

Predicción a escala genómica de componentes de *Saccharomyces cerevisiae* mediante análisis de balance de flujos

Prediction of genome scale of *Saccharomyces cerevisiae* by flux balance analysis

César Augusto Vargas García*, Henry Arguello Fuentes**, Rodrigo Gonzalo Torres Sáez***

Resumen

El microorganismo *Saccharomyces cerevisiae* cuenta con gran número de modelos biológicos conocidos como reconstrucciones, las cuales pueden ser a escala genómica. De estas reconstrucciones a escala genómica provienen los modelos matemáticos, también llamados modelos estequiométricos. 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, CO₂, 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 CO₂, con errores entre el 0.3% y 4.5% y la producción de etanol, con errores entre el 11% y 13%.

Palabras clave: analisis de balance de flujos, reconstrucción a escala genómica, iMM904, *S. cerevisiae*.

Abstract

Several biological models, named reconstructions, are used for the study of the *S. cerevisiae* microorganism. The reconstructions can be genomic scaled. Mathematical models are generated from the reconstructions and they are called stoichiometric models. The flux balance analysis (FBA) is one of the tools used for the analysis of these models. The FBA attempts to predict the evolution of the microorganism and the consumption and production of components like glucose, ethanol, glycerol, succinate, acetate and pyruvate. A Flux variability analysis (FVA) is used to determine the uniqueness of the FBA predictions. This paper shows the results of applying FBA to the iMM904 reconstruction of *S. cerevisiae* and compares them with experimental data from literature. The results in this paper show that it is possible to predict the evolution with errors between 11% and 28% ; the production of CO₂ with errors between 0.3% and 4.5%; and the production of ethanol with errors between 11% and 13%, using FBA for the iMM904 model.

Keywords: flux balance analysis, genome scale reconstructions, iMM904, *S. cerevisiae*.

Recibido: febrero 16 de 2012

Aprobado: junio 20 de 2012

* Ms(c) Ingeniería de Sistemas e Informática, Universidad Industrial de Santander, caugusto.vargas@gmail.com.

** PhD(c) Electrical and Computer Engineering, Docente Asistente Escuela de Ingeniería de Sistemas e Informática, Universidad Industrial de Santander. henarfu@uis.edu.co

*** PhD Bioquímica, Docente Asistente Facultad de Ciencias Básicas, Universidad Industrial de Santander, rtorres@uis.edu.co

Introducción

Debido a que las reservas de petróleo disminuyen rápidamente y dados los actuales problemas ambientales producidos por el uso de combustibles fósiles, el etanol se ha convertido en una fuente alternativa de energía renovable y limpia (Bai *et al.*, 2008). La producción de etanol es realizada por microorganismos como *S. cerevisiae* (conocida como levadura de cerveza). El uso de *S. cerevisiae* y no de otros microorganismos para la producción de etanol se debe a su tolerancia al etanol y a su capacidad de producción de biomasa para el consumo animal (Bai *et al.*, 2008).

Una forma de almacenar la información estructural de *S. cerevisiae* es a través de los modelos estequiométricos (Llaneras y Picó, 2008), 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. (tabla 1).

Las reconstrucciones base de la literatura son la iFF708 (Förster *et al.*, 2003), iND750 (Duarte *et al.*, 2009) e iMM904 (Mo *et al.*, 2009). 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 (Mo *et al.*, 2009). A partir de la iFF708, iND750 e iMM904 se han realizado nuevas reconstrucciones de *S. cerevisiae* tales como la iLL672 (Kuepfer *et al.*, 2005), la iIN800 (Nookaew *et al.*, 2008), la iZ900 (Zomorodi y Maranas, 2010) y la Yeast 4.0 (Herrgard *et al.*, 2008).

Tabla 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

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

La técnica más utilizada para el estudio de los modelos estequiométricos extraídos de estas reconstrucciones es el Análisis de Balance de Flujos - FBA (Kauffman *et al.*, 2003), (Lee *et al.*, 2006), (Raman y Chandra, 2009), (Orth *et al.*, 2010), (Gianchandani *et al.*, 2010). Esta técnica ha cobrado interés debido a que no requiere información de la dinámica del sistema (Kauffman *et al.*, 2003). En cambio, el FBA requiere de una función que represente el objetivo biológico del sistema estudiado. El estudio de funciones objetivo se ha realizado utilizando modelos estequiométricos simples y a escala genómica del microorganismo *E-coli* (Schuetz *et al.*, 2007), (Feist y Palsson, 2010) y otros microorganismos (Gong *et al.*, 2009), (Gong y Yu, 2010). También se ha estudiado modelos estequiométricos reducidos de *S. cerevisiae* (Burgard y Maranas, 2003), (Gianchandani *et al.*, 2008). Sin embargo hasta la fecha no se han explorado nuevas funciones objetivo del FBA usando redes a escala genómica de *S. cerevisiae*.

El trabajo realizado en (Famili *et al.*, 2003) 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 (Famili *et al.*, 2003) y (Hjersted y Henson, 2009) y se han realizado estudios posteriores como en (Bro *et al.*, 2006) y (van Berlo *et al.*, 2011) utilizando estos modelos. En contraste el modelo iMM904 carece de estudios como el de (Famili *et al.*, 2003) y (Hjersted y Henson, 2009). 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 (Lee *et al.*, 2000). 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 (Mahadevan y Schilling, 2003).

La literatura actual carece de un trabajo que estudie la producción y consumo de los componentes etanol, glicerol, CO₂ 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 (Mo *et al.*, 2009), son necesarios trabajos que exploren otros componentes además de la glucosa, tales como el etanol el cual es importante en la producción de biocombustibles.

El estudio de este modelo es importante debido a que la reconstrucción iMM904 es la más detallada dentro de las reconstrucciones que han sido base para las reconstrucciones actuales (Mo *et al.*, 2009). También se requiere la validación de la capacidad de predicción de la técnica FBA utilizando dicho modelo. Esto permite determinar la aplicabilidad de este modelo en futuros trabajos sobre el organismo *S. cerevisiae*, y además muestra cuales son los problemas de predicción, con el fin de proponer futuros estudios para solucionarlos.

El uso de FBA con modelos extraídos de reconstrucciones genómicas ha permitido evaluar genes esenciales de crecimiento, reducción, o no esenciales; y por tanto permite la identificación de objetivos en el desarrollo de medicamentos usados en tratamientos como el de la tuberculosis (Raman *et al.*, 2005). Otra de las aplicaciones del FBA y modelos genómicos es la predicción de respuesta de microorganismos ante eliminación o reducción de expresión genética. Estas predicciones son usadas para realizar ingeniería metabólica con el fin de obtener bioproductos deseados (Manichaikul *et al.*, 2009). Usar la reconstrucción iMM904 permite estudiar *S. cerevisiae* a un nivel de detalle mucho mayor que los modelos iFF708 e iND750.

El objetivo del presente artículo 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 artículo 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 CO₂ y etanol del organismo *S. cerevisiae*, y los correspondientes valores experimentales.

Este artículo está organizado de la siguiente forma: la sección de métodos explica el modelo estequiométrico y describe las técnicas FBA y FVA, y el por qué de la existencia de múltiples soluciones de un FBA. La sección de resultados muestra el resultado de aplicar las técnicas FBA y FVA al modelo estequiométrico obtenido de la reconstrucción iMM904. La sección final indica las conclusiones del presente trabajo.

Materiales y métodos

Modelo estequiométrico

En el modelo estequiométrico, la variación de la cantidad de un componente a través del tiempo v_j^E , se define como una combinación lineal

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

donde v_j son las velocidades de reacción que lo afectan, S_{ij} 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

$$\frac{d\mathbf{n}}{dt} = \mathbf{S}\mathbf{v} = \mathbf{0}_M, \quad (2)$$

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 muestra un ejemplo de un modelo estequiométrico. El sistema está compuesto por cuatro componentes cuyas cantidades son: ρ_1, ρ_2, n_1, n_2 y cuatro reacciones cuyas velocidades son: v_1, v_2, v_3 y v_4 . ρ_1 y ρ_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_1 y ρ_w $w = 1, \dots, W$, y cuyas aristas corresponden a las reacciones v_j (figura 1(a)).

El conjunto de reacciones que describe el sistema se muestra en la figura 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(b); por ejemplo, la reacción que ocurre a la velocidad v_1 toma una unidad del componente ρ_1 y produce dos unidades de n_1 .

Para construir la matriz $\bar{\mathbf{S}}$ se usa la matriz de incidencia $\bar{\mathbf{S}}$ de dimensiones $(W + M) \times N$, mostrada en la figura 1(c). Las primeras W filas de $\bar{\mathbf{S}}$ corresponden a los componentes ρ_w , y las restantes corresponden a los componentes n_i . Las columnas de $\bar{\mathbf{S}}$ corresponden a las reacciones con velocidades v_j .

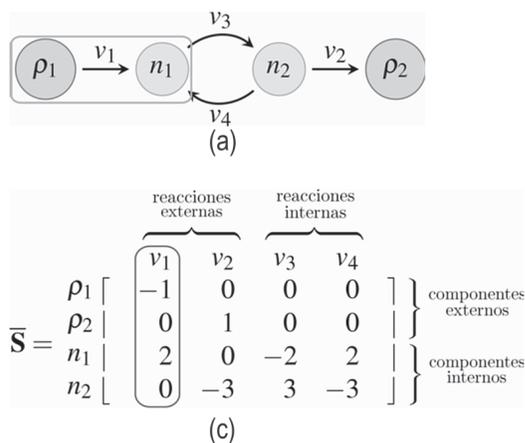


Figura 1: Ejemplo de modelo estequiométrico. (a) grafo del sistema, (b) reacciones químicas, (c) matriz de incidencia, (d) matriz estequiométrica y (e) vector de velocidades de reacción.

En la figura 1(c) la columna 1 de la matriz \bar{S} corresponde a la reacción con velocidad v_1 . Esta columna fue construida tomando los coeficientes que aparecen en la figura 1(b). En esta reacción no participan los componentes ρ_2 y n_2 , por tanto $\bar{S}_{2,1} = \bar{S}_{4,1} = 0$. El componente ρ_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 ρ_1 con signo negativo, es decir $\bar{S}_{1,1} = -1$. 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 $\bar{S}_{w+ij} < 0$ entonces v_j consume $|\bar{S}_{w+ij}|$ unidades del componente n_i , si $\bar{S}_{w+ij} > 0$ entonces la reacción v_j produce el componente n_i \bar{S}_{w+ij} unidades, y si $\bar{S}_{w+ij} = 0$ el componente n_i no participa en la reacción v_j . De forma análoga, \bar{S}_{w+ij} representa la participación del componente ρ_w en la reacción v_j .

La matriz S se obtiene eliminando las primeras W filas de \bar{S} que corresponden a componentes externos. La figura 1(d) muestra la matriz S resultado de eliminar las filas $w = 1, 2$ de \bar{S} , que corresponden a los componentes ρ_1 y ρ_2 . En S se cumple que $M < N$ ($M = 2$ y $N = 4$). Finalmente, la figura 1(e) muestra el vector de velocidades de reacción v .

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

$$\rho_1 \xrightarrow{v_1} 2n_1, \quad 2n_1 \xrightleftharpoons[v_4]{v_3} 3n_2, \quad 3n_2 \xrightarrow{v_2} \rho_2$$

(b)

$$S = \begin{matrix} & v_1 & v_2 & v_3 & v_4 \\ n_1 [& 2 & 0 & -2 & 2 \\ n_2 [& 0 & -3 & 3 & -3 \end{matrix}$$

(d)

$$v = [v_1 \quad v_2 \quad v_3 \quad v_4]^T$$

(e)

$$\frac{dn}{dt} = Sv = \mathbf{0}_M, \quad (3)$$

donde $\mathbf{0}_M$ es un vector de longitud M cuyos elementos son todos cero. La ecuación matricial (3) define un sistema de ecuaciones lineales subdeterminado con infinitas soluciones. Para delimitar el espacio solución de la ecuación (3) se definen los límites de cada una de las velocidades de reacción v_j del sistema mediante

$$\alpha_j \leq v_j \leq \beta_j, \quad j = 1, \dots, N, \quad (4)$$

donde α_j y β_j son los límites inferior y superior de v_j . Por último se establece una función objetivo

$$f = \mathbf{c}^T v, \quad (5)$$

donde $c = [c_1, \dots, c_N]$ 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 S , los vectores $\alpha = [\alpha_1, \dots, \alpha_N]$, $\beta = [\beta_1, \dots, \beta_N]$ y c , se busca un vector que sea solución del problema de programación lineal

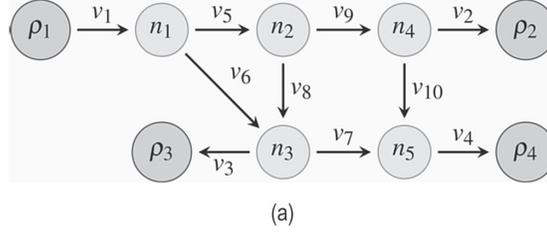
$$\begin{aligned} \max \quad & f = \mathbf{c}^T v \\ \text{s.a.} \quad & \\ & Sv = \mathbf{0}_M \\ & v_j \geq \alpha_j \\ & v_j \leq \beta_j \end{aligned} \quad (6)$$

la búsqueda de soluciones al problema (6) se conoce como FBA.

La figura 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 $n_1,$

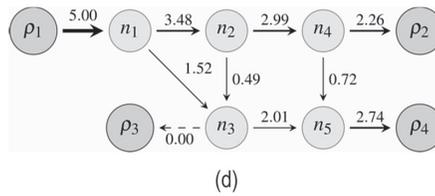
n_2, n_3, n_4 y n_5 . El grafo que describe el sistema se muestra en la figura 2(a). La matriz estequiométrica S se define en la figura 2(b). α , β y c se muestran en la figura 2(c). El valor negativo de α_1 indica que la dirección de la reacción con velocidad v_1 puede ser en cualquiera de dos sentidos: 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 (Raman y Chandra, 2009).



	reacciones externas				reacciones internas					
	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}
n_1	1	0	0	0	-1	-1	0	0	0	0
n_2	0	0	0	0	1	0	0	-1	-1	0
n_3	0	0	-1	0	0	1	-1	1	0	0
n_4	0	-1	0	0	0	0	0	0	1	-1
n_5	0	0	0	-1	0	0	1	0	0	1

β^T	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}
α^T	5	4	3	3	4	5	6	7	8	8
c^T	-2	0	0	0	0	0	0	0	0	0
v^T	0	1	0	1	0	0	0	0	0	0
	5.00	2.26	0.00	2.74	3.48	1.52	2.01	0.49	2.99	0.72



```
v=linprog(-c, [], [], S, zeros(M,1), alfa, beta)
c=c, S=S, 0_M = zeros(M,1), alpha=alfa, beta=beta
```

Figura 2: Principales elementos de un problema FBA. (a) grafo del sistema, (b) matriz estequiométrica S , (c) vector de velocidades máximas β y mínimas α , vector de función objetivo c y vector solución v encontrado al resolver el FBA. En (d) se representa gráficamente este resultado. El valor máximo de la función objetivo f es 5.

El vector de coeficientes \mathbf{c} define la función objetivo del sistema de la figura 2. En este ejemplo el vector contiene valores diferentes de $\mathbf{0}$ 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 p_2 y p_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 2(c) se obtuvo a través de programación lineal utilizando el paquete matemático MatLab®. La figura 2(d) muestra de forma gráfica la interpretación de este resultado. Por último la figura 2(e) muestra el código utilizado en MatLab para obtener esta solución.

Soluciones del FBA

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

$$P = \{ \mathbf{v} | S\mathbf{v} = \mathbf{0}_M = \alpha_j \leq v_j \leq \beta_j \} \quad (7)$$

Como lo demuestra (Gass, 1985), todo poliedro convexo, cerrado y acotado tiene la representación

$$P = \left\{ \mathbf{v} \mid \mathbf{v} = \sum_{k=1}^L \lambda_k \mathbf{u}_k, \lambda_k \geq 0, \sum_{k=1}^L \lambda_k = 1 \right\} \quad (8)$$

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 (Gass, 1985) indica que el vector que resuelve la ecuación (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 (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 $k \neq Q_r$, $\mathbf{c}^T \mathbf{u}_{Q_1} = \dots = \mathbf{c}^T \mathbf{u}_{Q_R}$. Más aún, los vectores de la forma

$$\mathbf{v} = \sum_{r=1}^R \lambda_r \mathbf{u}_{Q_r}, \lambda_r \geq 0, \sum_{r=1}^R \lambda_r = 1, \quad (9)$$

también son solución de la ecuación (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 (6). Si existe más de un punto extremo con el máximo valor de función objetivo posible, la ecuación (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 2. Se obtuvieron $L = 35$ puntos extremos mediante el Método de Descripción Doble (cdd) (Fukuda and Prodon, 1996), con implementación disponible para uso en MatLab. En la figura 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 $\mathbf{5}$, el cual corresponde a los puntos \mathbf{u}_{24} , \mathbf{u}_{33} , \mathbf{u}_{34} y \mathbf{u}_{35} .

En la figura 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 (6). El vector solución \mathbf{v} mostrado en la figura 2 se puede expresar como la combinación lineal convexa de los puntos extremos \mathbf{u}_{Q_r} indicados en la figura 3. Esto se indica en la figura 4. Para encontrar los valores λ_r mostrados en la figura 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 2.

Análisis de variabilidad de flujos – FVA

Debido a la existencia de múltiples soluciones de (6), se requiere determinar el intervalo de posibles valores de v_j . El Análisis Variabilidad de Flujos (Mahadevan y Schilling, 2003) 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

$$\begin{aligned} \min \quad & v_j \\ & v_j \\ \text{s.a.} \quad & \\ & S\mathbf{v} = \mathbf{0}_M \\ & v_j \geq \alpha_j \\ & v_j \leq \beta_j \\ & \mathbf{c}^T \mathbf{v} = f \end{aligned} \quad (10)$$

y la segunda optimización consiste en maximizar v_j resolviendo

$$\begin{aligned} \max \quad & v_j \\ & v_j \\ \text{s.a.} \quad & \\ & S\mathbf{v} = \mathbf{0}_M \\ & v_j \geq \alpha_j \\ & v_j \leq \beta_j \\ & \mathbf{c}^T \mathbf{v} = f \end{aligned} \quad (11)$$

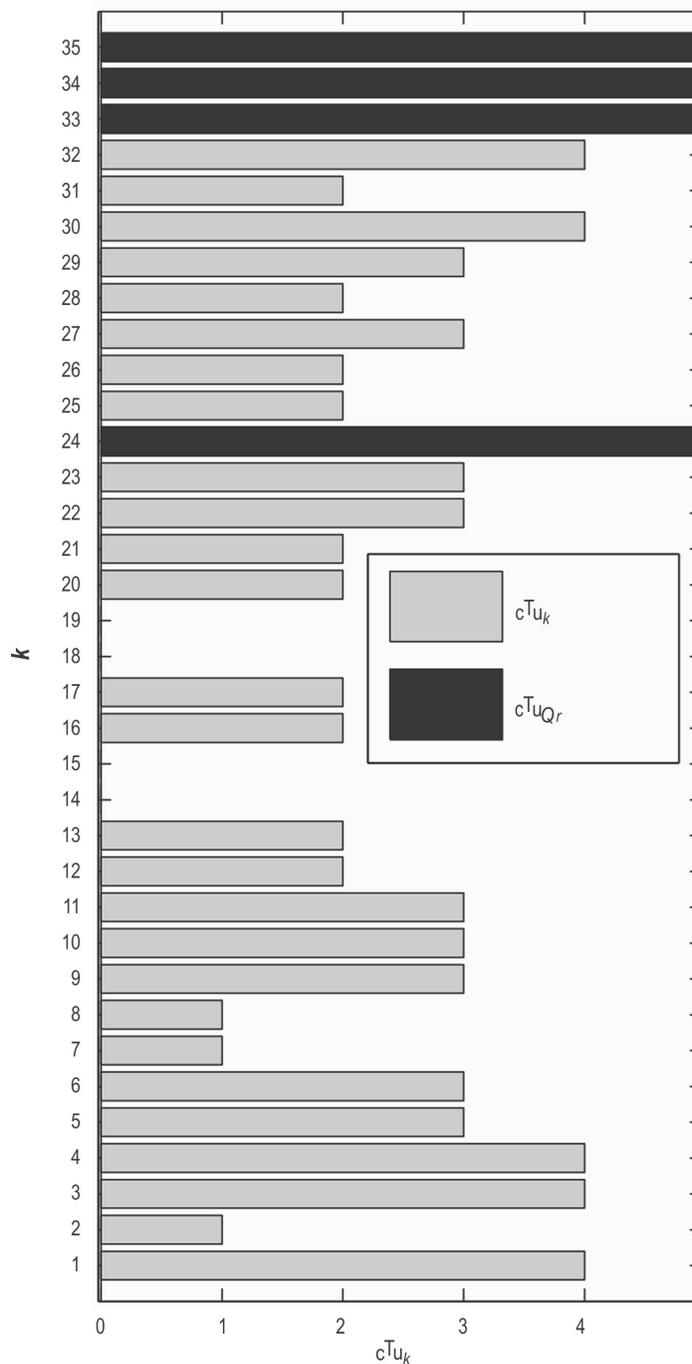


Figura 3: valores de función objetivo para cada punto extremo del problema presentado en la figura 1. En color negro están los puntos extremos u_k que tienen el valor máximo de función objetivo.

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 experimen-

tales usado en este trabajo fue tomado de (Nissen *et al.*, 1997) y se presenta en la tabla 2. 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_2), etanol (eth), glicerol (gly), succinato (succ), acetato (ace) y consumo de glucosa (glc).

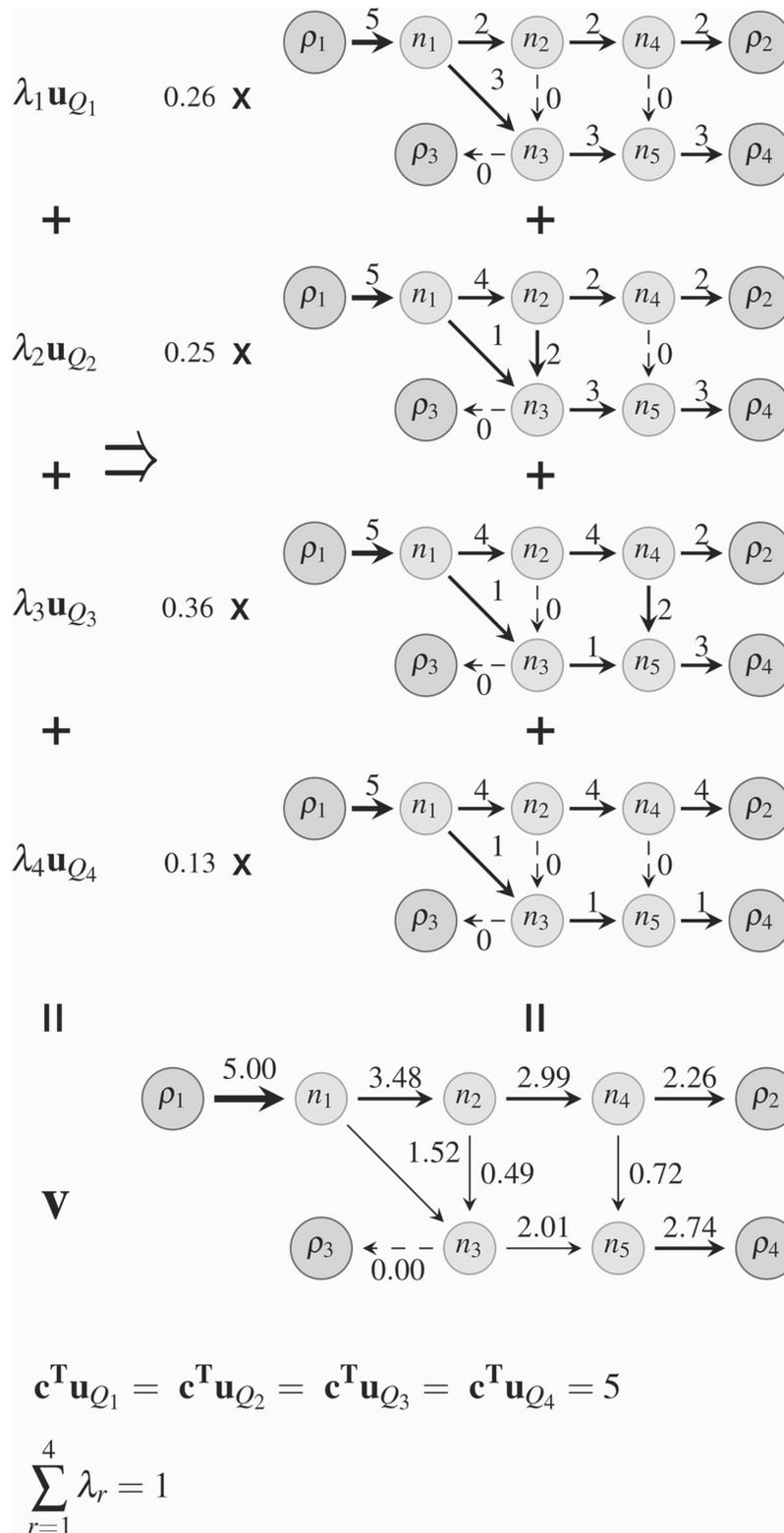


Figura 4: ilustración del conjunto solución de la ecuación (6) del problema de la figura 2. Izquierda: combinación lineal convexa de los puntos extremos que son solución de (6). Derecha: puntos extremos y valores de la combinación lineal convexa. Abajo: solución problema de la figura 2. Las funciones objetivo evaluadas en los puntos extremos mostrados son iguales al máximo valor posible.

Tabla 2: 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 (CO₂), crecimiento (gw)

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

La reconstrucción iMM904 cuenta con 1228 componentes internos y 1577 reacciones. Está disponible en formato SimPheny[®] (Mo et al., 2009).

Se define $v_{\langle \text{reacción} \rangle_j}$ como la velocidad de reacción de nombre $\langle \text{reacción} \rangle$, con j igual al índice de la columna de \mathbf{S} que corresponde a la reacción $\langle \text{reacción} \rangle$. Los valores $\alpha_{\langle \text{reacción} \rangle}$ y $\beta_{\langle \text{reacción} \rangle}$ definen los límites de la reacción $v_{\langle \text{reacción} \rangle}$. El valor $c_{\langle \text{reacción} \rangle}$ define el coeficiente que acompaña a $v_{\langle \text{reacción} \rangle}$ en la función objetivo. La correspondencia entre nombres de reacción en la iMM904 y velocidades $v_{\langle \text{reacción} \rangle}$ usados en las simulaciones del presente trabajo se muestran en la tabla 3. 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[®] (Schellenberger et al., 2011). La reconstrucción iMM904 también cuenta con información sobre las capacidades del sistema representadas por los vectores α y β . Excluyendo algunas reacciones, los límites de v_j son $\alpha_j = -1000 \left(\frac{mmol}{gDWh}\right)$ para reacciones reversibles, $\alpha_j = 0 \left(\frac{mmol}{gDWh}\right)$ para irreversibles y $\beta_j = -1000 \left(\frac{mmol}{gDWh}\right)$. La función objetivo incluida en esta reconstrucción es la maximización del crecimiento del sistema, representa-

da por \mathbf{c} , con $c_{gw} = 1$ y 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 (Feist y Palsson, 2010). El modelo está disponible para MatLab[®] y COBRA^{****}.

Las condiciones de las mediciones experimentales de la tabla 2 requieren modificación de algunas restricciones FBA que establece por defecto la reconstrucción iMM904. La condición anaerobia significa que el medio en el que se encuentra carece de oxígeno. En el FBA esto equivale a establecer $\alpha_{EX_02(e)} \beta_{EX_02(e)} = 1$. 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.

**** http://www.eecsgroup.com/CV_Research_files/cb_iMM904.mat

Tabla 3: correspondencias entre nombres de reacción usados en la reconstrucción iMM904 y las velocidades de reacción $v_{\text{reacción}}$ para este trabajo.

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)	$v_{EX_zymst(e)}$
hexadecanato (n-C16:1)	EX_hdcea(e)	$v_{EX_hdcea(e)}$
octadecanato (n-C18:0)	EX_ocdca(e)	$v_{EX_ocdca(e)}$
octadecanato (n-C18:1)	EX_ocdcea(e)	$v_{EX_ocdcea(e)}$
octadecanato (n-C18:2)	EX_ocdcya(e)	$v_{EX_ocdcya(e)}$

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 4. Esta lista no incluye todas las restricciones, solo aquellas relevantes para el conjunto de datos experimentales de la tabla 2.

Cada una de las cuatro condiciones experimentales expuestas en la tabla 2 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 2 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 2 se realizó un FBA y un FVA. Se determinó el error entre la predicción de las velocidad de reacción v_i y el correspondiente valor experimental v_i^E mediante $|(v_i - v_i^E) / v_i^E| \times 100\%$. 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 (10) y (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 determinar 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® Core i3 y memoria RAM de 4GB DDR3. Los problemas de programación lineal de este trabajo fueron solucionados utilizando el paquete de software libre GLPK (Makhorin, 2011) con interfaz disponible para MatLab® a través del paquete COBRA.

Resultados

Los resultados obtenidos se muestran en la figura 6. Las reacciones con las mejores predicciones median-

Tabla 4. restricciones α_j y β_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.

reacción	iMM904		FBA	
	α_j $\left(\frac{\text{mmol}}{\text{gDWh}}\right)$	β_j $\left(\frac{\text{mmol}}{\text{gDWh}}\right)$	α_j $\left(\frac{\text{mmol}}{\text{gDWh}}\right)$	β_j $\left(\frac{\text{mmol}}{\text{gDWh}}\right)$
$v_{EX_o2(e)}$	-2	1000	0	0
$v_{EX_ergst(e)}$	0	1000	-1000	1000
$v_{EX_zymst(e)}$	0	1000	-1000	1000
$v_{EX_hdcea(e)}$	0	1000	-1000	1000
$v_{EX_ocdca(e)}$	0	1000	-1000	1000
$v_{EX_ocdceat(e)}$	0	1000	-1000	1000
$v_{EX_ocdcya(e)}$	0	1000	-1000	1000
v_{ATPM}	1	1	1	1

te FBA en las cuatro condiciones experimentales son, en orden de exactitud CO_2 , con errores entre 0.3% y 4.5%; etanol, entre 11% y 13%; y crecimiento, entre 11% y 28% (figuras 5, 5 y 5). La predicción del crecimiento es superior al registrado en cada condición experimental de la tabla 2 (figura 5(a)).

La predicción de glicerol en todas las condiciones experimentales es de $0 \left(\frac{\text{mmol}}{\text{gDWh}}\right)$. El valor experimental del glicerol en todas las condiciones comparadas está entre 0 y $6 \left(\frac{\text{mmol}}{\text{gDWh}}\right)$. 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 $2 \left(\frac{\text{mmol}}{\text{gDWh}}\right)$. Los valores experimentales de succinato en todas las condiciones están entre 0.03 y $0.07 \left(\frac{\text{mmol}}{\text{gDWh}}\right)$. 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_2 son mejores que las registradas en (Famili *et al.*, 2003). Las predicciones de etanol tienen una exactitud similar a las registradas en dicho trabajo. Es importante resaltar que (Famili *et al.*, 2003) es el único estudio de *S. cerevisiae* mediante FBA y modelos genómicos que incluye un FVA.

La figura 5 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 1×10^{-5} (Mahadevan y Schilling, 2003). 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 2×10^{-5} y $3.48 \left(\frac{\text{mmol}}{\text{gDWh}}\right)$. 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 5(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 toenan los valores de la condición experimental 2 (figuras 5(d), 5(e) y 5(f)).

La figura 6 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 $36.54 \left(\frac{\text{mmol}}{\text{gDWh}}\right)$. para las reacciones con valor experimental. A diferencia de la figura 5, 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 $v_{EX_gly(e)}$, $v_{EX_succ(e)}$,

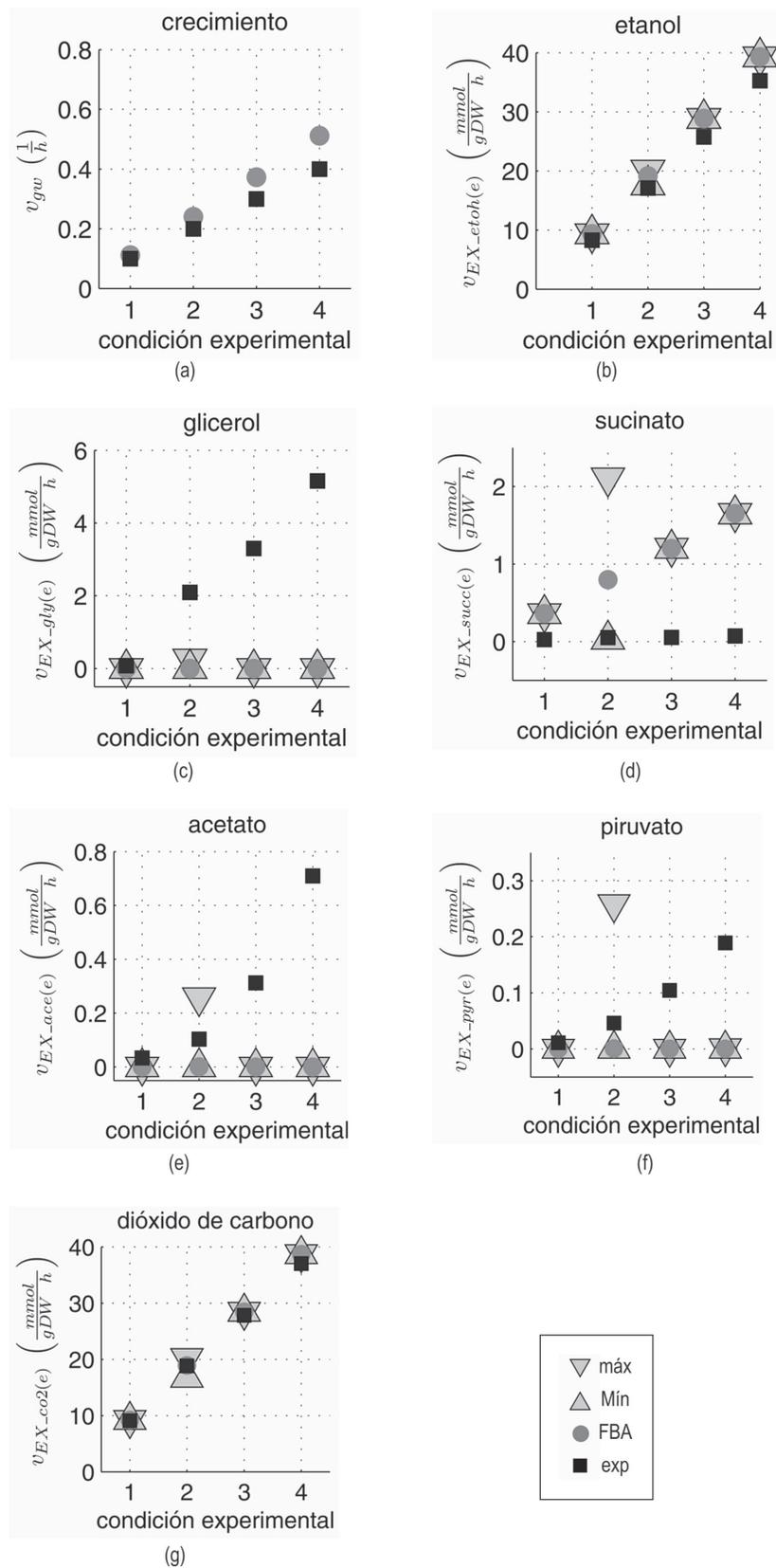


Figura 5: resultados de FBA y FVA bajo las condiciones experimentales de la tabla 3. (a) FBA del crecimiento del sistema. FBA y FVA de (b) etanol, (c) glicerol, (d) succinato, (e) acetato, (f) piruvato y (g) CO₂

$v_{EX_ace(e)}$ 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 comportamiento experimental. Es necesario un estudio de la técnica FBA y la reconstrucción iMM904 utilizando propuestas alternativas de funciones objetivo.

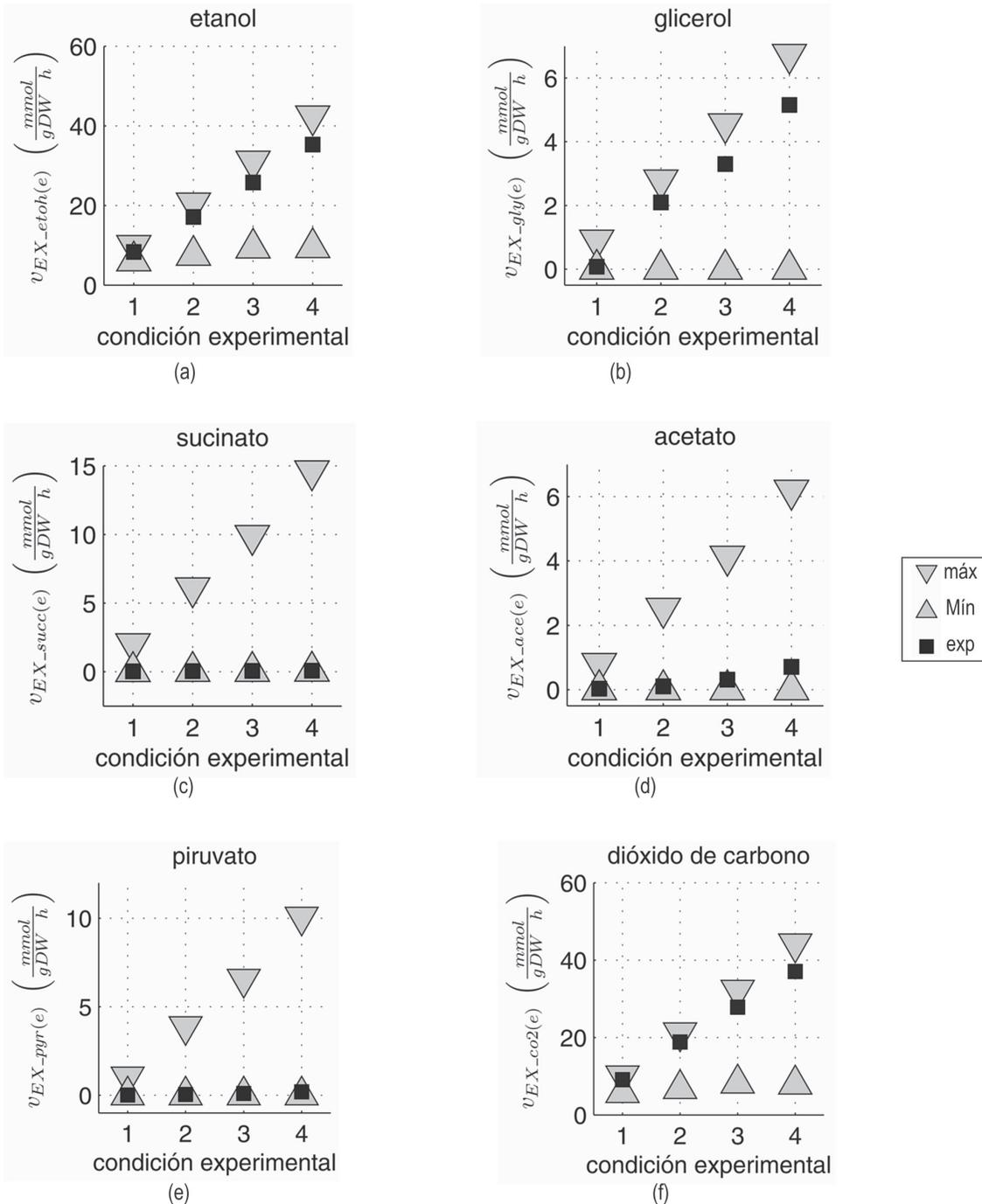


Figura 6: FVA bajo las condiciones experimentales de la tabla 3 y valor de función objetivo igual al valor experimental de crecimiento. FVA de (a) etanol, (b) glicerol, (c) succinato(d) acetato, (e) piruvato y (f) CO₂

Conclusiones

Las reacciones con las mejores predicciones FBA dentro del conjunto de datos de (Nissen *et al.*, 1997) son producción de CO₂ $v_{EX_co2(e)}$, 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 CO₂ con una exactitud superior a la registrada en la literatura, para las condiciones experimentales de (Nissen *et al.*, 1997). 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 (Nissen *et al.*, 1997). 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.

Referencias bibliográficas

- Bai, F. W., Anderson, W. A., & Moo-Young, M. 2008. Ethanol fermentation technologies from sugar and starch feedstocks. *Biotechnology Advances*, 26(1), 89–105.
- Bro, C., Regenber, B., Förster, J., & Nielsen, J. 2006. In silico aided metabolic engineering of *Saccharomyces cerevisiae* for improved bioethanol production. *Metabolic Engineering*, 8(2): 102–111.
- Burgard, A. P., & Maranas, C. D. 2003. Optimization-Based Framework for Inferring and Testing Hypothesized Metabolic Objective Functions. Wiley Periodicals.
- Duarte, N. C., Herrgård, M. J., & Palsson, B. O. 2004. Reconstruction and Validation of *Saccharomyces cerevisiae* iND750, a Fully Compartmentalized Genome-Scale Metabolic Model. *Genome Research*, 14(7): 1298–1309.
- Famili, I., Förster, J., Nielsen, J., & Palsson, B. O. 2003. *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*, 100(23):13134–13139.
- Feist, A. M., & Palsson, B. O. 2010. The biomass objective function. *Current opinion in microbiology*, 13(3): 344–349.
- Förster, J., Famili, I., Fu, P., Palsson, B. O., & Nielsen, J. 2003. Genome-Scale Reconstruction of the *Saccharomyces cerevisiae* Metabolic Network. *Genome Research*, 13(2): 244–253.
- Fukuda, K., & Prodon, A. 1996. Double description method revisited. In M. Deza, R. Euler, & I. Manoussakis (Eds.), *Combinatorics and Computer Science, Lecture Notes in Computer Science* (Vol. 1120, pp. 91–111).
- Gass, S. I. 1985. *Linear programming: methods and applications*. New York, Courier Dover Publications, 36 p.
- Gianchandani, E., Oberhardt, M., Burgard, A., Maranas, C., & Papin, J. 2008. Predicting biological system objectives de novo from internal state measurements. *Bioinformatics*, 9(1): 43.
- Gianchandani, E. P., Chavali, A. K., & Papin, J. A. 2010. The application of flux balance analysis in systems biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 2(3): 372–382.
- Gong, Zhaohua, Liu, C., Feng, E., & Zhang, Q. 2009. Computational method for inferring objective function of glycerol metabolism in *Klebsiella pneumoniae*. *Computational Biology and Chemistry*, 33(1): 1–6.
- Gong, Z., & Yu, Y. 2010. Infer objective function of glycerol metabolism in *Klebsiella pneumoniae* basing on bilevel programming. *Journal of Systems Science and Complexity*, 334–342.
- Herrgård, M. J., Swainston, N., Dobson, P., Dunn, W. B., Arga, K. Y., Arvas, M., Buthgen, N., *et al.* 2008. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat Biotech*, 26(10): 1155–1160.
- Hjersted, J. L., & Henson, M. a. 2009. Steady-state and dynamic flux balance analysis of ethanol production by *Saccharomyces cerevisiae*. *Systems Biology*, 3(3): 167–79.
- Kauffman, K. J., Prakash, P., & Edwards, J. S. 2003. Advances in flux balance analysis. *Current Opinion in Biotechnology*, 14(5): 491–496.
- Kuepfer, L., Sauer, U., & Blank, L. M. 2005. Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Research*, 15(10): 1421–1430.
- Lee, J. M., Gianchandani, E. P., & Papin, J. A. 2006. Flux balance analysis in the era of metabolomics. *Briefings in Bioinformatics*, 7(2): 140–150.
- Lee, S., Phalakornkule, C., Domach, M. M., & Grossmann, I. E. 2000. Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. *Computers and Chemical Engineering*.
- Llaneras, F., & Picó, J. 2008. Stoichiometric modelling of cell metabolism. *Journal of Bioscience and Bioengineering*, 105(1): 1–11.
- Mahadevan, R., & Schilling, C. H. 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering*.
- Makhorin, A. 2011. GLPK (GNU Linear Programming Kit), v. 4.47. Retrieved from <http://www.gnu.org/software/glpk/>.

- Manichaikul, A., Ghamsari, L., Hom, E. F. Y., Lin, C., Murray, R. R., Chang, R. L., Balaji, S., et al. 2009. Metabolic network analysis integrated with transcript verification for sequenced genomes. *Nature Methods*, 6(8): 589–592.
- Mo, M., Palsson, B., & Herrgard, M. 2009. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *Systems Biology / BMC*, 3(1): 37.
- Nissen, T. L., Schulze, U., Nielsen, J., & Villadsen, J. 1997. Flux Distributions in Anaerobic, Glucose-Limited Continuous Cultures of *Saccharomyces Cerevisiae*. *Microbiology*, 143(1): 203–218.
- Nookaew, I., Jewett, M., Meechai, A., Thammarongtham, C., Laoteng, K., Cheevadhanarak, S., Nielsen, J., et al. 2008. The genome-scale metabolic model iN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. *Systems Biology*, 2(1): 71.
- Orth, J. D., Thiele, I., & Palsson, B. U. 2010. What is flux balance analysis? *Nature biotechnology*, 28(3): 245–8.
- Raman, K., & Chandra, N. 2009. Flux balance analysis of biological systems: applications and challenges. *Brief Bioinform*, 10(4): 435–449.
- Raman, K., Rajagopalan, P., & Chandra, N. 2005. Flux Balance Analysis of Mycolic Acid Pathway: Targets for Anti-Tubercular Drugs. *Computational Biology*, 1(5), e46.
- Schellenberger, J., Que, R., Fleming, R. M. T., Thiele, I., Orth, J. D., Feist, A. M., Zielinski, D. C., et al. 2011. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protocols*, 6(9): 1290–1307.
- Schuetz, R., Kuepfer, L., & Sauer, U. 2007. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Molecular Systems Biology*.
- van Berlo, R. J. P., de Ridder, D., Daran, J.-M., Daran-Lapujade, P. a S., Teusink, B., & Reinders, M. J. T. 2011. Predicting metabolic fluxes using gene expression differences as constraints. *transactions on computational biology and bioinformatics*, 8(1): 206–16.
- Zomorodi, A., & Maranas, C. 2010. Improving the iMM904 *S. cerevisiae* metabolic model using essentiality and synthetic lethality data. *Systems Biology*, 4(1): 178.