

# Balance de Flujos Metabólicos en *Saccharomyces cerevisiae* basado en Compartimentalización Intracelular

## Metabolic Flows Balance in *Saccharomyces cerevisiae* based on Intracellular Compartmentalization

César Augusto Vargas García\*, Carlos Eduardo García Sánchez\*\*, Henry Arguello Fuentes\*\*\*, Rodrigo Gonzalo Torres Sáez\*\*\*\*

### Resumen

Una de las técnicas más utilizadas para la predicción de producción de bioproductos y distribución intracelular de flujos de microorganismos es el Análisis de Balance de Flujos - FBA por sus siglas en inglés. El FBA requiere de una función objetivo que represente el objetivo biológico del microorganismo estudiado. En este trabajo se propone un nuevo tipo de funciones objetivo basada en la combinación de objetivos de compartimentos físicos presentes en el microorganismo estudiado. Este tipo de funciones objetivo son examinadas junto con un modelo estequiométrico extraído de la reconstrucción iMM904 del microorganismo *S. cerevisiae*. Su desempeño se compara con la función objetivo más usada en la literatura, la maximización de biomasa, en condiciones experimentales anaeróbicas en cultivos continuos y aeróbicas en cultivos tipo lote. La función objetivo propuesta en este trabajo 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 lote 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.

**Palabras clave:** Análisis de Balance de Flujos, FBA, iMM904, *S. cerevisiae*, Función Objetivo basada en Compartimentos.

### Abstract

Flux Balance Analysis - FBA - is one of the most used techniques in prediction of microorganism bioproducts. It requires an 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 outperforms in 98% growth predictions and 70% ethanol predictions compared with growth maximization.

**Key words:** Flux Balance Analysis, FBA, iMM904 reconstruction, Compartment based Objective Function

**Recibido:** junio 15 de 2013

**Aprobado:** octubre 29 de 2013

\* Msc Ingeniería de Sistemas e Informática, Universidad Industrial de Santander, caugusto.vargas@gmail.com.

\*\* PhD Ingeniería Química, Universidad Industrial de Santander, carlos.garcia6@correo.uis.edu.co.

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

\*\*\*\* Profesor Asociado, Escuela de Química, Facultad de Ciencias, Edificio Camilo Torres, Universidad Industrial de Santander, rtorres@uis.edu.co

## Introducción

En los últimos años el Análisis de Balance de Flujos (FBA) ha permitido estudiar la obtención 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 tiene aplicaciones que van desde la ingeniería metabólica hasta el descubrimiento de medicamentos (Gianchandani *et al.*, 2010).

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 (Raman y Chandra, 2009). El microorganismo *Escherichia coli* cuenta con numerosos estudios sobre la mejor función objetivo que describe su comportamiento bajo diferentes condiciones experimentales, incluyendo anaeróbica y 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 (Feist y Palsson, 2010). En contraste el microorganismo *Saccharomyces cerevisiae* tiene solo un estudio que busca nuevas funciones objetivo y ninguno usando modelos a escala genómica (Feist y Palsson, 2010).

*S. cerevisiae* (conocida como levadura de cerveza) se utiliza para la producción de alimentos, químicos y combustibles, y ha sido pieza clave en el área de la biología eucariota debido a que conserva ciertas propiedades de organismos más complejos como el humano (Österlund *et al.*, 2011). 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 (Österlund *et al.*, 2011).

La reconstrucción a escala genómica iMM904 es hasta ahora la más detallada y confiable en la reproducción de condiciones experimentales de *S. cerevisiae* (Mo *et al.*, 2009). Supera en información genética a sus predecesoras iND750 (Duarte *et al.*, 2004) e iFF708 (Förster *et al.*, 2003).

La búsqueda de la mejor función objetivo del FBA se realiza de dos formas: a través de programación binivel (Terzer *et al.*, 2009) y mediante evaluación sistemática de funciones (Feist y Palsson, 2010). 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 (Gianchandani *et al.*, 2008) debido a que su complejidad es NP-dura<sup>1</sup> (Colson *et al.*, 2007).

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 cuál 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 (Feist y Palsson, 2010) 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.

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.

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 lote es la maximización de producción de biomasa (Gianchandani *et al.*, 2008). Sin embargo, el estudio no se realizó sobre un modelo estequiométrico extraído de una reconstrucción a escala genómica.

El presente trabajo 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 anaeróbi-

<sup>1</sup> Problema de tiempo polinomial no determinístico. Implica que el tiempo que toma solucionar dicho problema incrementa de forma polinomial en relación a su tamaño (número de variables).

cas en cultivos continuos<sup>2</sup> y aeróbicas tipo lote<sup>3</sup> en fase exponencial<sup>4</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 Gianchandani *et al.* (2008) y Feist y Palsson (2010) y usada en el FBA tradicional en condiciones aeróbicas y anaeróbicas.

Este artículo está organizado de la siguiente forma: la sección de métodos explica el modelo estequiométrico y describe la técnica FBA. Además, esta sección presenta la forma de construir funciones objetivos basada en compartimentalización y el lugar que ocupa en la formulación FBA. La sección de resultados muestra el producto de aplicar la técnica FBA al modelo estequiométrico obtenido de la reconstrucción iMM904 utilizando las funciones objetivo propuestas en este trabajo y la función de maximización de biomasa.

## Materiales y Métodos

### 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

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

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.

El conjunto de variaciones de todos los componentes  $n_i$  donde  $i = 1, \dots, M$  se representan por medio del sistema de ecuaciones lineales

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

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  $M < N$ .

### Análisis de Balance de Flujos – FBA

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

$$\frac{d\mathbf{n}}{dt} = \mathbf{S}\mathbf{v} = \mathbf{0}_M \quad (3)$$

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

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

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

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

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

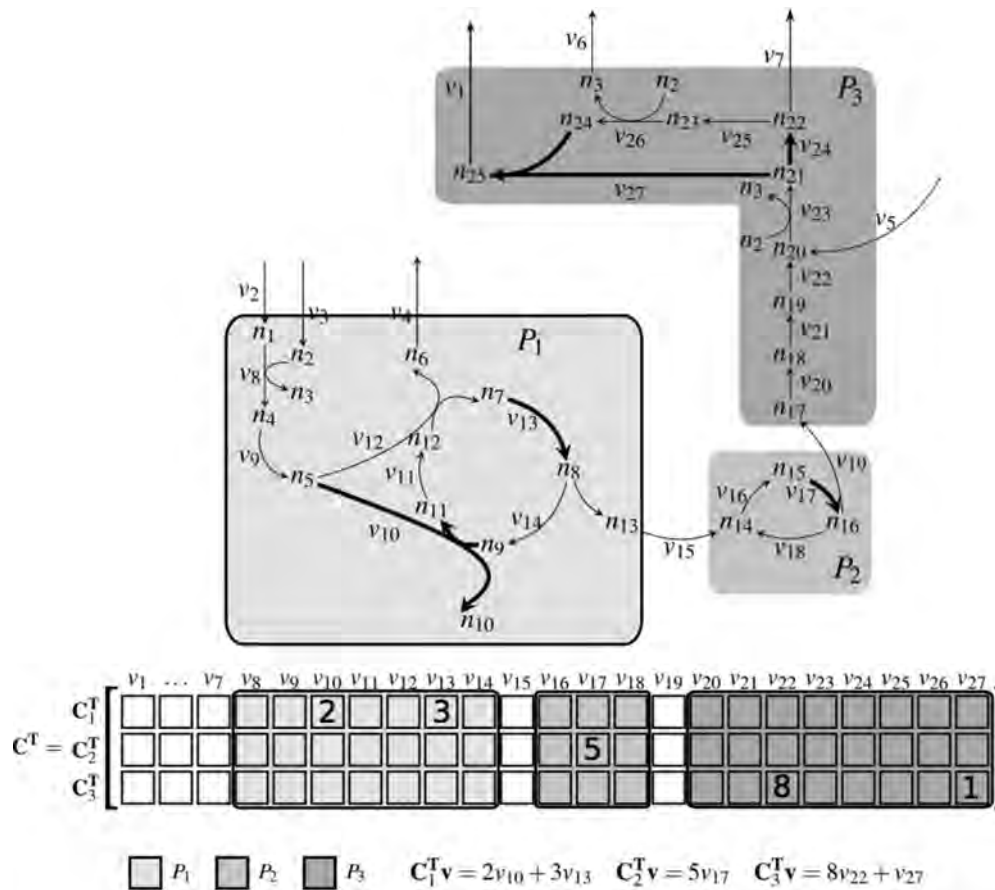
### Funciones Objetivo Basadas en Compartimentos

Sea el compartimento  $P_l$  donde  $l = 1, \dots, L$  es 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

<sup>2</sup> En cultivos continuos el microorganismo estudiado está en estado de equilibrio estacionario. La tasa de crecimiento es constante así como el consumo de nutrientes y la producción de bioproductos.

<sup>3</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>4</sup> En fase exponencial las tasas de crecimiento, consumo de sustratos y formación de bioproductos es constante y está al máximo de su capacidad.



**Figura 1.** Sistema compuesto por 3 compartimentos:  $P_1$ ,  $P_2$  y  $P_3$ . Arriba, diagrama que representa el sistema. Abajo, funciones objetivo  $C_1^T v$ ,  $C_2^T v$  y  $C_3^T v$  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.

del compartimento  $P_1$ . Si  $j \in P_1$  y la velocidad  $v_1$  no está incluida dentro de la función objetivo del compartimento  $P_1$  entonces  $C_{j,l} = 0$ . Por el contrario, si  $j \in P_1$  y  $v_j$  hace parte de la función objetivo del compartimento  $P_1$  entonces  $C_{j,l} \neq 0$ . Por último  $C_{j,l} \neq 0$  si  $j \notin P_1$ .

La figura 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  $v_{16}, \dots, v_{18}$ ,  $P_3$  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 compartimento alguno. La figura 1 también muestra los vectores columna  $C_1$ ,  $C_2$  y  $C_3$  los cuales representan las funciones objetivo de los compartimentos  $P_1$ ,  $P_2$  y  $P_3$ , respectivamente. Es importante 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  $C_1^T v$ .

Dados  $S$ ,  $\alpha$ ,  $\beta$ , la matriz  $C = [C_1, \dots, C_k]$  de dimensiones  $N \times K$  y el vector  $w = [w_1, \dots, w_k]^T$ , el presente trabajo propone buscar el vector que es solución del problema de programación lineal

$$\begin{aligned} \max_v \quad & w^T C^T v \\ \text{s.a.} \quad & Sv = 0_M \\ & v_j \geq \alpha_j \\ & v_j \leq \beta_j \end{aligned} \tag{7}$$

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

### Simulaciones

La reconstrucción iMM904 cuenta con 1228 componentes internos y 1577 reacciones distribuidos en 8



compartimentos, disponible en formato SimPheny™. Cuenta con información sobre las capacidades del sistema representadas por los vectores  $\alpha$  y  $\beta$ . Excluyendo algunas reacciones, los límites de  $v_j$ , siendo PS el peso seco, son  $\alpha_j = -1000 \frac{\text{mmol}}{\text{g PS} \cdot \text{h}}$  para reacciones reversibles,  $\alpha_j = 0 \frac{\text{mmol}}{\text{g PS} \cdot \text{h}}$  para irreversibles y  $\beta_j = 1000 \frac{\text{mmol}}{\text{g PS} \cdot \text{h}}$ . La función objetivo por defecto en esta reconstrucción es la maximización de biomasa o crecimiento del sistema (Feist y Palsson, 2010), representada por  $\mathbf{c}$ , con  $c_{\text{gw}} = 1$  y  $c_j = 0$  para toda reacción diferente a la de crecimiento  $v_{\text{gw}}$ . El modelo estequiométrico  $\mathbf{S}^{1228 \times 1577}$  y los vectores  $\alpha$  y  $\beta$  se extrajeron de la reconstrucción iMM904 utilizando el paquete COBRA (Schellenberger *et al.*, 2011) disponible para MatLab®.

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 propusieron 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  $\mathbf{C}_1$ ,  $\mathbf{C}_2$  y  $\mathbf{C}_3$ , respectivamente.

Para la construcción de las funciones  $\mathbf{C}_1$  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  $\mathbf{C}_2$  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  $\mathbf{C}_3$  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  $P_3$ .

Estas funciones objetivo están relacionados principalmente con la producción y consumo de ATP y poder reductor, debido a que estos cofactores juegan un papel importante en la explicación del comportamiento celular. Otros objetivos considerados están asociados con la producción de biomasa, debido a la relevancia de este objetivo en el modelado FBA. La mayoría de los objetivos comparados en trabajos como Schuetz *et al.* (2007) y Knorr *et al.* (2007) quedan incluidos entre las combinaciones estudiadas en este trabajo. Su efectividad ha sido estudiada en la literatura de forma individual sin tener en cuenta compartimentos físicos del microorganismo estudiado (Feist y Palsson, 2010).

Se propuso un nuevo compartimento virtual representado por  $P_4$  cuyas funciones objetivo, denominadas  $\mathbf{C}_4$  son la maximización y la minimización de biomasa. Este nuevo compartimento permite probar combinaciones de estas dos funciones y otros objetivos propuestos en el citosol. La tabla 1 lista las funciones objetivo propuestas en este trabajo para cada compartimento.

Se construyeron en total 1259 matrices  $\mathbf{C}$  que corresponden a las posibles combinaciones de compartimentos y funciones objetivo de cada uno de estos. Por cada matriz  $\mathbf{C}$  se construyó el vector  $\mathbf{w} = \mathbf{1}_K$ , donde  $\mathbf{1}_K$  es un vector con todos sus elementos iguales a uno y con una longitud  $K$  igual al número de compartimentos incluidos en  $\mathbf{C}$ .

Por cada matriz construida se formularon 12 instancias del problema (7) que representan 12 condiciones experimentales, divididas en 4 condiciones anaeróbicas en quimiostato (Nissen *et al.*, 1997) y 8 aeróbicas tipo lote (Heyland *et al.*, 2009). 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. El crecimiento se presentó por medio de la tasa específica de crecimiento. Esta tasa es igual a la tasa de dilución en cultivos continuos con unidades  $\frac{\text{g biomasa formada}}{\text{g biomasa total} \cdot \text{h}}$ . En cultivos tipo lote, aunque no hay tasa de dilución, se conserva las mismas unidades que en el caso continuo. Las tablas 2 y 3 muestran las condiciones y componentes estudiados.

Cada una de las instancias del problema (7) formuladas para cada una de las matrices  $\mathbf{C}$  se resuelve utilizando el software GLPK (Makhorin, 2011) con interfaz disponible para MatLab® a través del paquete COBRA, mediante un equipo con sistema operativo Linux. En total se resolvieron  $12 \times 1259 = 15108$  instancias del problema (7).

Con el fin de facilitar la notación se define como la velocidad de reacción de nombre  $\langle rX \rangle$  a la cual le corresponde la columna  $j$  de  $\mathbf{S}$ , i.e.  $v_{\langle rX \rangle} = v_j$ . Los valores  $\alpha_{\langle rX \rangle}$  y  $\beta_{\langle rX \rangle}$  definen los límites de la reacción  $v_{\langle rX \rangle}$ . El valor  $c_{\langle rX \rangle}$  define el coeficiente que acompaña a  $v_{\langle rX \rangle}$  en la función objetivo. La correspondencia entre nombres de reacción en la iMM904 y velocidades  $v_{\langle rX \rangle}$  usados en las simulaciones del presente trabajo se muestran en la tabla 4. Las reacciones listadas son externas, salvo  $v_{\text{gw}}$  y  $v_{\text{ATPM}}$ . Por convención, en la reconstrucción iMM904 todas las reacciones externas producen componentes externos.

**Tabla 1.** Posibles funciones objetivo para los compartimentos P1, P2, P3 y P4. ↑ maximización, ↓ minimización, → consumo, ← producción, (trans) transporte de cofactores.

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

**Tabla 2.** Resultados experimentales en condiciones anaeróbicas extraídos de Nissen *et al.* (1997): valores de consumo de glucosa, producción de etanol y crecimiento (tasa de dilución).

Cond #	Glucosa $\frac{mmol}{gPS * h}$	Etanol $\frac{mmol}{gPS * h}$	Crecimiento $\frac{1}{h}$
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

**Tabla 3.** Resultados experimentales en condiciones aeróbicas tipo lote extraídos de Heyland *et al.* (2009): valores de consumo de glucosa, producción de etanol, producción de dióxido de carbono y crecimiento.

Cond #	Glucosa $\frac{mmol}{gPS * h}$	Etanol $\frac{mmol}{gPS * h}$	CO <sub>2</sub> $\frac{mmol}{gPS * h}$	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

**Tabla 4.** Correspondencias entre nombres de reacción usados en la reconstrucción iMM904 y las velocidades de reacción para este trabajo.

Reacción	iMM904	FBA
etanol	EX_etoh(e)	$v_{EX\_etoh(e)}$
glucosa	EX_glc(e)	$v_{EX\_glc(e)}$
dioxido 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)	$v_{EX\_zymst(e)}$
hexadecenoato (n-C16:1)	EX_hdcea(e)	$v_{EX\_hdcea(e)}$
octadecanoato (n-C18:0)	EX\_ocdca(e)	$v_{EX\_ocdca(e)}$
octadecenoato (n-C18:1)	EX\_ocdcea(e)	$v_{EX\_ocdcea(e)}$
octadecanoato (n-C18:2)	EX\_ocdcya(e)	$v_{EX\_ocdcya(e)}$

Las condiciones experimentales anaeróbicas de Nissen *et al.* (1997) requieren modificación de algunas restricciones en el problema (7) que establece por defecto la reconstrucción iMM904. En el problema (7) la condición anaeróbica equivale a establecer  $\alpha_{EX\_o2(e)} = \beta_{EX\_o2(e)} = 0$ . Debido al estado anaeróbico es necesario que el sistema tenga acceso a componentes que en ausencia de oxígeno no se pueden producir. Estos componentes son ergosterol, zymosterol y algunos ácidos grasos. Dicho acceso es permitido relajando las restricciones sobre 6 reacciones de intercambio;  $v_{EX\_ergst(e)}$ ,  $v_{EX\_zymst(e)}$ ,  $v_{EX\_hdcea(e)}$ ,  $v_{EX\_ocdca(e)}$ ,  $v_{EX\_ocdcea(e)}$  y  $v_{EX\_ocdcya(e)}$ ; de modo que pasen de ser irreversibles (sólo producción) a ser reversibles (consumo o producción). Por convención, en la reconstrucción iMM904, las reacciones correspondientes a estos componentes son irreversibles. Convertir en reversibles estas reacciones es equivalente a que el límite inferior sea  $\alpha_{<rx>} = -1000 \frac{mmol}{g\ PS * h}$  en cada una de las velocidades de reacción de estos componentes.

La lista de restricciones por defecto de la iMM904 y las restricciones del problema (7) extraídas de Nissen *et al.* (1997) se presentan en la tabla 5. Esta lista no incluye todas las restricciones, sólo aquellas relevantes para el conjunto de datos experimentales de dicho trabajo.

En las condiciones aeróbicas tipo lote (Heyland *et al.*, 2009) se modificó el límite inferior  $\alpha_{EX\_o2(e)}$  de consumo de oxígeno por defecto de la reconstrucción iMM904 el cual pasó de  $-2 \frac{mmol}{g\ PS * h}$  a  $-1000 \frac{mmol}{g\ PS * h}$ . Se cambió el límite inferior de consumo oxígeno para que los resultados obtenidos no fueran producto de esa restricción

(Hjersted y Henson, 2009) si no del desempeño de las funciones objetivo. Además el consumo de oxígeno puede ser superior a  $2 \frac{mmol}{g\ PS * h}$ , ya que valores de hasta  $10 \frac{mmol}{g\ PS * h}$  son comunes experimentalmente.

El mantenimiento interno de energía del sistema, representado en el modelo por  $v_{ATPM}$  se conserva igual al provisto por la reconstrucción para todas las condiciones experimentales estudiadas. Para todas las condiciones experimentales 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  $e_{gw}$  del error en la predicción de crecimiento  $e_{gw} = \left| \frac{v_{gw} - v_{gw}^E}{v_{gw}^E} \right|$ , con  $v_{gw}$  como el valor de crecimiento predicho y  $v_{gw}^E$  su correspondiente valor experimental. También se determinó el promedio  $\bar{e}_{EX\_etoh(e)}$  del error en la predicción de etanol  $\bar{e}_{EX\_etoh(e)} = \left| \frac{v_{EX\_etoh(e)} - v_{EX\_etoh(e)}^E}{v_{EX\_etoh(e)}^E} \right|$ , con  $v_{EX\_etoh(e)}$  como el valor predicho de etanol y  $v_{EX\_etoh(e)}^E$  su correspondiente valor experimental.

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 (Feist y Palsson, 2010), y los resultados experimentales registrados en Nissen *et al.* (1997)

**Tabla 5.** Restricciones  $\alpha_j$  y  $\beta_j$  presentes en la reconstrucción iMM904 y en el problema (7) de este trabajo para las condiciones experimentales de Nissen *et al.* (1997). En negrilla se resaltan cambios entre los valores de la reconstrucción y los utilizados en la formulación de (7).

Reacción	iMM904		FBA	
	$\alpha_j$	$\beta_j$	$\alpha_j$	$\beta_j$
	$\frac{mmol}{g PS * h}$	$\frac{mmol}{g PS * h}$	$\frac{mmol}{g PS * h}$	$\frac{mmol}{g PS * h}$
$V_{EX\_co2(e)}$	<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
$V_{EX\_hdcea(e)}$	<b>0</b>	1000	<b>-1000</b>	1000
$V_{EX\_ocdcya(e)}$	<b>0</b>	1000	<b>-1000</b>	1000
$V_{ATPM}$	<b>1</b>	1	1	1

y Heyland *et al.* (2009). Estas comparaciones se hicieron en términos del error promedio en la predicción de crecimiento  $\bar{e}_{gw}$  y de etanol  $\bar{e}_{EX\_etoh(e)}$ .

## Resultados

La tabla 6 muestra las diez matrices **C** que producen los promedios de error más bajos en la predicción de crecimiento  $e_{gw}$ , para cada una de las condiciones experimentales de las Tablas 1 y 2.

Las seis primeras matrices **C** usadas en condiciones anaeróbicas tienen promedios iguales  $e_{gw}$  (diferencias menores a  $1 \times 10^{-2}$ ). Las funciones objetivo **C**<sub>1</sub> y **C**<sub>4</sub> son la misma para estas seis matrices. La función objetivo **C**<sub>2</sub> no aparece en alguna de estas. Esto significa que incluir alguna de las funciones del compartimento **C**<sub>2</sub> no mejora la predicción de crecimiento. La diferencia en las primeras seis matrices radica en la función objetivo **C**<sub>3</sub>. Cada una de las alternativas de función objetivo **C**<sub>3</sub> aparece en los seis primeros lugares. También existen una matriz **C** en la que no se incluye alguna función objetivo **C**<sub>3</sub> (Fila 1 de resultados en condiciones anaeróbicas en la tabla 6). Lo anterior implica que el menor promedio de error de predicción de crecimiento se obtiene construyendo **C**<sub>ana</sub> con minimización de producción de NADH como función objetivo **C**<sub>1</sub> y maximización de biomasa como función **C**<sub>4</sub>. **C**<sub>ana</sub> reduce el error en la predicción de crecimiento  $e_{gw}$  en un 10% con respecto a la predicción obtenida con la función objetivo de maximización de biomasa. El menor valor de  $e_{gw}$  encontrado es de 0.19.

En condiciones aeróbicas en cultivos tipo lote se observa que son seis matrices las que obtienen el menor

promedio  $e_{gw}$ , para las cuales el promedio 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  $e_{gw}$ . 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. Por lo tanto la matriz **C**<sub>aer</sub>, compuesta por esta combinación de compartimentos y funciones objetivo, permite obtener la mejor predicción. El mínimo valor  $e_{gw}$  obtenido para este tipo de condiciones experimentales es 0.1.

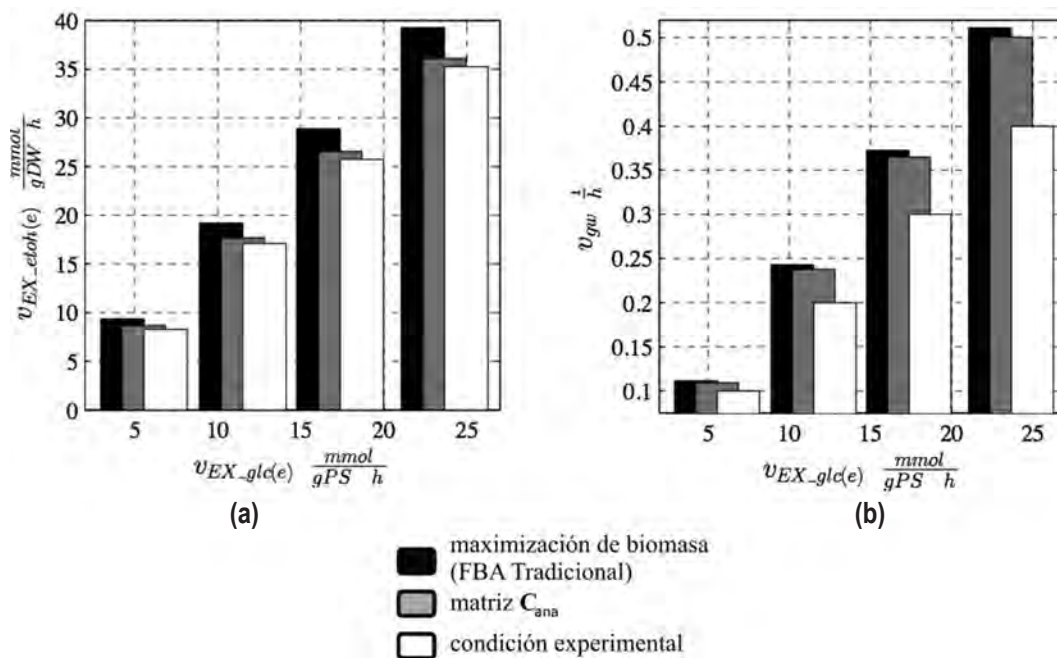
En condiciones anaeróbicas, la función objetivo maximización de biomasa obtuvo un error de predicción promedio de crecimiento  $\bar{e}_{gw}$  de 0.21. La matriz **C**<sub>ana</sub> obtuvo un  $\bar{e}_{gw}$  de 0.19. Esto implica que se logró una reducción del % en el valor de . En cuanto a la producción 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**<sub>ana</sub> obtuvo un  $\bar{e}_{EX\_etoh(e)}$  de 0.03, lo que significa una reducción del 75% en el valor de  $\bar{e}_{EX\_etoh(e)}$ . La figura 2 muestra el desempeño de la función objetivo maximización de producción de biomasa, la matriz **C**<sub>ana</sub> y el comportamiento experimental en condiciones anaeróbicas. Los valores obtenidos de  $e_{gw}$  y  $e_{EX\_etoh(e)}$  para cada condición anaeróbica son mostrados en la tabla 7.

En condiciones aeróbicas tipo lote, la función objetivo maximización de biomasa obtuvo un error de predicción promedio de crecimiento  $\bar{e}_{gw}$  de 4.04. La matriz **C**<sub>ana</sub> obtuvo un  $\bar{e}_{gw}$  de 0.1, logrando una reducción del 98% en el valor de  $\bar{e}_{gw}$ . En cuanto a la producción de



**Tabla 6.** Diez matrices **C** con los promedios de error en la predicción de crecimiento  $\bar{e}_{gw}$  más bajos de todas las combinaciones de compartimentos y funciones objetivo. (-) Función objetivo de compartimento no incluida en **C**.

tipo condición	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>	$\bar{e}_{gw}$
anaerobia	↓←NADH	-	-	↑←biomasa	<b>0.19</b>
	↓←NADH	-	↓←NADH y NADPH	↑←biomasa	<b>0.19</b>
	↓←NADH	-	↓→ATP	↑←biomasa	<b>0.19</b>
	↓←NADH	-	↑←Ácidos Grasos	↑←biomasa	<b>0.19</b>
	↓←NADH	-	↑←NADH y NADPH	↑←biomasa	<b>0.19</b>
	↓←NADH	-	↑→ATP	↑←biomasa	<b>0.19</b>
	-	-	-	↑←biomasa	0.21
	-	-	↓←NADH y NADPH	↑←biomasa	0.21
	-	-	↓→ATP	↑←biomasa	0.21
	-	-	↑←Ácidos Grasos	↑←biomasa	0.21
aerobia (Lote)	↓←NADH	↓→NADH y NADPH	-	↑←biomasa	<b>0.10</b>
	↓←NADH	↓→NADH y NADPH	↑←Ácidos Grasos	↑←biomasa	<b>0.10</b>
	↓←NADH	↓→NADH y NADPH	↑→ATP	↑←biomasa	<b>0.10</b>
	↓←NADH	↓→NADH y NADPH	↓←NADH y NADPH	↑←biomasa	<b>0.10</b>
	↓←NADH	↓→NADH y NADPH	↓→ATP	↑←biomasa	<b>0.10</b>
	↓←NADH	↓→NADH y NADPH	↑←NADH y NADPH	↑←biomasa	<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



**Figura 2.** Resultados obtenidos mediante función objetivo maximización de biomasa, la matriz **C<sub>ana</sub>** y condiciones experimentales anaeróbicas. a) Comportamiento del crecimiento del sistema y b) comportamiento del etanol vs consumo de glucosa.

**Tabla 7.** Resultados obtenidos con la matriz y maximización de biomasa en condiciones anaeróbicas.

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>

etanol, la función objetivo maximización de biomasa obtuvo un error de predicción promedio  $\bar{e}_{EX\_etoh(e)}$  de  $C_{aer}$  obtuvo un  $\bar{e}_{EX\_etoh(e)}$  de 0.30, lo que significa una reducción del 70% en el valor de  $\bar{e}_{EX\_etoh(e)}$ . La figura 3 muestra el desempeño de la función objetivo maximización de producción de biomasa,  $C_{aer}$  y el comportamiento experimental en condiciones aeróbicas tipo lote. Los valores obtenidos de  $e_{gw}$  y  $e_{EX\_etoh(e)}$  para cada condición aeróbica tipo lote son mostrados en la tabla 8.

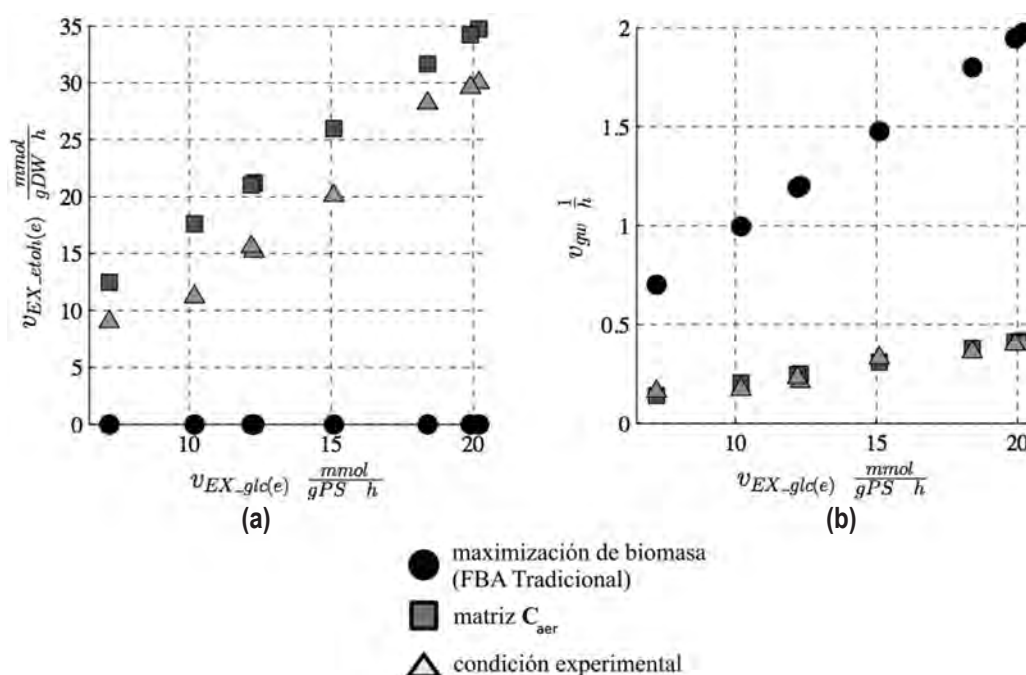
### Conclusiones

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 75%

en las condiciones anaeróbicas de Nissen *et al.* (1997). Esta combinación también reduce en un % 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 lote planteadas por Heyland *et al.* (2009). 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



**Figura 3.** Resultados obtenidos mediante función objetivo maximización de biomasa, la matriz  $C_{ana}$  y condiciones experimentales aeróbicas tipo lote. a) Comportamiento del crecimiento del sistema y b) comportamiento del etanol vs consumo de glucosa.

**Tabla 8.** Resultados obtenidos con la matriz y maximización de biomasa en condiciones aeróbicas.

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>

biomasa mejora las predicciones obtenidas por la función objetivo maximización de biomasa, propuesta en la literatura por Feist y Palsson (2010) y en la reconstrucción a escala genómica iMM904 (Mo et al., 2009), en condiciones experimentales anaeróbicas en cultivo continuo (Nissen et al., 1997) y aeróbicas tipo lote (Heyland et al., 2009).

### Referencias bibliográficas

Colson B., Marcotte P., Savard G. 2007. An overview of bilevel optimization. *Annals of Operations Research*. 153(1): 235-256.

Duarte N.C., Herrgård M.J., Palsson B. Ø. 2004. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized Genome-Scale metabolic model. *Genome Research*. 14 (7):1298-1309.

Feist A.M. y Palsson B.Ø. 2010. The biomass objective function. *Current Opinion in Microbiology*. 13(3):344-349.

Förster J., Famili I., Fu P., Palsson B.Ø., Nielsen J. 2003. Genome-Scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Research*. 13(2):244-253.

Gianchandani E.P., Oberhardt M., Burgard A., Maranas C., Papin J. 2008. Predicting biological system objectives de novo from internal state measurements. *BMC 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.

Heyland J., Fu J., Blank L.M. 2009. Correlation between TCA cycle flux and glucose uptake rate during respiro-fermentative growth of *Saccharomyces cerevisiae*. *Microbiology*. 155(12):3827-3837.

Hjersted J., Henson M. 2009. Steady-state and dynamic flux balance analysis of ethanol production by *Saccharomyces cerevisiae*. *IEE Systems Biology*. 3(3):167-179.

Knorr A.L., Jain R., Srivastava R. 2007. Bayesian-based selection of metabolic objective functions. *Bioinformatics*. 23(3): 351-357.

Makhorin A. 2011. GLPK (GNU Linear Programming Kit). v.4.47. <http://www.gnu.org/software/glpk>

Mo M.L., Palsson B.Ø., Herrgård M. 2009. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Systems Biology*. 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.

Raman, K., Chandra N. 2009. Flux balance analysis of biological systems: applications and challenges. *Briefings in Bioinformatics*. 10(4): 435-449.

Österlund T., Nookaew I., Nielsen J. 2011. Fifteen years of large scale metabolic modeling of yeast: Developments and impacts. *Biotechnology Advances*. 30(5):979-988.

Schellenberger J., Que R., Fleming R.M., Thiele I., Orth J.D., Feist A.M., Zielinski D.C., Bordbar A., Lewis N.E., Rahmadian S., Kang J., Hyduke D.R., Palsson B.Ø. 2011. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature 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*. 3:119. doi:10.1038/msb4100162

Terzer M., Maynard. N.D., Covert M.W., Stelling J. 2009. Genome scale metabolic networks. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*. 1(3):285-297.