 \quad v_3 \quad v_4 \quad v_5 \quad v_6 \quad v_7 \quad v_8 \quad v_9 \quad v_{10} ] \\ \alpha^T = [ \quad 5 \quad 4 \quad 3 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 8 ] \\ \mathbf{c}^T = [ \quad -2 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 ] \\ \mathbf{v}^T = [ \quad 0 \quad 1 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 ] \\ \mathbf{v}^T = [ \quad 5.00 \quad 2.26 \quad 0.00 \quad 2.74 \quad 3.48 \quad 1.52 \quad 2.01 \quad 0.49 \quad 2.99 \quad 0.72 ] \end{array}$$

(c)



(d)

```
v=linprog(-c, [], [], S, zeros(M,1), alfa, beta)
```

```
c=c, S=S, 0_M = zeros(M,1), alpha=alfa, beta=beta
```

(e)

**Figura 1.2:** Principales elementos de un problema FBA. (a) grafo del sistema, (b) matriz estequiométrica  $\mathbf{S}$ , (c) vector de velocidades máximas  $\beta$  y mínimas  $\alpha$ , vector de función objetivo  $\mathbf{c}$  y vector solución  $\mathbf{v}$  encontrado al resolver el FBA. En (d) se representa gráficamente este resultado. El valor máximo de la función objetivo  $f$  es 5.

representación

---


$$P = \left\{ \mathbf{v} \mid \mathbf{v} = \sum_{k=1}^L \lambda_k \mathbf{u}_k, \lambda_k \geq 0, \sum_{k=1}^L \lambda_k = 1 \right\} \quad (1.8)$$

donde  $\mathbf{u}_k$   $k = 1, \dots, L$  se conoce como los puntos extremos o vértices del poliedro  $P$ . Un punto  $\mathbf{u}_k$  es extremo si no existen dos puntos  $\mathbf{a}$  y  $\mathbf{b} \in P$  tales que  $\mathbf{u}_k = \theta \mathbf{a} + (1 - \theta) \mathbf{b}$ . Esta representación de  $P$  se conoce como la representación  $V$ .

El teorema estudiado en [25] indica que el vector que resuelve la ecuación (1.6) es el punto extremo  $\mathbf{u}_Q$ ,  $Q \in \{1, \dots, L\}$  tal que  $\mathbf{c}^T \mathbf{u}_Q > \mathbf{c}^T \mathbf{u}_k$  con  $k \neq Q$ . En caso de existir múltiples soluciones, los vectores que resuelven la ecuación (1.6) son los puntos extremos  $\mathbf{u}_{Q_r}$   $Q_r \in \{1, \dots, L\}$ ,  $r = 1, \dots, R$  tales que  $\mathbf{c}^T \mathbf{u}_{Q_r} > \mathbf{c}^T \mathbf{u}_k$  con  $k \neq Q_r$  y  $\mathbf{c}^T \mathbf{u}_{Q_1} = \dots = \mathbf{c}^T \mathbf{u}_{Q_R}$ . Más aún, los vectores de la forma

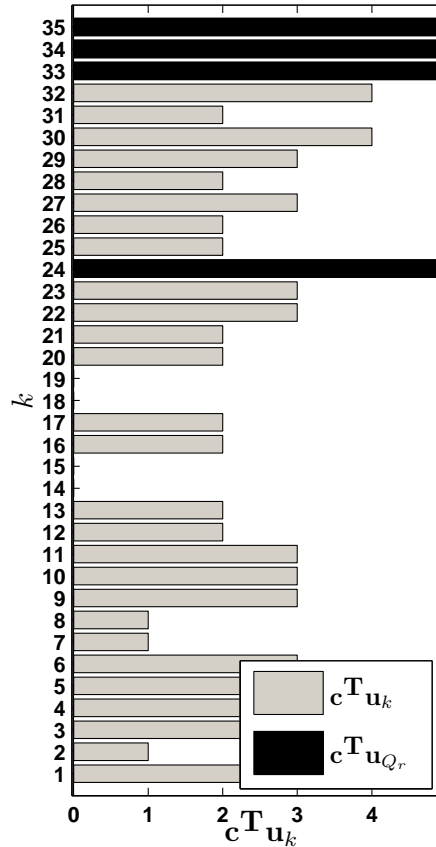
$$\mathbf{v} = \sum_{r=1}^R \lambda_r \mathbf{u}_{Q_r}, \lambda_r \geq 0, \sum_{r=1}^R \lambda_r = 1, \quad (1.9)$$

también son solución de la ecuación (1.6).

Los teoremas anteriormente descritos son equivalentes a establecer que, el punto extremo con el mayor valor de función objetivo es solución única de la ecuación (1.6). Si existe más de un punto extremo con el máximo valor de función objetivo posible, la ecuación (1.6) tiene infinitas soluciones, todas descritas por combinaciones lineales convexas de estos puntos extremos.

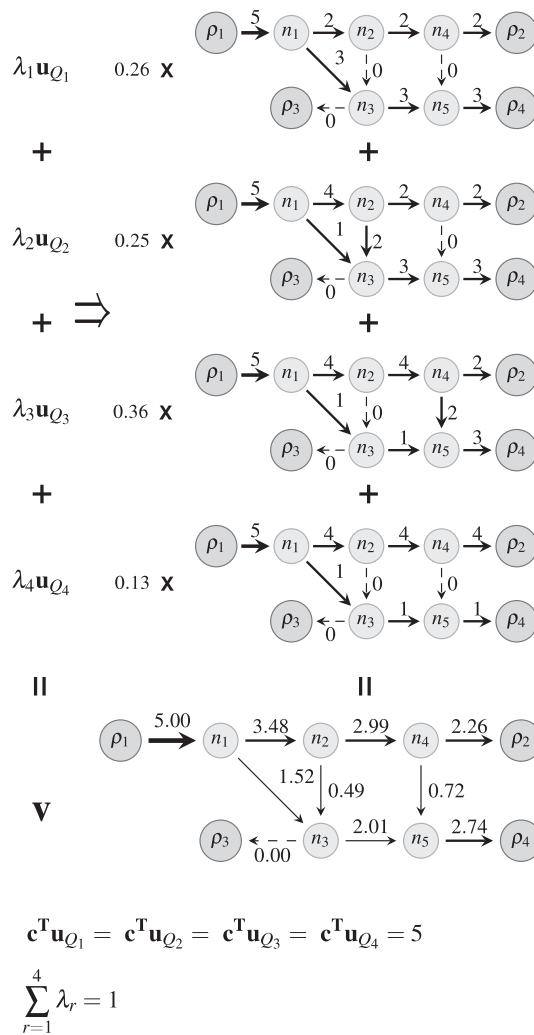
La existencia de más de una solución se ilustra a través del estudio de los puntos extremos del problema de la Figura 1.2. Se obtuvieron  $L = 35$  puntos extremos mediante el Método de Descripción Doble (cdd) [26], con implementación disponible para uso en MatLab. En la Figura 1.3 se muestran los valores de  $\mathbf{c}^T \mathbf{u}_k$ ,  $k = 1, \dots, L$ . El máximo valor de  $\mathbf{c}^T \mathbf{u}_k$  es 5, el cual corresponde a los

puntos  $\mathbf{u}_{24}$ ,  $\mathbf{u}_{33}$ ,  $\mathbf{u}_{34}$  y  $\mathbf{u}_{35}$ .



**Figura 1.3:** valores de función objetivo para cada punto extremo del problema presentado en la Figura 1.2. En color negro están los puntos extremos  $\mathbf{u}_k$  que tienen el valor máximo de función objetivo.

En la Figura 1.4 se muestran los puntos  $\mathbf{u}_{Q_r}$ ,  $r = 1, \dots, 4$  con  $Q_1 = 24$ ,  $Q_2 = 33$ ,  $Q_3 = 34$  y  $Q_4 = 35$ , que son los puntos extremos solución de la ecuación (1.6). El vector solución  $\mathbf{v}$  mostrado en la Figura 1.2 se puede expresar como la combinación lineal convexa de los puntos extremos  $\mathbf{u}_{Q_r}$  indicados en la Figura 1.3. Esto se indica en la Figura 1.4. Para encontrar los valores  $\lambda_r$  mostrados en la Figura 1.4, se hizo  $\lambda = \mathbf{U}^+ \mathbf{v}$ , donde  $\lambda = [\lambda_1, \dots, \lambda_4]^T$ ,  $\mathbf{U} = [\mathbf{u}_{Q_1}, \dots, \mathbf{u}_{Q_4}]$ ,  $\mathbf{U}^+$  es la pseudoinversa de  $\mathbf{U}$  y  $\mathbf{v}$  es el vector solución encontrado en la Figura 1.2.



**Figura 1.4:** ilustración del conjunto solución de la ecuación (1.6) del problema de la Figura 1.2. Izquierda: combinación lineal convexa de los puntos extremos que son solución de (1.6). Derecha: puntos extremos y valores de la combinación lineal convexa. Abajo: solución problema de la Figura 1.2. Las funciones objetivo evaluadas en los puntos extremos mostrados son iguales al máximo valor posible.

---

## 1.4. Análisis de Variabilidad de Flujos - FVA

Debido a la existencia de múltiples soluciones de (1.6), se requiere determinar el intervalo de posibles valores de  $v_j$ . El Análisis Variabilidad de Flujos [20] consiste en optimizar el valor de  $v_j$  mediante dos problemas de optimización separados. La primera optimización consiste en minimizar  $v_j$  a través de

$$\begin{aligned} & \underset{v_j}{\text{mín}} \\ & \text{s.a.} \end{aligned} \quad \begin{aligned} \mathbf{Sv} &= \mathbf{0}_M \\ v_j &\geq \alpha_j \\ v_j &\leq \beta_j \\ \mathbf{c}^T \mathbf{v} &= f \end{aligned} \quad (1.10)$$

y la segunda optimización consiste en maximizar  $v_j$  resolviendo

$$\begin{aligned} & \underset{v_j}{\text{máx}} \\ & \text{s.a.} \end{aligned} \quad \begin{aligned} \mathbf{Sv} &= \mathbf{0}_M \\ v_j &\geq \alpha_j \\ v_j &\leq \beta_j \\ \mathbf{c}^T \mathbf{v} &= f \end{aligned} \quad (1.11)$$

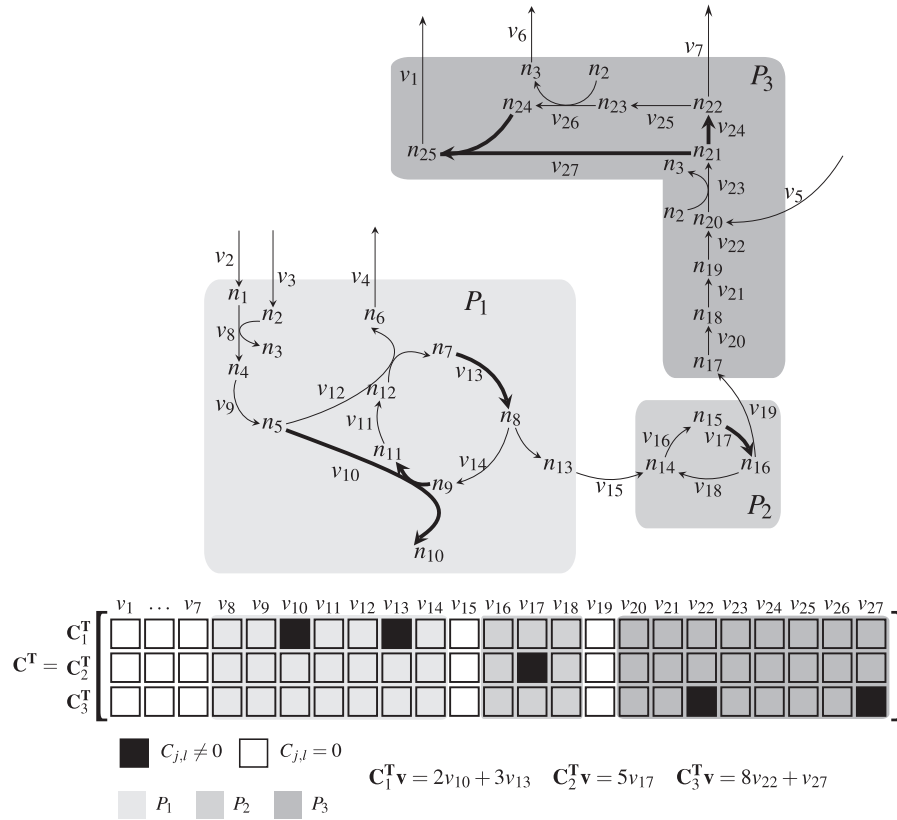
## Capítulo 2

# Funciones Objetivo Basadas en Compartimentos de Microorganismos

### 2.1. Funciones Objetivo Basadas en Compartimentos

Sea el compartimento  $P_l$   $l = 1, \dots, L$  un conjunto de índices  $j$  de  $\mathbf{v}$  se define la combinación lineal  $\mathbf{C}_l^T \mathbf{v}$ , donde  $\mathbf{C}_l$  es un vector columna de dimensión  $N$ . Esta combinación se conoce como la función objetivo del compartimento  $P_l$ . Si  $j \in P_l$  y la velocidad  $v_j$  no está incluida dentro de la función objetivo del compartimento  $P_l$  entonces  $C_{j,l} = 0$ . Por el contrario, si  $j \in P_l$  y  $v_j$  hace parte de la función objetivo del compartimento  $P_l$  entonces  $C_{j,l} \neq 0$ . Por último  $C_{j,l} = 0$  si  $j \notin P_l$ .

La Figura 2.1 muestra un ejemplo de un sistema con 3 compartimentos:  $P_1$ ,  $P_2$  y  $P_3$ . El conjunto  $P_1$  contiene las reacciones  $v_8, \dots, v_{14}$ ,  $P_2$  contiene las reacciones  $v_{16}, \dots, v_{18}$ , y  $P_3$  las reacciones  $v_{20}, \dots, v_{27}$ . Las reacciones  $v_1, \dots, v_7$  no pertenecen a compartimento alguno y se conocen como reacciones de transporte externas del sistema. Las reacciones  $v_{15}$  y  $v_{19}$  representan el intercambio de componentes entre compartimentos y tampoco pertenecen a compartimen-



**Figura 2.1:** Sistema compuesto por 3 compartimentos:  $P_1$ ,  $P_2$  y  $P_3$ . Arriba, diagrama que representa el sistema. Abajo, funciones objetivo  $\mathbf{C}_1^T \mathbf{v}$ ,  $\mathbf{C}_2^T \mathbf{v}$  y  $\mathbf{C}_3^T \mathbf{v}$  de dichos compartimentos. La función objetivo del compartimento  $P_1$  está compuesta por las velocidades de reacción  $v_{10}$  y  $v_{13}$ , resaltadas en el diagrama y en la función objetivo (cuadros negros del vector  $\mathbf{C}_1^T$ ).

to alguno. La Figura 2.1 también muestra los vectores columna  $\mathbf{C}_1$ ,  $\mathbf{C}_2$  y  $\mathbf{C}_3$  los cuales representan las funciones objetivo de los compartimentos  $P_1$ ,  $P_2$  y  $P_3$  respectivamente. Cabe aclarar que aunque las reacciones  $v_8, \dots, v_{14}$  pertenecen al compartimento  $P_1$ , sólo las reacciones  $v_{10}$  y  $v_{13}$  se incluyen en la función objetivo  $\mathbf{C}_1^T \mathbf{v}$ .

Dados  $\mathbf{S}$ ,  $\alpha$ ,  $\beta$ , la matriz  $\mathbf{C} = [\mathbf{C}_{l_1}, \dots, \mathbf{C}_{l_K}]$  de dimensiones  $N \times K$  y el vector  $\mathbf{w} = [w_1, \dots, w_K]^T$ , el presente trabajo propone buscar el vector  $\mathbf{v}$  que es solución del problema de programación lineal

---

$$\begin{aligned} & \underset{\mathbf{v}}{\text{máx}} && \mathbf{w}^T \mathbf{C}^T \mathbf{v} \\ & \text{s.a.} && \\ & && \mathbf{Sv} = \mathbf{0}_M \\ & && v_j \geq \alpha_j \\ & && v_j \leq \beta_j, \end{aligned} \tag{2.1}$$

donde  $w_k \geq 0$  con  $k = 1, \dots, K$ . El valor  $w_k$  representa el aporte del objetivo del compartimento  $P_{l_k}$  al objetivo global del sistema.  $\mathbf{w}^T \mathbf{C}^T$  representa una nueva función objetivo obtenida de la combinación de las funciones objetivo de  $K$  compartimentos presentes en el sistema estudiado.

# Capítulo 3

## Resultados

### 3.1. FBA Tradicional usando la Reconstrucción iMM904

#### Simulaciones

Las simulaciones realizadas buscan determinar la capacidad de predicción del FBA usando el modelo estequiométrico obtenido de la reconstrucción iMM904. Para este fin, se compara las velocidades de reacciones de consumo/producción de componentes obtenidas a través del FBA, con las respectivas mediciones experimentales. El conjunto de mediciones experimentales usado en este trabajo fue tomado de [27] y se presenta en la Tabla 3.1. Consta de 4 experimentos bajo condiciones anaeróbicas, en los cuales varía la velocidad de crecimiento (gw) del sistema. A cada velocidad de crecimiento, le corresponde la velocidad de producción de dióxido de carbono (CO<sub>2</sub>), etanol (eth), glicerol (gly), succinato (succ), acetato (ace) y consumo de glucosa (glc).

La reconstrucción iMM904 cuenta con 1228 componentes internos y 1577 reacciones. Está disponible en formato SimPheny<sup>®</sup> [5].

Se define  $v_{\langle reaccion \rangle}$ , como la velocidad de reacción de nombre  $\langle reaccion \rangle$ , con  $j$  igual al índice de la columna de  $\mathbf{S}$  que corresponde a la reacción  $\langle$

cond	eth $\left(\frac{mmol}{gDW\ h}\right)$	gly $\left(\frac{mmol}{gDW\ h}\right)$	succ $\left(\frac{mmol}{gDW\ h}\right)$	glc $\left(\frac{mmol}{gDW\ h}\right)$	ace $\left(\frac{mmol}{gDW\ h}\right)$	pyr $\left(\frac{mmol}{gDW\ h}\right)$	CO <sub>2</sub> $\left(\frac{mmol}{gDW\ h}\right)$	gw $\left(\frac{1}{h}\right)$
1	8.77	1.01	0.03	5.88	0.04	0.01	9.60	0.101
2	16.84	2.06	0.05	11.31	0.10	0.05	18.53	0.190
3	24.82	3.18	0.05	16.75	0.30	0.10	26.83	0.281
4	33.07	4.83	0.07	22.18	0.67	0.19	34.73	0.369

**Cuadro 3.1:** Conjunto de datos 1. Abreviaturas: (cond) número condición experimental, producción de etanol (eth), producción de glicerol (gly), producción de succinato (succ), consumo de glucosa (glc), producción de acetato, producción de piruvato (pyr), producción de dióxido de carbono (CO<sub>2</sub>), crecimiento (gw).

$reaccion >$ . Los valores  $\alpha_{<reaccion>}$  y  $\beta_{<reaccion>}$  definen los límites de la reacción  $v_{<reaccion>}$ . El valor  $c_{<reaccion>}$  define el coeficiente que acompaña a  $v_{<reaccion>}$  en la función objetivo. La correspondencia entre nombres de reacción en la iMM904 y velocidades  $v_{<reaccion>}$  usados en las simulaciones del presente trabajo se muestran en la Tabla 3.2. Las reacciones listadas son externas, salvo  $v_{gw}$  y  $v_{ATPM}$ . Por convención, en la reconstrucción iMM904 todas las reacciones externas producen componentes externos.

El modelo estequiométrico  $\mathbf{S}^{1228 \times 1577}$  se construye utilizando el paquete COBRA disponible para MatLab® [28]. La reconstrucción iMM904 también cuenta con información sobre las capacidades del sistema representadas por los vectores  $\alpha$  y  $\beta$ . Excluyendo algunas reacciones, los límites de  $v_j$  son  $\alpha_j = -1000 \frac{mmol}{gDW\ h}$  para reacciones reversibles,  $\alpha_j = 0 \frac{mmol}{gDW\ h}$  para irreversibles y  $\beta_j = 1000 \frac{mmol}{gDW\ h}$ . La función objetivo incluida en esta reconstrucción es la maximización del crecimiento del sistema, representada por  $\mathbf{c}$ , con  $c_{gw} = 1$  y  $c_j = 0$  para toda reacción diferente a la de crecimiento. Este vector  $\mathbf{c}$  se interpreta como que el objetivo es la maximización de crecimiento del sistema. La reconstrucción iMM904 incluye la función objetivo  $\mathbf{c}^T \mathbf{v}$  descrita anteriormente debido a que es la más aceptada en la literatura [21].

Las condiciones de las mediciones experimentales de la Tabla 3.1 requieren modificación de algunas restricciones FBA que establece por defecto la

reacción	iMM904	FBA
etanol	EX_etoh(e)	$v_{EX\_etoh(e)}$
glicerol	EX_gly(e)	$v_{EX\_gly(e)}$
sucinato	EX_succ(e)	$v_{EX\_succ(e)}$
glucosa	EX_glc(e)	$v_{EX\_glc(e)}$
acetato	EX_ace(e)	$v_{EX\_ace(e)}$
piruvato	EX_pyr(e)	$v_{EX\_pyr(e)}$
dióxido de carbono	EX_co2(e)	$v_{EX\_co2(e)}$
crecimiento	biomass	$v_{gw}$
oxígeno	EX_o2(e)	$v_{EX\_o2(e)}$
mantenimiento ATP	ATPM	$v_{ATPM}$
ergosterol	EX_ergst(e)	$v_{EX\_ergst(e)}$
zymosterol	EX_zymst(e)	$v_{EX\_zymst(e)}$
hexadecenato (n-C16:1)	EX_hdcea(e)	$v_{EX\_hdcea(e)}$
octadecanato (n-C18:0)	EX_ocdca(e)	$v_{EX\_ocdca(e)}$
octadecenato (n-C18:1)	EX_ocdcea(e)	$v_{EX\_ocdcea(e)}$
octadecenato (n-C18:2)	EX_ocdcya(e)	$v_{EX\_ocdcya(e)}$

**Cuadro 3.2:** correspondencias entre nombres de reacción usados en la reconstrucción iMM904 y las velocidades de reacción  $v_{<reaccion>}$  para este trabajo.

reconstrucción iMM904. La condición anaerobia significa que el medio en el que se encuentra carece de oxígeno. En el FBA esto equivale a establecer  $\alpha_{EX\_o2(e)} = \beta_{EX\_o2(e)} = 0$ . Debido al estado anaerobio es necesario que el sistema tenga acceso a componentes que en ausencia de oxígeno no se pueden producir. Por defecto las reacciones correspondientes a estos componentes solo producen y no consumen, es decir dichas reacciones son irreversibles. Para permitir que el sistema tenga acceso a componentes fundamentales se convierten en reversibles.

El mantenimiento interno de energía del sistema, representado en el modelo por  $v_{ATPM}$  se conserva igual al provisto por la reconstrucción. La lista de restricciones por defecto de la iMM904 y las restricciones del FBA establecidas en este trabajo se presentan en la Tabla 3.7. Esta lista no incluye todas las restricciones, solo aquellas relevantes para el conjunto de datos experimentales de la Tabla 3.1.

	iMM904		FBA	
	$\alpha_j$ $\left(\frac{\text{mmol}}{\text{gDW h}}\right)$	$\beta_j$ $\left(\frac{\text{mmol}}{\text{gDW h}}\right)$	$\alpha_j$ $\left(\frac{\text{mmol}}{\text{gDW h}}\right)$	$\beta_j$ $\left(\frac{\text{mmol}}{\text{gDW h}}\right)$
$v_{EX\_o2(e)}$	<b>-2</b>	<b>1000</b>	<b>0</b>	<b>0</b>
$v_{EX\_ergst(e)}$	<b>0</b>	1000	<b>-1000</b>	1000
$v_{EX\_zymst(e)}$	<b>0</b>	1000	<b>-1000</b>	1000
$v_{EX\_hdcea(e)}$	<b>0</b>	1000	<b>-1000</b>	1000
$v_{EX\_ocdca(e)}$	<b>0</b>	1000	<b>-1000</b>	1000
$v_{EX\_ocdcea(e)}$	<b>0</b>	1000	<b>-1000</b>	1000
$v_{EX\_ocdcya(e)}$	<b>0</b>	1000	<b>-1000</b>	1000
$v_{ATPM}$	1	1	1	1

**Cuadro 3.3:** restricciones  $\alpha_j$  y  $\beta_j$  presentes en la reconstrucción iMM904 y en el FBA de este trabajo. En negrilla se resaltan cambios entre los valores de la reconstrucción y los utilizados en el FBA de este trabajo.

Cada una de las cuatro condiciones experimentales expuestas en la Tabla 3.1 establece una entrada de glucosa. Esta entrada es representada fijando  $\alpha_{EX\_glc(e)}$  y  $\beta_{EX\_glc(e)}$  a cada uno de los valores experimentales. Por defecto todas las reacciones externas de la iMM904 producen componentes externos. La reacción  $v_{EX\_glc(e)}$  es externa y produce el componente de glucosa (glc(e)). El valor experimental de glucosa que aparece en la Tabla 3.1 es de consumo y no producción por tanto, el valor de  $\alpha_{EX\_glc(e)}$  y  $\beta_{EX\_glc(e)}$  es el valor experimental con signo negativo.

Por cada condición experimental de la Tabla 3.1 se realizó un FBA y un FVA. Se determinó el error entre la predicción de las velocidad de reacción  $v_j$  y el correspondiente valor experimental  $v_j^E$  mediante  $|(v_j - v_j^E)/v_j^E| \times 100\%$ . También se midió la variabilidad de cada una de las reacciones en cada una de las condiciones experimentales. Esta variabilidad se obtuvo restando el valor mínimo del máximo de la reacción obtenidos resolviendo las ecuaciones (1.10) y (1.11) para aquellas con información experimental disponible.

---

También se realizó un FVA fijando el valor de la función objetivo  $f = v_{gw}$  al valor de crecimiento de cada condición experimental del sistema. Al fijar la función objetivo al valor experimental de crecimiento se busca determina los efectos de un estado subóptimo del sistema sobre la variabilidad de las demás reacciones con valor experimental.

Las simulaciones se realizaron en un equipo con procesador Intel<sup>®</sup> Core i3<sup>™</sup> y memoria RAM de 4GB DDR3. Los problemas de programación lineal de este trabajo fueron solucionados utilizando el paquete de software libre GLPK [29] con interfaz disponible para MatLab<sup>®</sup> a través del paquete COBRA.

## Resultados

Los resultados obtenidos se muestran en la Figura 3.1. Las reacciones con las mejores predicciones mediante FBA en las cuatro condiciones experimentales son, en orden de exactitud CO<sub>2</sub>, con errores entre 0,3 % y 4,5 %; etanol, entre 11 % y 13 %; y crecimiento, entre 11 % y 28 % (Figuras 3.1(g), 3.1(b) y 3.1(a)). La predicción del crecimiento es superior al registrado en cada condición experimental de la Tabla 3.1 (Figura 3.1(a)).

La predicción de glicerol en todas las condiciones experimentales es de  $0 \frac{mmol}{gDW h}$ . El valor experimental del glicerol en todas las condiciones comparadas está entre 0 y  $6 \frac{mmol}{gDW h}$ . Esto implica que el error de predicción de glicerol excede el 100 %. La predicción de succinato en todas las condiciones experimentales está entre 0 y  $2 \frac{mmol}{gDW h}$ . Los valores experimentales de succinato en todas las condiciones están entre 0.03 y  $0.07 \frac{mmol}{gDW h}$ . Por lo tanto, el error en la predicción de succinato supera el 100 %. Observaciones similares a las anteriores muestran que el error en las predicciones de acetato y piruvato exceden el 100 %.

La exactitud de las predicciones de CO<sub>2</sub> son mejores que las registradas en [15]. Las predicciones de etanol tienen una exactitud similar a las registradas en dicho trabajo. Es importante resaltar que [15] es el único estudio de S.

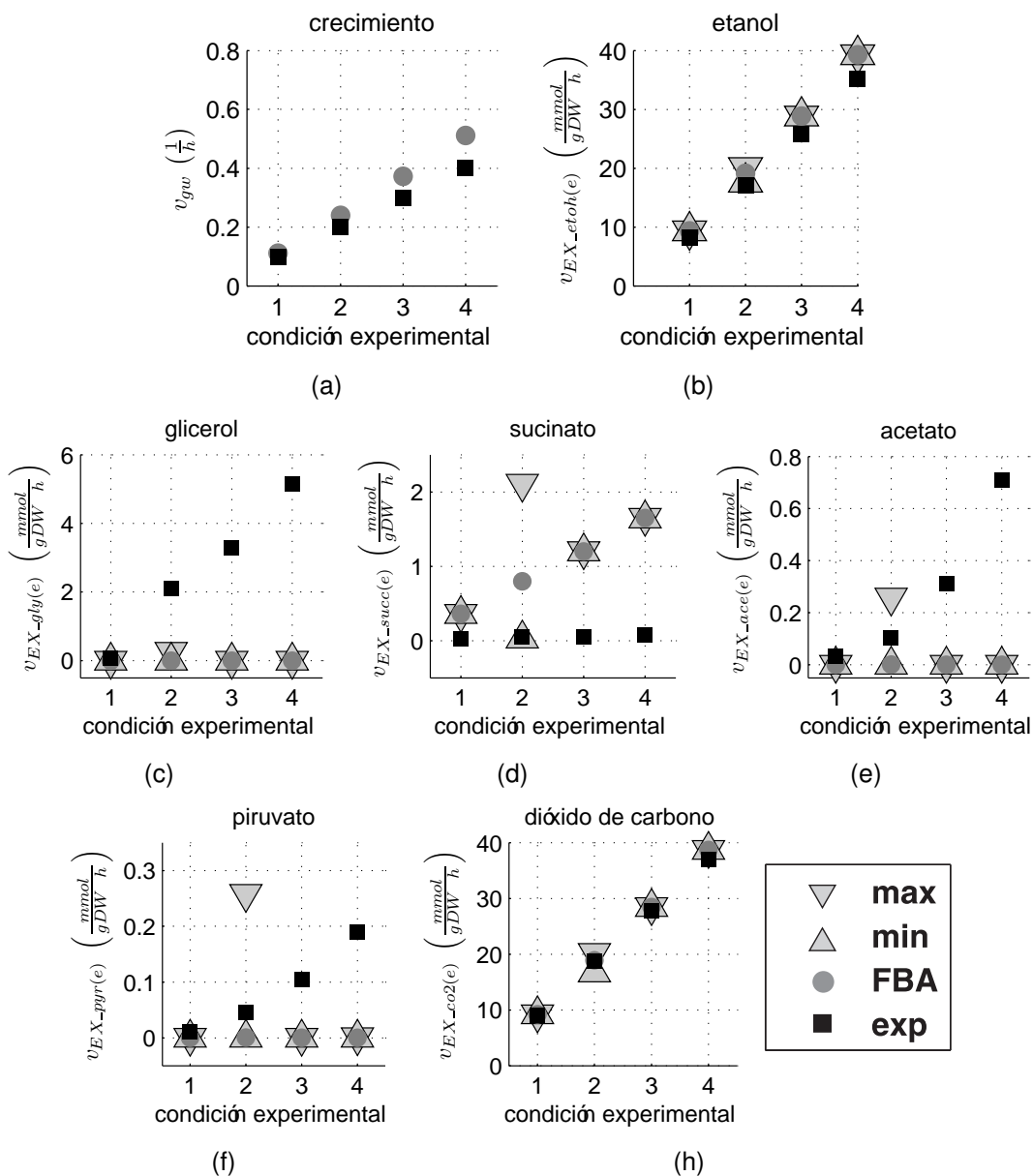
---

cerevisiae mediante FBA y modelos genómicos que incluye un FVA.

La Figura 3.1 muestra también los resultados de variabilidad de las soluciones FBA para cada condición experimental. Se considera que no hay variabilidad cuando la diferencia entre el valor máximo y mínimo de una reacción en una predicción es inferior a  $1 \times 10^{-5}$  [20]. Si no hay variabilidad en las velocidades de reacción estudiadas quiere decir que todas las soluciones posibles tienen los mismos valores de velocidades de reacción estudiadas. Si todas las velocidades  $v_j$  no tienen variabilidad entonces la solución es única. A mayor variabilidad, mayor número de soluciones al problema FBA. La variabilidad de las reacciones con mediciones experimentales osciló entre  $2 \times 10^{-3}$  y  $3,48 \frac{mmol}{gDW\ h}$ . Las reacciones  $v_{EX\_gly(e)}$  y  $v_{EX\_ace(e)}$  tuvieron variabilidad únicamente en la condición experimental 2. La reacción  $v_{EX\_pyr(e)}$  no tuvo variabilidad en las condiciones 1 y 3. La Figura 3.1(c) muestra que el valor experimental de  $v_{EX\_gly(e)}$  no está entre las predicciones máxima y mínima. Las predicciones máxima y mínima del succinato, acetato y piruvato solo contienen los valores de la condición experimental 2 (Figuras 3.1(d), 3.1(e) y 3.1(f)).

La Figura 3.2 muestra los resultados de realizar FVA fijando el valor de función objetivo al valor experimental de  $v_{gw}$ . La variabilidad aumentó, oscilando entre 0,81 y  $36,54 \frac{mmol}{gDW\ h}$  para las reacciones con valor experimental. A diferencia de la Figura 3.1, los valores de las condiciones experimentales están entre el valor mínimo y máximo de cada reacción. Estos resultados requieren un valor de crecimiento diferente al óptimo para explicar el comportamiento experimental de  $v_{EX\_gly(e)}$ ,  $v_{EX\_succ(e)}$ ,  $v_{EX\_ace(e)}$  y  $v_{EX\_pyr(e)}$ . Este valor de crecimiento es subóptimo comparado con el obtenido por el FBA. Lo anterior implica que fijar el crecimiento a un valor subóptimo, permite que la diferencia entre la predicción FBA de crecimiento  $v_{gw}$  y su respectivo valor experimental pueda ser utilizada en la producción de componentes como el glicerol  $v_{EX\_gly(e)}$ .

Aunque la exactitud de las predicciones es similar a la encontrada en la literatura, los FVA realizados en este trabajo muestran que se requiere una función objetivo que permita al FBA explicar completamente el comportamien-



**Figura 3.1:** resultados de FBA y FVA bajo las condiciones experimentales de la Tabla 3.1. (a) FBA del crecimiento del sistema. FBA y FVA de (b) etanol, (c) glicerol, (d) succinato, (e) acetato, (f) piruvato y (g) CO<sub>2</sub>

to experimental. Es necesario un estudio de la técnica FBA y la reconstrucción iMM904 utilizando propuestas alternativas de funciones objetivo.

---

## 3.2. FBA usando Funciones Objetivo basadas en Compartimentos

### Simulaciones

Se utilizaron los compartimentos citosol, mitocondria y peroxisoma disponibles en la reconstrucción IMM904, los cuales se denominarán de aquí en adelante  $P_1$ ,  $P_2$  y  $P_3$ . Para el compartimento  $P_1$  se propuso 13 posibles funciones objetivo, para el compartimento  $P_2$  4 posibles funciones objetivo y para el compartimento  $P_3$  5 posibles funciones objetivo. Estas funciones objetivo se denominarán de aquí en adelante  $C_1$ ,  $C_2$  y  $C_3$  respectivamente.

Para la construcción de las funciones  $C_1$  se eligió la maximización y minimización de las velocidades de reacción que corresponden a la producción y/o consumo de los componentes nadh, nadph, atp, etanol, dióxido de carbono, glicerol, succinato y acetato los cuales están presentes en el compartimento  $P_1$ . Las funciones  $C_2$  consisten en la maximización y minimización de las velocidades de producción y/o consumo de los componentes nadh, nadph y atp del compartimento  $P_2$ . Por último para las funciones objetivo  $C_3$  se eligió la maximización y minimización de las reacciones que producen y/o consumen ácidos grasos, nadh, nadph y atp que se encuentran en el compartimento  $P_3$ . La efectividad de estas funciones objetivo ha sido estudiada en la literatura de forma individual sin tener en cuenta compartimentos físicos del microorganismo estudiado[21].

Se propuso un nuevo compartimento representado por  $P_4$  cuyas funciones objetivo, denominadas  $C_4$  son la maximización y la minimización de biomasa. La Tabla 3.4 lista las funciones objetivo propuestas en este trabajo para cada compartimento.

Se construyeron en total 1259 matrices  $C$  que corresponden a las posibles combinaciones de compartimentos y funciones objetivo de cada uno de estos. Por cada matriz  $C$  se construyó el vector  $w = \mathbf{1}_K$ , donde  $\mathbf{1}_K$  es un vector de

<b>C<sub>1</sub></b>	↑ ← nadh y nadph	<b>C<sub>2</sub></b>	↑ ← atp
	↑ → nadh y nadph		↑ ← nadh y nadph
	↓ ← nadh y nadph		↓ → nadh y nadph
	↓ → nadh y nadph		↑ trans atp citosol
	↓ ← nadh		
	↑ ← atp	<b>C<sub>3</sub></b>	↑ ← ácidos grasos
	↑ → atp		↑ → atp
	↓ → atp		↓ → atp
	↑ ← acetato		↑ ← nadh y nadph
	↑ ← CO <sub>2</sub>		↓ ← nadh y nadph
	↑ ← etanol	<b>C<sub>4</sub></b>	↓ ← biomasa
	↑ ← glycerol		↑ ← biomasa
↑ ← succinato			

**Cuadro 3.4:** posibles funciones objetivo para los compartimentos  $P_1$ ,  $P_2$ ,  $P_3$  y  $P_4$ . ↑ maximización, ↓ minimización, → consumo, ← producción, trans transporte.

1s con longitud  $K$  igual al número de compartimentos incluidos en  $\mathbf{C}$ .

Por cada matriz  $\mathbf{C}$  construida se formularon 12 instancias del problema (2.1) que representan 12 condiciones experimentales, divididas en 4 condiciones anaeróbicas en quimiostato [27] y 8 aeróbicas tipo batch [30]. La condición anaeróbica significa que el medio en el que se encuentra el microorganismo carece de oxígeno y la aeróbica implica que hay presencia de oxígeno. Las Tablas 3.5 y 3.6 muestran las condiciones y componentes estudiados.

cond #	glucosa $\left(\frac{mmol}{gDW \cdot h}\right)$	etanol $\left(\frac{mmol}{gDW \cdot h}\right)$	crecimiento $\left(\frac{1}{h}\right)$
1	5,56	8,28	0,10
2	11,50	17,12	0,20
3	17,37	25,74	0,30
4	23,65	35,27	0,40

**Cuadro 3.5:** resultados experimentales en condiciones anaeróbicas extraídos de [27].

Cada una de las 12 instancias del problema (2.1) formuladas para cada

cond #	glucosa ( $\frac{mmol}{gDW h}$ )	etanol ( $\frac{mmol}{gDW h}$ )	co <sub>2</sub> ( $\frac{mmol}{gDW h}$ )	crecimiento ( $\frac{1}{h}$ )
1	7,2	9,0	15,4	0,16
2	10,2	11,2	20,6	0,17
3	12,2	15,6	23,1	0,23
4	12,3	15,2	24,0	0,21
5	15,1	20,1	25,5	0,33
6	18,4	28,2	32,0	0,36
7	19,9	29,6	33,8	0,40
8	20,2	30,0	34,4	0,40

**Cuadro 3.6:** resultados experimentales en condiciones aeróbicas tipo batch extraídos de [30].

una de las matrices **C** se resolvió utilizando el software GLPK [29] con interfaz disponible para MatLab® a través del paquete COBRA, mediante un equipo con procesador Intel® Core i3™ y memoria RAM de 4GB DDR3. En total se resolvieron  $12 \times 1259 = 15108$  instancias del problema (2.1).

La lista de restricciones por defecto de la IMM904 y las restricciones del problema (2.1) extraídas de [27] se presentan en la Tabla 3.7. Esta lista no incluye todas las restricciones, solo aquellas relevantes para el conjunto de datos experimentales de [27].

En las condiciones aeróbicas tipo batch ([30]) se modificó el límite inferior  $\alpha_{EX_{o2}(e)}$  de consumo de oxígeno por defecto de la reconstrucción IMM904 el cual pasó de  $-2 \frac{mmol}{gDW h}$  a  $-1000 \frac{mmol}{gDW h}$ . Se cambió el límite inferior de consumo oxígeno para que los resultados obtenidos no fueran producto de esa restricción [16] sino del desempeño de las funciones objetivo. Además el consumo de oxígeno puede ser superior a  $2 \frac{mmol}{gDW h}$  (valores de hasta  $10 \frac{mmol}{gDW h}$  son comunes).

El mantenimiento interno de energía del sistema, representado en el modelo por  $v_{APTM}$  se conserva igual al provisto por la reconstrucción para todas las condiciones experimentales estudiadas. Para todas las condiciones expe-

	iMM904		FBA	
	$\alpha_j$	$\beta_j$	$\alpha_j$	$\beta_j$
	$\left(\frac{mmol}{gDW\ h}\right)$	$\left(\frac{mmol}{gDW\ h}\right)$	$\left(\frac{mmol}{gDW\ h}\right)$	$\left(\frac{mmol}{gDW\ h}\right)$
$v_{EX\_o2}(e)$	<b>-2</b>	<b>1000</b>	<b>0</b>	<b>0</b>
$v_{EX\_ergst}(e)$	<b>0</b>	1000	<b>-1000</b>	1000
$v_{EX\_zymst}(e)$	<b>0</b>	1000	<b>-1000</b>	1000
$v_{EX\_hdcea}(e)$	<b>0</b>	1000	<b>-1000</b>	1000
$v_{EX\_ocdcea}(e)$	<b>0</b>	1000	<b>-1000</b>	1000
$v_{EX\_ocdcea}(e)$	<b>0</b>	1000	<b>-1000</b>	1000
$v_{EX\_ocdcya}(e)$	<b>0</b>	1000	<b>-1000</b>	1000
$v_{ATPM}$	1	1	1	1

**Cuadro 3.7:** restricciones  $\alpha_j$  y  $\beta_j$  presentes en la reconstrucción iMM904 y en el problema (2.1) de este trabajo para las condiciones experimentales de [27]. En negrilla se resaltan cambios entre los valores de la reconstrucción y los utilizados en la formulación de (2.1).

rimentales se fijó el valor de consumo de glucosa  $v_{EX\_glc}(e)$  al correspondiente valor registrado. Las demás restricciones se mantuvieron igual a las suministradas por la reconstrucción iMM904.

Se midió el desempeño de la predicción de cada una de las matrices **C** en los dos grupos de condiciones experimentales. Este desempeño se representó en términos del promedio del error en la predicción de crecimiento  $e_{gw} = |(v_{gw} - v_{gw}^E)/v_{gw}^E|$  donde  $v_{gw}^E$  corresponde al valor experimental de crecimiento. También se determinó el promedio del error en la predicción de etanol  $e_{EX\_etoh}(e) = |(v_{EX\_etoh}(e) - v_{EX\_etoh}^E)/v_{EX\_etoh}^E|$  donde  $v_{EX\_etoh}^E$  corresponde al valor experimental de etanol.

Por último, se compararon las predicciones de etanol y crecimiento obtenidas por la matriz **C** con el menor error promedio en la predicción de crecimiento  $e_{gw}$ , las predicciones de estos bioproductos obtenidas por la función objetivo del FBA tradicional, la maximización de biomasa [21], y los resultados experimentales registrados en [27] y [30]. Estas comparaciones se hicieron en términos del error promedio en la predicción de crecimiento  $e_{gw}$  y de etanol

tipo condicion	$C_1$	$C_2$	$C_3$	$C_4$	$\bar{e}_{gw}$
anaero	↓ ← <b>nadh</b>	-	-	↑ ← <b>biomasa</b>	<b>0,19</b>
	↓ ← <b>nadh</b>	-	↓ ← nadh y nadph	↑ ← <b>biomasa</b>	<b>0,19</b>
	↓ ← <b>nadh</b>	-	↓ → atp	↑ ← <b>biomasa</b>	<b>0,19</b>
	↓ ← <b>nadh</b>	-	↑ ← acidos grasos	↑ ← <b>biomasa</b>	<b>0,19</b>
	↓ ← <b>nadh</b>	-	↑ ← nadh y nadph	↑ ← <b>biomasa</b>	<b>0,19</b>
	↓ ← <b>nadh</b>	-	↑ → atp	↑ ← <b>biomasa</b>	<b>0,19</b>
	-	-	-	↑ ← biomasa	0,21
	-	-	↓ ← nadh y nadph	↑ ← biomasa	0,21
	-	-	↓ → atp	↑ ← biomasa	0,21
	-	-	↑ ← acidos grasos	↑ ← biomasa	0,21
aero (batch)	↓ ← <b>nadh</b>	↓ → <b>nadh y nadph</b>	-	↑ ← <b>biomasa</b>	<b>0,10</b>
	↓ ← <b>nadh</b>	↓ → <b>nadh y nadph</b>	↑ ← acidos grasos	↑ ← <b>biomasa</b>	<b>0,10</b>
	↓ ← <b>nadh</b>	↓ → <b>nadh y nadph</b>	↑ → atp	↑ ← <b>biomasa</b>	<b>0,10</b>
	↓ ← <b>nadh</b>	↓ → <b>nadh y nadph</b>	↓ ← nadh y nadph	↑ ← <b>biomasa</b>	<b>0,10</b>
	↓ ← <b>nadh</b>	↓ → <b>nadh y nadph</b>	↓ → atp	↑ ← <b>biomasa</b>	<b>0,10</b>
	↓ ← <b>nadh</b>	↓ → <b>nadh y nadph</b>	↑ ← nadh y nadph	↑ ← <b>biomasa</b>	<b>0,10</b>
	↓ ← nadh y nadph	-	↓ ← nadh y nadph	↑ ← biomasa	1,00
	↓ → nadh y nadph	-	↓ ← nadh y nadph	↑ ← biomasa	1,00
	↓ ← nadh y nadph	-	-	↑ ← biomasa	1,00
	↓ → nadh y nadph	-	-	↑ ← biomasa	1,00

**Cuadro 3.8:** diez matrices  $C$  con los promedios de error en la predicción de crecimiento  $e_{gw}$  ( $\bar{e}_{gw}$ ) más bajos de todas las combinaciones de compartimentos y funciones objetivo. (-) Función objetivo de compartimento no incluida en  $C$ .

$e_{EX\_etoh}(e)$ .

## Resultados

La Tabla 3.8 muestra las diez matrices  $C$  que producen los promedios de error más bajos en la predicción de crecimiento  $\bar{e}_{gw}$ , para cada una de las condiciones experimentales de las Tablas 3.5 y 3.6.

Las seis primeras matrices  $C$  usadas en condiciones anaeróbicas tienen promedios iguales  $\bar{e}_{gw}$  (diferencias menores a  $1 \times 10^{-2}$ ). Las funciones objetivo  $C_1$  y  $C_4$  son la misma para estas seis matrices. La función objetivo  $C_2$  no se incluye. La diferencia en las primeras seis matrices radica en la función objetivo  $C_3$ . Cada una de las alternativas de función objetivo  $C_3$  aparece en los seis primeros lugares. También existen una matriz  $C$  en la que no se incluye alguna función objetivo  $C_3$  (Fila 1 de resultados en condiciones anaeróbicas en la Tabla 3.8). Lo anterior implica que el menor promedio de error de predicción

---

de crecimiento se obtiene construyendo **C** con minimización de producción de nadh como función objetivo **C**<sub>1</sub> y maximización de biomasa como función **C**<sub>4</sub>. La combinación de la función objetivo maximización de biomasa en el compartimento  $P_4$  con la minimización de producción de nadh en el compartimento  $P_1$  reduce el error en la predicción de crecimiento  $\bar{e}_{gw}$  en un 10% con respecto a la predicción obtenida con la función objetivo de maximización de biomasa. El menor valor de  $\bar{e}_{gw}$  encontrado es de 0,19.

En condiciones aeróbicas en cultivos tipo batch se observa que son seis matrices las que obtienen el menor promedio  $\bar{e}_{gw}$ . Para estas seis matrices el promedio  $\bar{e}_{gw}$  es el mismo. Estas matrices difieren en la función objetivo **C**<sub>3</sub>. Lo anterior indica que incluir o no cualquier alternativa de la función objetivo **C**<sub>3</sub> no afecta significativamente el valor de  $\bar{e}_{gw}$ . Las seis matrices tienen como función objetivo **C**<sub>1</sub> la minimización de la producción de nadh, en **C**<sub>2</sub> la minimización de consumo de nadh y nadph y en **C**<sub>4</sub> la maximización de la producción de biomasa. El mínimo valor  $\bar{e}_{gw}$  obtenido para este tipo de condiciones experimentales es 0,1.

En condiciones anaeróbicas la función objetivo maximización de biomasa obtuvo un error de predicción promedio de crecimiento  $\bar{e}_{gw}$  de 0,21. La matriz **C** encontrada obtuvo un  $\bar{e}_{gw}$  de 0,19. Esto implica que la matriz **C** logró una reducción del 10% en el valor de  $\bar{e}_{gw}$ . En cuanto a la producción de etanol la función objetivo de maximización de biomasa obtuvo un error de predicción promedio de etanol  $\bar{e}_{EX\_etoh(e)}$  de 0,12. La matriz **C** encontrada obtuvo un  $\bar{e}_{EX\_etoh(e)}$  de 0,03. La matriz **C** logró una reducción del 75% en el valor de  $\bar{e}_{EX\_etoh(e)}$ . La Figura 3.3 muestra el desempeño de la función objetivo maximización de producción de biomasa, la matriz **C** encontrada y el comportamiento experimental en condiciones anaeróbicas. Los resultados para cada condición anaeróbica son mostrados en la Tabla 3.9.

En condiciones aeróbicas tipo batch la función objetivo maximización de biomasa obtuvo un error de predicción promedio de crecimiento  $\bar{e}_{gw}$  de 4,04. La matriz **C** encontrada obtuvo un  $\bar{e}_{gw}$  de 0,1. Esto implica que la matriz **C**

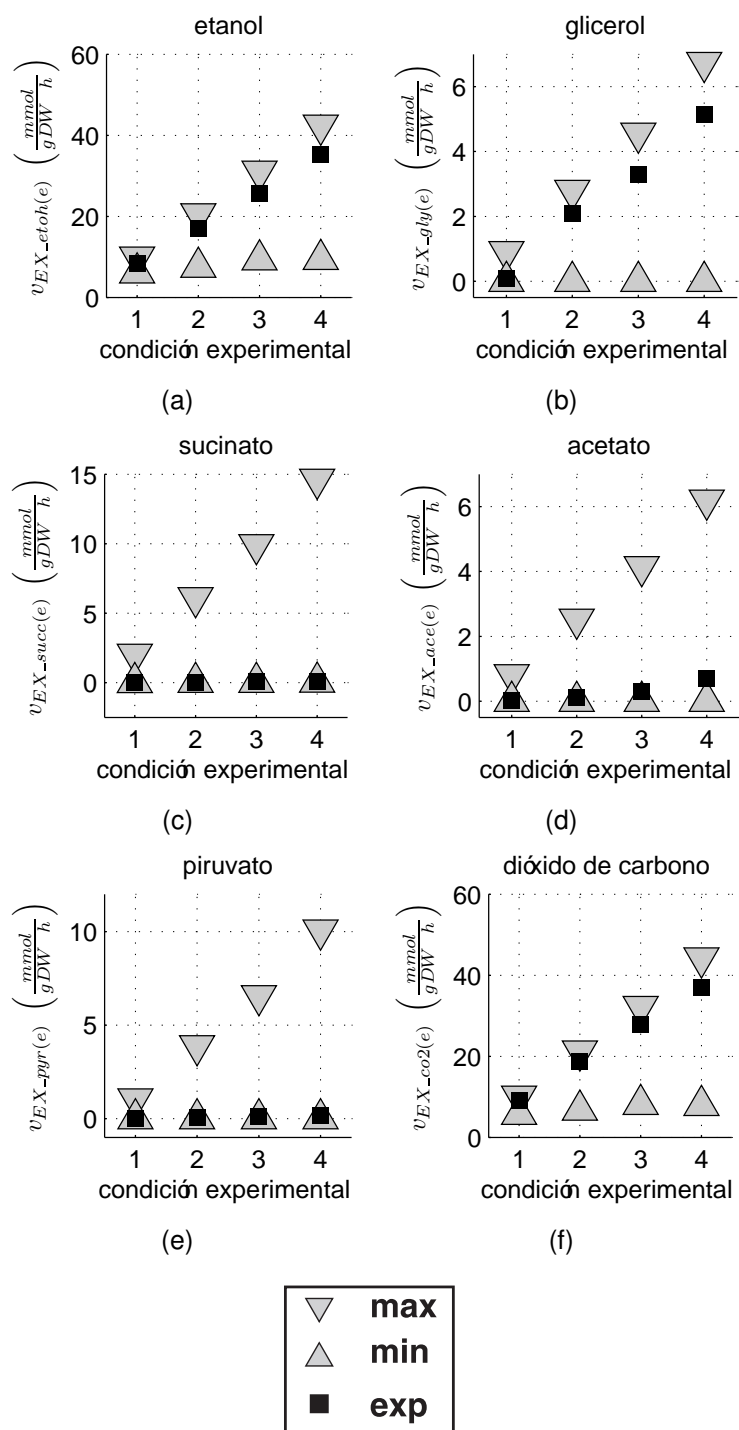
cond #	biomasa		propuesta	
	$e_{gw}$	$e_{EX\_etoh(e)}$	$e_{gw}$	$e_{EX\_etoh(e)}$
1	0,12	0,13	<b>0,09</b>	<b>0,05</b>
2	0,21	0,12	<b>0,19</b>	<b>0,03</b>
3	0,24	0,12	<b>0,22</b>	<b>0,03</b>
4	0,28	0,11	<b>0,25</b>	<b>0,02</b>

**Cuadro 3.9:** resultados obtenidos con la función objetivo propuesta en condiciones anaeróbicas

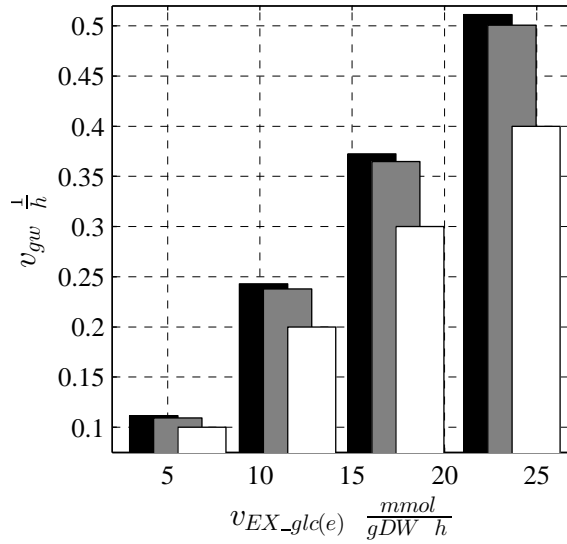
cond #	biomasa		propuesta	
	$e_{gw}$	$e_{EX\_etoh(e)}$	$e_{gw}$	$e_{EX\_etoh(e)}$
1	3,38	1,00	<b>0,12</b>	<b>0,39</b>
2	4,85	1,00	<b>0,21</b>	<b>0,57</b>
3	4,72	1,00	<b>0,19</b>	<b>0,40</b>
4	4,18	1,00	<b>0,07</b>	<b>0,35</b>
5	3,47	1,00	<b>0,07</b>	<b>0,29</b>
6	4,00	1,00	<b>0,05</b>	<b>0,12</b>
7	3,94	1,00	<b>0,04</b>	<b>0,16</b>
8	3,86	1,00	<b>0,02</b>	<b>0,16</b>

**Cuadro 3.10:** resultados obtenidos con la función objetivo propuesta en condiciones aeróbicas

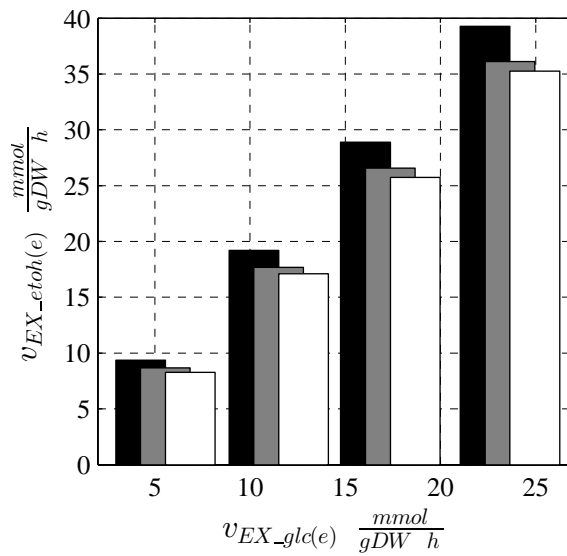
logró una reducción del 98 % en el valor de  $\bar{e}_{gw}$ . En cuanto a la producción de etanol la función objetivo maximización de biomasa obtuvo un error de predicción promedio  $\bar{e}_{EX\_etoh(e)}$  de 1. La matriz **C** encontrada obtuvo un  $\bar{e}_{EX\_etoh(e)}$  a 0,30. Esto implica que la matriz **C** logró una reducción del 70 % en el valor de  $\bar{e}_{EX\_etoh(e)}$ . La Figura 3.4 muestra el desempeño de la función objetivo maximización de producción de biomasa, la matriz **C** encontrada y el comportamiento experimental en condiciones aeróbicas tipo batch. Los valores obtenidos de  $e_{gw}$  y  $e_{EX\_etoh(e)}$  para cada condición aeróbica tipo batch son mostrados en la Tabla 3.10.



**Figura 3.2:** FVA bajo las condiciones experimentales de la Tabla 3.1 y valor de función objetivo igual al valor experimental de crecimiento. FVA de (a) etanol, (b) glicerol, (c) succinato, (d) acetato, (e) piruvato y (f) CO<sub>2</sub>



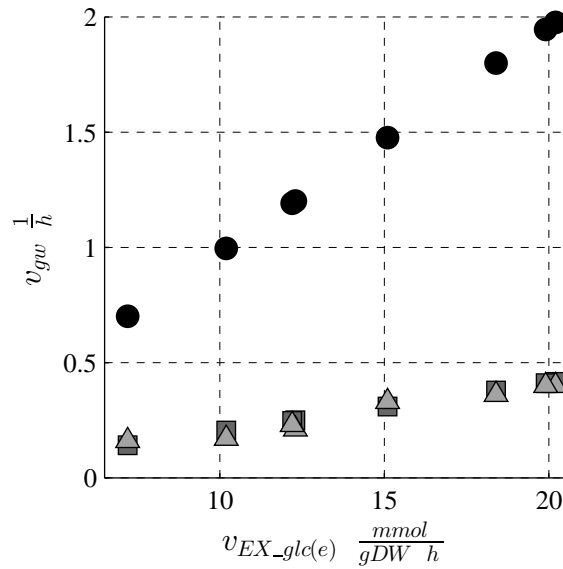
(a)



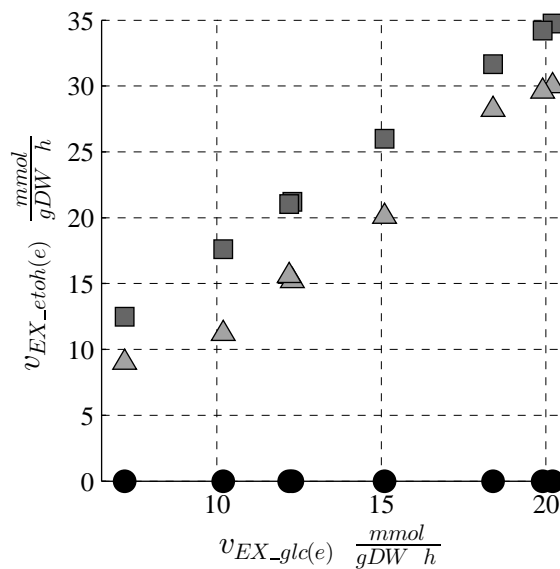
(b)

- maximización de biomasa (FBA Tradicional)
- matriz C encontrada
- condición experimental

**Figura 3.3:** resultados obtenidos mediante función objetivo maximización de biomasa, la matriz **C** encontrada y condiciones experimentales anaeróbicas. a) Comportamiento del crecimiento del sistema y b) comportamiento del etanol vs consumo de glucosa.



(a)



(b)

- maximización de biomasa (FBA Tradicional)
- matriz C encontrada
- ▲ condición experimental

**Figura 3.4:** resultados obtenidos mediante función objetivo maximización de biomasa, la matriz **C** encontrada y condiciones experimentales aeróbicas tipo batch. a) Comportamiento del crecimiento del sistema y b) comportamiento del etanol vs consumo de glucosa.

# Capítulo 4

## Conclusiones

Las reacciones con las mejores predicciones FBA dentro del conjunto de datos de [27] son producción de CO<sub>2</sub>  $v_{EX\_co2(e)}$ , con errores entre el 0.3% y el 4.5%; producción de etanol  $v_{EX\_etoh(e)}$ , con errores entre el 11% y 13% y crecimiento  $v_{gw}$  con errores entre el 11% y 28%. El FBA y la reconstrucción iMM904 predice la producción de CO<sub>2</sub> con una exactitud superior a la registrada en la literatura, para las condiciones experimentales de [27]. La predicción de etanol obtenida es similar a trabajos anteriores.

Los resultados de los FVA realizados en este trabajo permitieron determinar que el FBA y la reconstrucción iMM904 falla en la predicción de glicerol  $v_{EX\_gly(e)}$ , succinato  $v_{EX\_succ(e)}$ , acetato  $v_{EX\_ace(e)}$  y piruvato  $v_{EX\_pyr(e)}$  bajo las condiciones experimentales descritas en [27]. La no variabilidad en la predicción de estos componentes soporta esta afirmación. El aumento de la variabilidad obtenida al realizar el estudio de FVA con un valor subóptimo de función objetivo confirma este resultado.

Este trabajo detalló la formulación FBA aplicada a la reconstrucción iMM904. A diferencia de los trabajos encontrados en la literatura, este trabajo explica cada una de las restricciones utilizadas en el FBA y mostró su equivalencia en las condiciones experimentales utilizadas. Este trabajo describió y mostró la importancia de utilizar el FVA para el análisis de resultados arrojados por el FBA.

---

Combinar la maximización de producción de biomasa con la minimización de producción de nadh en el compartimento citosol de *S. cerevisiae* reduce el error promedio en la predicción de crecimiento en un 10 % en las condiciones anaeróbicas de [27]. Esta combinación también reduce en un 75 % el promedio del error en la predicción de producción de etanol.

Si a la función objetivo encontrada en condiciones anaeróbicas se agrega la minimización de consumo de nadh y nadph en el compartimento mitocondria de *S. cerevisiae*, se reduce el promedio del error en la predicción de crecimiento en un 98 % en las condiciones aeróbicas tipo batch de [30]. Esta nueva función objetivo reduce el promedio del error en la predicción de etanol en un 70 % con respecto a la función de maximización de biomasa.

Agregar funciones objetivo de compartimentos de *S. cerevisiae* a la función objetivo de maximización de biomasa mejora las predicciones obtenidas por la función objetivo maximización de biomasa, propuesta en la literatura [21] y en la reconstrucción a escala genómica iMM904 [5], en condiciones experimentales anaeróbicas en cultivo continuo [27] y aeróbicas tipo batch [30].

# Capítulo 5

## Apéndice A: Código usado en la producción de los resultados

```
function [state fval out error norma]=main(cb_model,obj_fun,fun_comb,t_num)

state=zeros(size(fun_comb,1),1);
fval=nan(size(fun_comb,1),1);
out=nan(size(fun_comb,1),length(cb_model.E));
error=out;
norma=fval;

changeCobraSolver('glpk','lp');

for i=1:size(fun_comb,1)
    cb_model.c=get_c(obj_fun,fun_comb(i,:))';
    rstl=optimizeCbModel(cb_model);
```

---

```

state(i)=rstl.stat;
if rstl.stat>0
    fval(i)=rstl.f;
    out(i,:)=rstl.x(cb_model.E)';
    error(i,:)=abs(abs(rstl.x(cb_model.E))-cb_model.vexp)./cb_model.vexp;
    norma(i,:)=norm(abs(rstl.x(cb_model.E(1:end-1)))-cb_model.vexp(1:end-1))
end
if mod(i,50)==0
    fid=fopen(t_num,'w+');
    fprintf(fid,'%d',i);
    fclose(fid);
end
end
end

```

```

function c=get_c(obj_fun,comb)

obj_cmp=obj_fun(comb>0);
C=cell2mat(cellfun(@get_fun,...
    obj_cmp,num2cell(comb(comb>0))...
    ,'Uniformoutput',false)');

c=ones(1,length(comb(comb>0)))*C;

```

```

function obj_fun=get_fun(comp,fun_index)

```

---

```
obj_fun=comp(fun_index,:);
```

# Referencias

- [1] Tobias Österlund, Intawat Nookaew, and Jens Nielsen. Fifteen years of large scale metabolic modeling of yeast: Developments and impacts. *Bio-technology Advances*, 2011. [11](#)
- [2] Francisco Llaneras and Jesús Picó. Stoichiometric modelling of cell metabolism. *Journal of Bioscience and Bioengineering / Science Direct*, 105(1):1–11, January 2008. [11](#)
- [3] Jochen Forster, Iman Famili, Patrick Fu, Bernhard O. Palsson, and Jens Nielsen. Genome-Scale reconstruction of the saccharomyces cerevisiae metabolic network. *Genome Research / Genome Research*, 13(2):244–253, February 2003. [11](#)
- [4] Natalie C. Duarte, Markus J. Herrgård, and Bernhard O. Palsson. Reconstruction and validation of saccharomyces cerevisiae iND750, a fully compartmentalized Genome-Scale metabolic model. *Genome Research*, 14(7):1298 –1309, July 2004. [11](#)
- [5] Monica Mo, Bernhard Palsson, and Markus Herrgard. Connecting extra-cellular metabolomic measurements to intracellular flux states in yeast. *BMC Systems Biology*, 3(1):37, 2009. [11](#), [14](#), [29](#), [47](#)
- [6] Lars Kuepfer, Uwe Sauer, and Lars M. Blank. Metabolic functions of duplicate genes in saccharomyces cerevisiae. *Genome Research / Genome Research*, 15(10):1421 –1430, October 2005. [12](#)

- 
- [7] Intawat Nookaew, Michael Jewett, Asawin Meechai, Chinae Thammamongtham, Kobkul Laoteng, Supapon Cheevadhanarak, Jens Nielsen, and Sakarindr Bhumiratana. The genome-scale metabolic model iIN800 of *saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. *Systems Biology / BMC*, 2(1):71, 2008. [12](#)
- [8] A. Zomorodi and C. Maranas. Improving the iMM904 *s. cerevisiae* metabolic model using essentiality and synthetic lethality data. *Systems Biology / BMC*, 4(1):178, 2010. [12](#)
- [9] Markus J Herrgard, Neil Swainston, Paul Dobson, Warwick B Dunn, K Yalcin Arga, Mikko Arvas, Nils Buthgen, Simon Borger, Roeland Costenoble, Matthias Heinemann, Michael Hucka, Nicolas Le Novere, Peter Li, Wolfram Liebermeister, Monica L Mo, Ana Paula Oliveira, Dina Petranovic, Stephen Pettifer, Evangelos Simeonidis, Kieran Smallbone, Irena Spasie, Dieter Weichart, Roger Brent, David S Broomhead, Hans V Westerhoff, Betul Kurdar, Merja Penttila, Edda Klipp, Bernhard O Palsson, Uwe Sauer, Stephen G Oliver, Pedro Mendes, Jens Nielsen, and Douglas B Kell. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat Biotech / Nature*, 26(10):1155–1160, October 2008. [12](#)
- [10] Kenneth J Kauffman, Purusharth Prakash, and Jeremy S Edwards. Advances in flux balance analysis. *Current Opinion in Biotechnology / Science Direct*, 14(5):491–496, October 2003. [12](#)
- [11] Jong Min Lee, Erwin P. Gianchandani, and Jason A. Papin. Flux balance analysis in the era of metabolomics. *Briefings in Bioinformatics / Oxford Journals*, 7(2):140–150, June 2006. [12](#)
- [12] Karthik Raman and Nagasuma Chandra. Flux balance analysis of biological systems: applications and challenges. *Brief Bioinform*, 10(4):435–449, 2009. [12](#), [13](#), [20](#)

- 
- [13] Jeffrey D Orth, Ines Thiele, and Bernhard O. Palsson. What is flux balance analysis? *Nature biotechnology / Nature*, 28(3):245–8, March 2010. [12](#)
- [14] Erwin P Gianchandani, Arvind K Chavali, and Jason A Papin. The application of flux balance analysis in systems biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 2(3):372–382, May 2010. [12](#)
- [15] Iman Famili, Jochen Förster, Jens Nielsen, and Bernhard O. Palsson. *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proceedings of the National Academy of Sciences / PNAS*, 100(23):13134 –13139, November 2003. [12](#), [13](#), [33](#)
- [16] J.L. Hjersted and M.A. Henson. Steady-state and dynamic flux balance analysis of ethanol production by *saccharomyces cerevisiae*. *IET Systems Biology*, 3:167, 2009. [12](#), [13](#), [38](#)
- [17] Christoffer Bro, Birgitte Regenber, Jochen Forster, and Jens Nielsen. In silico aided metabolic engineering of *saccharomyces cerevisiae* for improved bioethanol production. *Metabolic Engineering / Elsevier*, 8(2):102–111, March 2006. [13](#)
- [18] Rogier J P van Berlo, Dick de Ridder, Jean-Marc Daran, Pascale a S Daran-Lapujade, Bas Teusink, and Marcel J T Reinders. Predicting metabolic fluxes using gene expression differences as constraints. *transactions on computational biology and bioinformatics / IEEE, ACM*, 8(1):206–16, 2011. [13](#)
- [19] Sangbum Lee, Chan Phalakornkule, Michael M. Domach, and Ignacio E. Grossmann. Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. *Computers and Chemical Engineering / Elsevier*, 2000. [13](#)
- [20] R. Mahadevan and C.H. Schilling. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering / Elsevier*, 2003. [13](#), [25](#), [34](#)

- 
- [21] Adam M. Feist and Bernhard O. Palsson. The biomass objective function. *Current opinion in microbiology*, 13(3):344–349, June 2010. PMID: 20430689 PMCID: PMC2912156. [13](#), [14](#), [15](#), [30](#), [36](#), [39](#), [47](#)
- [22] Marco Terzer, Nathaniel D Maynard, Markus W Covert, and Jörg Stelling. Genome scale metabolic networks. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 1(3):285–297, June 2009. [13](#)
- [23] Erwin Gianchandani, Matthew Oberhardt, Anthony Burgard, Costas Maranas, and Jason Papin. Predicting biological system objectives de novo from internal state measurements. *BMC Bioinformatics*, 9(1):43, 2008. [13](#), [14](#), [15](#)
- [24] Benoît Colson, Patrice Marcotte, and Gilles Savard. An overview of bilevel optimization. *Annals of Operations Research*, 153(1):235–256, April 2007. [13](#)
- [25] Saul I. Gass. *Linear programming: methods and applications*. New York, Courier Dover Publications, p 36, 1985. [20](#), [22](#)
- [26] Komei Fukuda and Alain Prodon. Double description method revisited. In Michel Deza, Reinhardt Euler, and Ioannis Manoussakis, editors, *Combinatorics and Computer Science*, volume 1120 of *Lecture Notes in Computer Science*, pages 91–111. Springer Berlin / Heidelberg, 1996. [22](#)
- [27] Torben L. Nissen, Ulrik Schulze, Jens Nielsen, and John Villadsen. Flux distributions in anaerobic, Glucose-Limited continuous cultures of *saccharomyces cerevisiae*. *Microbiology*, 143(1):203 –218, January 1997. [29](#), [37](#), [38](#), [39](#), [46](#), [47](#)
- [28] Jan Schellenberger, Richard Que, Ronan M T Fleming, Ines Thiele, Jeffrey D Orth, Adam M Feist, Daniel C Zielinski, Aarash Bordbar, Nathan E Lewis, Sorena Rahmanian, Joseph Kang, Daniel R Hyduke, and Bernhard O Palsson. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA toolbox v2.0. *Nat. Protocols*, 6(9):1290–1307, 2011. [30](#)

- 
- [29] Andrew Makhorin. GLPK (gnu linear programming kit), v. 4.47, September 2011. [33](#), [38](#)
- [30] Jan Heyland, Jianan Fu, and Lars M Blank. Correlation between TCA cycle flux and glucose uptake rate during respiro-fermentative growth of *saccharomyces cerevisiae*. *Microbiology*, 155(12):3827–3837, December 2009. [37](#), [38](#), [39](#), [47](#)