

**ANÁLISIS DE BALANCE DE *FLUXES* DE *Saccharomyces cerevisiae*  
MEDIANTE LA UTILIZACIÓN DE NUEVAS FUNCIONES OBJETIVO, COMPUESTAS  
POR COMBINACIONES DE OBJETIVOS DE LOS COMPARTIMENTOS CELULARES**

**CARLOS EDUARDO GARCÍA SÁNCHEZ**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FÍSICOQUÍMICAS  
ESCUELA DE INGENIERÍA QUÍMICA  
DOCTORADO EN INGENIERÍA QUÍMICA  
BUCARAMANGA  
2013**

**ANÁLISIS DE BALANCE DE *FLUXES* DE *Saccharomyces cerevisiae*  
MEDIANTE LA UTILIZACIÓN DE NUEVAS FUNCIONES OBJETIVO, COMPUESTAS  
POR COMBINACIONES DE OBJETIVOS DE LOS COMPARTIMENTOS CELULARES**

**CARLOS EDUARDO GARCÍA SÁNCHEZ**  
Ingeniero Químico

Tesis para optar al  
título de Doctor en Ingeniería Química

Director  
**Ph.D. RODRIGO GONZALO TORRES SAEZ**  
Grupo de Investigación en Bioquímica y Microbiología

**UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FÍSICOQUÍMICAS  
ESCUELA DE INGENIERÍA QUÍMICA  
DOCTORADO EN INGENIERÍA QUÍMICA  
BUCARAMANGA  
2013**

A mis padres y hermanos.  
A Yenny y Clarita.

## **AGRADECIMIENTOS**

A la Universidad Industrial de Santander, por la beca de sostenimiento brindada. Gracias a la existencia de estas becas, y del subsidio en los costos educativos en los programas de posgrado, me fue posible asumir este gran reto.

A los profesores y trabajadores de la Escuela de Ingeniería Química, por el apoyo, la orientación, la ayuda y la formación recibida.

A mi director Rodrigo Torres, por la confianza depositada, la enseñanza dada, y las valiosas discusiones y correcciones a lo largo de todo el camino recorrido en mi estudio doctoral. Fue un honor haberlo tenido como director.

A César Vargas, por aceptar hacer parte de esta investigación.

A mi familia, por todo.

A mis amigos, por estar ahí.

A Dios.

## CONTENIDO

INTRODUCCIÓN	pág. 17
1. MODELADO DE LOS PROCESOS CELULARES	19
1.1. CONSTRUCCIÓN DE MODELOS CELULARES	24
1.1.1. Tipos de modelos celulares	26
1.1.2. Técnicas de medición	33
1.1.2.1. Genómica	34
1.1.2.2. Proteómica	34
1.1.2.3. Metabolómica	35
1.1.2.4. Transcriptómica	36
1.1.2.5. Metagenómica	36
1.1.2.6. Marcación isotópica	37
1.2. MODELADO DE LA RED METABÓLICA	38
1.2.1. Características estructurales de las redes metabólicas	40
1.2.2. Construcción de modelos metabólicos a escala genómica	41
1.2.3. Modelado cinético	43
1.2.3.1. Análisis de Control Metabólico (MCA)	46
1.2.3.2. Modelos de acción de masas generalizados (GMA)	47
1.2.3.3. Sistema-S	48

1.2.3.4. Cinética lin-log	49
1.2.4. Modelado estequiométrico	50
1.2.4.1. Análisis sistémicos y estructurales	57
1.2.4.2. Determinación cuantitativa de distribuciones de <i>fluxes</i>	59
2. MODELOS METABÓLICOS DE <i>Saccharomyces cerevisiae</i>	63
2.1.  MODELOS DEL METABOLISMO CENTRAL	63
2.2.  MODELOS A ESCALA GENÓMICA	65
2.2.1. iFF708	65
2.2.2. iND750	66
2.2.3. iLL672	67
2.2.4. iIN800	67
2.2.5. iMM904	68
2.2.6. Red consensuada v1	68
2.2.7. iAZ900	69
2.2.8. Red consensuada v4	70
3. ANÁLISIS DE BALANCE DE <i>FLUX</i> (FBA)	71
3.1.  FUNCIONES OBJETIVO	73
3.1.1. Comparación de funciones objetivo	74
3.1.2. Generación de funciones objetivo a partir de datos	75
3.2.  VARIACIONES DEL FBA	76
3.2.1. FBA dinámico	76

3.2.2.	Incorporación de información regulatoria	76
3.2.3.	Incorporación de restricciones termodinámicas	77
3.2.4.	Análisis de variabilidad de <i>flux</i> (FVA)	79
3.2.5.	MOMA - <i>Minimization of metabolic adjustment</i>	80
3.2.6.	ROOM - <i>Regulatory on/off minimization</i>	81
4.	MÉTODOS, MODELOS Y DATOS	83
4.1.	HIPÓTESIS DE TRABAJO	83
4.2.	OBJETIVOS	84
4.2.1.	OBJETIVO GENERAL	84
4.2.2.	OBJETIVOS ESPECÍFICOS	84
4.3.	SELECCIÓN DEL MODELO METABÓLICO	85
4.4.	SELECCIÓN DE DATOS EXPERIMENTALES	85
4.5.	EVALUACIÓN DEL ERROR EN LAS PREDICCIONES	86
4.6.	COMPARTIMENTOS Y OBJETIVOS – PRIMERA VERSIÓN	88
4.7.	COMPARTIMENTOS Y OBJETIVOS – VERSIÓN FINAL	91
5.	RESULTADOS Y DISCUSIÓN	94
5.1.	RESULTADOS PRIMERA VERSIÓN	94
5.2.	RESULTADOS VERSIÓN FINAL	97
5.3.	DISCUSIÓN	111
6.	CONCLUSIONES	119
7.	FUTURAS DIRECCIONES	121

BIBLIOGRAFÍA	122
ANEXOS	144

## TABLAS

Tabla 1. Factores que restringen la función metabólica.	pág. 23
Tabla 2. Clasificación de modelos celulares.	27
Tabla 3. Objetivo de las diferentes técnicas “ómicas”.	33
Tabla 4. Ejemplo de matriz estequiométrica (corresponde al modelo del metabolismo central de <i>Saccharomyces cerevisiae</i> de la Figura 4).	52
Tabla 5. Conjuntos de datos experimentales usados para evaluar la precisión de las predicciones de las diferentes funciones objetivo.	86
Tabla 6. Posibles objetivos de los compartimentos celulares evaluados con la primera versión de las combinaciones.	90
Tabla 7. Posibles objetivos de los compartimentos celulares evaluados en la versión final del trabajo.	93
Tabla 8. Resumen de los resultados obtenidos en el presente estudio.	109
Tabla 9. Resumen de estudios de comparación del desempeño de diferentes funciones objetivo en el FBA.	116

## FIGURAS

	pág.
Figura 1. Representación gráfica de una red metabólica.	29
Figura 2. Representación gráfica de una ruta de la red de señales.	30
Figura 3. Representación gráfica de una red regulatoria.	31
Figura 4. Modelo del metabolismo central de <i>Saccharomyces cerevisiae</i> .	51
Figura 5. Aplicaciones del modelado estequiométrico del metabolismo, junto con el modelo matemático básico de cada técnica.	56
Figura 6. Modelo del metabolismo central de <i>S. cerevisiae</i> construido por Nissen <i>et al.</i> , 1997.	64
Figura 7. Mejores funciones objetivo en FBA y sus errores de predicción, para el caso de crecimiento anaeróbico de <i>S. cerevisiae</i> en cultivo continuo, usando las combinaciones de objetivos de la versión inicial.	95
Figura 8. Mejores funciones objetivo en FBA y sus errores de predicción, para el caso de crecimiento aeróbico de <i>S. cerevisiae</i> en cultivo continuo, con <i>flux</i> de intercambio de oxígeno conocido, usando las combinaciones de objetivos de la versión inicial.	96
Figura 9. Mejores funciones objetivo en FBA y sus errores de predicción, para el caso de crecimiento aeróbico de <i>S. cerevisiae</i> en cultivo continuo, sin datos acerca del valor del <i>flux</i> de intercambio de oxígeno, usando las combinaciones de objetivos de la versión inicial.	96
Figura 10. Mejores funciones objetivo en FBA y sus errores de predicción, para el caso de crecimiento anaeróbico de <i>S. cerevisiae</i> en cultivo continuo, en la versión final del trabajo.	100
Figura 11. Mejores funciones objetivo en FBA y sus errores de predicción, para el caso de crecimiento aeróbico de <i>S. cerevisiae</i> en cultivo continuo conociendo el valor del <i>flux</i> de intercambio de oxígeno, en la versión final del trabajo.	101

Figura 12. Mejores funciones objetivo en FBA y sus errores de predicción, para el caso de crecimiento aeróbico de <i>S. cerevisiae</i> en cultivo continuo, sin conocer el valor del <i>flux</i> de intercambio de oxígeno, en la versión final del trabajo.	102
Figura 13. Mejores funciones objetivo en FBA y sus errores de predicción, para el caso de crecimiento aeróbico de <i>S. cerevisiae</i> en fase exponencial de crecimiento en cultivo por lotes, sin conocer el valor del <i>flux</i> de intercambio de oxígeno, en la versión final del trabajo.	103
Figura 14. Mejores funciones objetivo en FBA y sus errores de predicción, para el caso de crecimiento aeróbico de <i>S. cerevisiae</i> en cultivo continuo, con tasas de dilución iguales o menores a $0,15 \text{ h}^{-1}$ , en la versión final del trabajo. El <i>flux</i> de oxígeno es conocido.	104
Figura 15. Mejores funciones objetivo en FBA y sus errores de predicción, para el caso de crecimiento aeróbico de <i>S. cerevisiae</i> en cultivo continuo, con tasas de dilución mayores a $0,15 \text{ h}^{-1}$ y no superiores a $0,28 \text{ h}^{-1}$ , en la versión final del trabajo. El <i>flux</i> de oxígeno es conocido.	105
Figura 16. Mejores funciones objetivo en FBA y sus errores de predicción, para el caso de crecimiento aeróbico de <i>S. cerevisiae</i> en cultivo continuo, con tasas de dilución mayores a $0,28 \text{ h}^{-1}$ , en la versión final del trabajo. El <i>flux</i> de oxígeno es conocido.	106
Figura 17. Mejores funciones objetivo en FBA y sus errores de predicción, para el caso de crecimiento aeróbico de <i>S. cerevisiae</i> en cultivo continuo, con glucosa como sustrato, en la versión final del trabajo. El <i>flux</i> de oxígeno es conocido.	107
Figura 18. Mejores funciones objetivo en FBA y sus errores de predicción, para el caso de crecimiento aeróbico de <i>S. cerevisiae</i> en cultivo continuo, con sustratos diferentes a glucosa, en la versión final del trabajo. El <i>flux</i> de oxígeno es conocido.	108

## ANEXOS

	pág.
Anexo A. Conjuntos de datos experimentales usados en el estudio.	144
Anexo B. Vectores de pesos relativos usados para generar todas las funciones objetivo probadas.	147
Anexo C. Contenido del trabajo <i>Computational Flux Balance Analysis (FBA) of new representative objective functions using a multiple compartmental objective approach and its application to Saccharomyces cerevisiae biological behavior</i> , presentado en el evento <i>ISCB Latin America Conference on Bioinformatics</i> , celebrado en Santiago de Chile en marzo de 2012.	151
Anexo D. Distinción obtenida por el trabajo <i>Computational Flux Balance Analysis (FBA) of new representative objective functions using a multiple compartmental objective approach and its application to Saccharomyces cerevisiae biological behavior</i> .	156
Anexo E. Primera página del artículo <i>Predictive Potential of Flux Balance Analysis of Saccharomyces cerevisiae Using as Optimization Function Combinations of Cell Compartmental Objectives</i> , divulgado en la publicación <i>PLoS One</i> .	157
Anexo F. Primera página del artículo <i>Balance de Flujos Metabólicos en Saccharomyces cerevisiae basado en Compartimentalización Intracelular</i> , aceptado para publicación en la <i>Revista Colombiana de Biotecnología</i>	158
Anexo G. Primera página del artículo <i>Exploration of the effect of input data on the modeling of cellular objective in Flux Balance Analysis (FBA)</i> , aceptado para publicación en <i>Advances in Computational Biology: Proceedings of the Second Colombian Computational Biology Congress</i>	159
Anexo H. Listado de reacciones involucradas en cada uno de los objetivos de los compartimentos celulares propuestos, en el modelo iMM904. Se presenta, para cada objetivo de cada compartimento, el número de la reacción en el modelo y la reacción completa en la que se incluye el(los) metabolito(s) que conforman el objetivo.	160

## RESUMEN

**TÍTULO:** ANÁLISIS DE BALANCE DE *FLUXES* DE *Saccharomyces cerevisiae* MEDIANTE LA UTILIZACIÓN DE NUEVAS FUNCIONES OBJETIVO, COMPUESTAS POR COMBINACIONES DE OBJETIVOS DE LOS COMPARTIMENTOS CELULARES \*

**AUTOR:** CARLOS EDUARDO GARCÍA SÁNCHEZ \*\*

**PALABRAS CLAVE:** FBA, función objetivo, potencial predictivo, *fluxes* metabólicos.

El objetivo principal del análisis de balance de *flux* (*Flux Balance Analysis* – FBA) es obtener predicciones cuantitativas de los *fluxes* metabólicos de un organismo. Para ello, es necesario usar una función objetivo apropiada para garantizar buenas estimaciones de esas velocidades de reacción.

En este estudio, se evaluó el desempeño predictivo del FBA, usando funciones objetivo surgidas de la combinación lineal de diferentes objetivos celulares supuestos. Este enfoque es más apropiado para las células eucarióticas, debido a la multiplicidad de compartimentos celulares que presentan. Por esta razón, se usó en este estudio *Saccharomyces cerevisiae* como organismo modelo, y su red metabólica fue representada usando el modelo metabólico a escala genómica denominado iMM904. Dado que el objetivo fue evaluar el potencial predictivo del FBA utilizando el tipo de funciones objetivo descritas, el consumo de sustrato y oxígeno fueron los únicos valores experimentales que se usaron como datos de entrada para el FBA. Además, se utilizó información experimental acerca del crecimiento celular y de la excreción de metabolitos al ambiente para evaluar la calidad de las predicciones.

La calidad de las predicciones obtenidas con el FBA dependió drásticamente del conocimiento de la tasa de consumo de oxígeno. En la mayoría de clasificaciones estudiadas, las mejores predicciones se obtuvieron con ‘maximización de producción de biomasa’ como función objetivo, así como con algunas combinaciones de objetivos que la incluían. En el caso particular del modelado de crecimiento en fase exponencial en un cultivo por lotes sin conocimiento de la tasa de consumo de oxígeno, la función objetivo ‘maximización de producción de biomasa más minimización de producción de NADH en el citosol más minimización de consumo de NAD(P)H en la mitocondria’ arrojó mejores predicciones de los *fluxes* que las obtenidas usando cualquier otra función objetivo explorada en este trabajo.

---

\* Tesis doctoral.

\*\* Facultad de Ingenierías Físicoquímicas, Escuela de Ingeniería Química. Director de tesis: Rodrigo Gonzalo Torres Saez.

## ABSTRACT

**TITLE:** FLUX BALANCE ANALYSIS OF *Saccharomyces cerevisiae* USING NEW OBJECTIVE FUNCTIONS, ARISING FROM COMBINATION OF CELLULAR COMPARTMENTS OBJECTIVES \*

**AUTHOR:** CARLOS EDUARDO GARCÍA SÁNCHEZ \*\*

**KEYWORDS:** FBA, objective function, predictive potential, metabolic fluxes.

The main objective of flux balance analysis (FBA) is to obtain quantitative predictions of metabolic fluxes of an organism, and it is necessary to use an appropriate objective function to guarantee a good estimation of those fluxes. FBA has become the most used technique to estimate the metabolic reaction rates with determined environmental conditions.

In this study, a predictive performance of FBA was evaluated, using objective functions arising from the linear combination of different cellular objectives. This approach is most suitable for eukaryotic cells, owing to their multiplicity of cellular compartments. For this reason, *Saccharomyces cerevisiae* was used as model organism, and its metabolic network was represented using the genome-scale metabolic model iMM904. As the objective was to evaluate the predictive performance from the FBA using the kind of objective function previously described, substrate uptake and oxygen consumption were the only input data used for the FBA. Experimental information about cellular growth and exchange of metabolites with the environment was used to assess the quality of the predictions.

The quality of the predictions obtained with the FBA depends greatly on the knowledge of the oxygen uptake rate. For the most of studied classifications, the best predictions were obtained with "maximization of growth", and with some combinations that include this objective. However, in the case of exponential growth with unknown oxygen exchange flux, the objective function "maximization of growth, plus minimization of NADH production in cytosol, plus minimization of NAD(P)H consumption in mitochondrion" gave much more accurate estimations of fluxes than the obtained with any other objective function explored in this study.

---

\* Ph.D. thesis.

\*\* Facultad de Ingenierías Físicoquímicas, Escuela de Ingeniería Química. Thesis director: Rodrigo Gonzalo Torres Saez.

## INTRODUCCIÓN

Han pasado varios siglos desde que Robert Hooke usó por primera vez la expresión *cell*, “célula” en español, para describir la unidad básica estructural del corcho. El término eventualmente se convirtió en la palabra para denotar la unidad básica estructural y funcional de cualquier ser vivo. No obstante, se ha avanzado mucho respecto a la determinación de la estructura genética y otros tipos de información de la célula, tras cuatro siglos de actividades investigativa a pequeña escala y la más reciente aparición de técnicas más poderosas, que han permitido leer la codificación genética de los organismos y medir gran cantidad de datos de diversas variables celulares en diferentes condiciones ambientales. A pesar de esto, aún hay un largo camino por recorrer para conseguir el objetivo de lograr un completo entendimiento y modelado de la célula y sus procesos internos (Mori, 2004).

El metabolismo es el proceso celular más estudiado, pues consiste en el modelado de las reacciones a través de las cuales la célula consigue la energía y conforma sus estructuras. El metabolismo es el conjunto de transformaciones químicas catalizadas por enzimas que le permiten a un organismo vivo mantener su vida; en el metabolismo, el organismo toma algunas sustancias de su entorno (sustratos), y a partir de ellas obtiene energía y moléculas pequeñas que ensambla para conformar los diferentes organelos, tejidos, paredes, etc. que necesita para permanecer vivo (Cox y Nelson, 2005).

Durante mucho tiempo, el estudio del metabolismo consistió en la determinación experimental de las reacciones que lleva a cabo una célula. Debido al alcance de cada experimento, que solía limitarse a unas pocas reacciones a la vez, el metabolismo se descubría ruta por ruta. Una ruta metabólica se define como cualquier secuencia de pasos de reacción bioquímicos factibles y observables, que conectan a un conjunto específico de metabolitos de entrada y salida (Koffas *et al.*, 1999).

La aparición relativamente reciente de las técnicas llamadas de alto rendimiento (*high-throughput technologies*), con las cuales se puede obtener gran cantidad de datos celulares en un tiempo corto, ha permitido incrementar sustancialmente el conocimiento sobre el metabolismo y otros procesos celulares. Ello permitió (e hizo necesario) analizar las células desde un punto de vista sistémico, considerando no sólo los elementos individuales sino también la interacción entre ellos (Palsson, 2006).

En este contexto, han surgido métodos que apuntan a la obtención de modelos predictivos del metabolismo, dependientes de las condiciones ambientales o genéticas a las que esté sometida una célula. La técnica de modelado predictivo más importante es el análisis de balance de *fluxes*, o FBA (*flux balance analysis*), en la cual se utiliza un criterio

de optimización para resolver el problema subdeterminado consistente en hallar las velocidades de reacción metabólicas de una célula en estado estacionario (Raman y Chandra, 2009).

Las predicciones obtenidas con el FBA dependen fundamentalmente de la calidad del modelo metabólico, de las restricciones que se incluyan en el problema de optimización, y de la adecuación de la función objetivo para representar el comportamiento de la célula. El presente trabajo es una contribución al análisis de funciones con significado biológico que pueden plantearse como funciones objetivo en un FBA. Se pretende, de esta manera, lograr un avance en el modelado del comportamiento celular, y en su aplicación a la determinación y predicción de las distribuciones de velocidades de reacciones intracelulares (o *fluxes* metabólicos).

En el presente trabajo, se proponen una nueva clase de funciones objetivo, que representan combinaciones de objetivos de diferentes organelos celulares. Muchas de las funciones objetivo propuestas previamente quedan incluidas entre las combinaciones planteadas, lo que resulta útil en la comparación del desempeño de las funciones. Esta comparación se hizo usando datos experimentales existentes obtenidos de la literatura disponible, provenientes de cultivos celulares de *Saccharomyces cerevisiae* bajo diferentes condiciones experimentales. En la medida en que exista mayor confiabilidad y variedad de condiciones experimentales en las mediciones disponibles, el tipo de análisis realizado en este trabajo podrá determinar mejor la representación matemática de los objetivos celulares, así como la dependencia de éstos con las condiciones ambientales y genéticas.

## 1. MODELADO DE LOS PROCESOS CELULARES

Durante la segunda mitad del siglo XX, la biología estuvo fuertemente influenciada por enfoques reduccionistas, con lo cual se generó gran cantidad de información sobre componentes individuales celulares y algunos subsistemas, obteniéndose principalmente sus composiciones químicas y, en ocasiones, también sus funciones biológicas (Palsson, 2006). Sin embargo, la interacción existente entre la gran cantidad de componentes celulares impide que las funciones de la célula puedan ser explicadas simplemente por la caracterización individual de sus componentes (Papin *et al.*, 2003). Por esta razón, se ha incrementado el interés en años recientes por el desarrollo de una visión integral de la célula a través de la biología de sistemas, debido a que esta nueva disciplina integra los enfoques experimental y computacional para explicar y predecir, de manera holística, el comportamiento celular bajo condiciones ambientales determinadas (Borodina y Nielsen, 2005). La disponibilidad de gran cantidad de análisis exhaustivos ha permitido esta nueva era de la biología, reconociendo claramente que los organismos están intrínsecamente compuestos de múltiples sistemas que a través de su interacción moldean y confieren propiedades a los sistemas vivos (Ishii *et al.*, 2004), y entendiendo que comprender las relaciones causales entre genotipo y fenotipo requiere una ampliación de las técnicas tradicionales de biología molecular (Mori, 2004).

Se considera que la biología de sistemas implica una interacción iterativa entre experimentación, desarrollo tecnológico, teoría y modelado computacional, y que es precisamente la presencia del modelado computacional la que la distingue de la más tradicional y reduccionista biología molecular (Kell, 2006). Por lo anterior, herramientas fundamentales para lograr la meta de la biología de sistemas son la manipulación experimental, el modelado matemático (el conocimiento acumulado sobre un organismo debe ser ensamblado en un modelo matemático armonioso) y las simulaciones computacionales (el análisis holístico y la aplicación de los modelos matemáticos, así como su validación, requieren el uso de computadores) (Ishii *et al.*, 2004). En otras palabras, se requiere recurrir a conceptos y técnicas de otras disciplinas, entre ellas física, matemática, análisis numérico, procesos estocásticos, teoría de control e informática (Mori, 2004).

Desde hace unas 4 décadas, el desarrollo de las técnicas de ADN recombinante ha introducido una nueva dimensión a la manipulación de las rutas metabólicas, ofreciendo, por primera vez, la capacidad de construir configuraciones metabólicas específicas (Koffas *et al.*, 1999). Gracias a esto, ha surgido la ingeniería metabólica, que es una disciplina científica y tecnológica que trata sobre la generación de fenotipos deseables, usualmente para la síntesis eficiente de moléculas biológicas (Leonard *et al.*, 2007). La ingeniería

metabólica se puede describir como el campo que combina el análisis sistemático de las redes metabólicas (y de otros tipos de redes, como redes de señales y redes regulatorias) con las técnicas de biología molecular para mejorar las propiedades celulares por medio del diseño racional e implementación de modificaciones genéticas (Koffas *et al.*, 1999).

Entre los problemas que apunta resolver la aplicación de la ingeniería metabólica se encuentran: el incremento del rendimiento de productos obtenidos en algunos bioprocesos, el alto costo o limitada disponibilidad de ciertos sustratos, y el exceso de producción de metabolitos secundarios o subproductos que disminuyen los rendimientos de producto en células (Reed *et al.*, 2003).

La ingeniería metabólica es una aplicación de la biología de sistemas, pues es soportada por métodos que modifican el genoma, miden *fluxes* y analizan el estado de la célula, de modo que en la arquitectura celular pueden determinarse e identificarse blancos efectivos para manipulación genética (Raab *et al.*, 2005). De esta manera, es evidente que el modelado y la simulación son herramientas útiles para el diseño de organismos con características fenotípicas deseadas (Ishii *et al.*, 2005). Además, una gran cantidad de productos comercialmente importantes se derivan de procesos en los que intervienen microorganismos (Gombert y Nielsen, 2000; Brock *et al.*, 2002); éstos han sido usados en la producción de metabolitos primarios y secundarios, proteínas recombinantes, biopolímeros, entre otros (Park *et al.*, 2005). Las aplicaciones no se limitan a procesos industriales, sino que se ha evaluado el uso efectivo de la ingeniería metabólica en aplicaciones biomédicas (Raab *et al.*, 2005).

El modelado numérico del comportamiento de los microorganismos, para aplicaciones industriales, se ha basado tradicionalmente en parámetros macroscópicos como la velocidad específica de crecimiento (Blanch, 1981). Estos parámetros cinéticos se han usado principalmente en el diseño de bio-reactores y la optimización de procesos. En algunos casos, la manipulación de los parámetros de los procesos industriales ha permitido obtener rendimientos de producto superiores al 90%, como es el caso de la producción de etanol por fermentación de azúcar usando *S. cerevisiae* (Bai *et al.*, 2008). Como se ha mencionado previamente, otra alternativa de mejoramiento de los bioprocesos consiste en la generación de alteraciones en los microorganismos, de modo que estos tengan características más adecuadas para la aplicación industrial particular (Doble *et al.*, 2004). Este objetivo puede alcanzarse mediante la manipulación de las vías metabólicas, para así cambiar las propiedades bioquímicas de las células e incrementar la productividad de los microorganismos (Palsson, 2006).

Tanto la biología de sistemas, en su rol de ciencia básica, como la ingeniería metabólica, que es una ciencia aplicada, tendrían un hito histórico si se lograra representar un modelo que cubriera todos los procesos que se llevan a cabo en una célula. Sin embargo, incluso una célula microbiana sencilla puede usar miles de reacciones diferentes durante su ciclo

de vida, lo que implica que construir un modelo celular completo es un reto que exige avances en las tecnologías y las técnicas de modelado existentes (Ishii *et al.*, 2004).

Algunos aspectos importantes para el avance de la integración, almacenamiento y análisis de datos en modelos a gran escala son la estandarización del vocabulario, la asignación correcta de interacciones funcionales, la generación de conceptos biológicos a partir de la integración y visualización de datos, y el desarrollo de software y algoritmos adecuados (Smid *et al.*, 2005).

El metabolismo celular fue el primer proceso celular bioquímico que generó interés de estudiar detalladamente la biología de sistemas, pues las transformaciones que efectúa la célula sobre las sustancias que consume permiten entender la obtención de energía y su funcionamiento como sistema biológico abierto. En este sentido, a la tasa neta de conversión de un metabolito precursor en un producto en una ruta metabólica se le conoce habitualmente dentro de la biología de sistemas como *flux* metabólico (Iyer *et al.*, 2007). En otras palabras, un *flux* metabólico es la velocidad de una reacción bioquímica celular. Los *fluxes* pueden considerarse como los determinantes funcionales de la fisiología celular, debido a que la mayoría de reacciones que afectan a una célula quedan incluidas en esa definición (Leonard *et al.*, 2007). La presentación ordenada de todos los *fluxes* de una célula conduce a la creación de la red metabólica, que se define como una representación abstracta del metabolismo de una célula, en la cual se incluyen los metabolitos, las reacciones catalizadas por enzimas, y la relación entre estos dos elementos (Zhao *et al.*, 2006).

Otro de los procesos celulares fundamentales es la síntesis de proteínas, que básicamente conforman todas las enzimas de la red metabólica que permiten a la célula llevar a cabo todas las reacciones que necesita para obtener energía y sintetizar las macromoléculas biológicas que la conforman. Generalmente, este proceso de síntesis de proteínas se lleva a cabo en dos etapas: primero ocurre la transcripción, que consiste en la creación de un ácido ribonucleico mensajero (mRNA – *messenger ribonucleic acid*) que transcribe la misma información contenida en un gen de ácido desoxirribonucleico (DNA – *deoxyribonucleic acid*) donde está codificada la proteína que se va a producir, y luego la traducción, etapa en la cual el mRNA se enlaza con un ribosoma, y permite traducir el mensaje que codifica para una secuencia de aminoácidos de una proteína específica (Cox y Nelson, 2005).

Los niveles de proteínas en las células dependen de los niveles presentes de mRNA, pero también de la regulación de la traducción (*translational regulation*) y de la degradación selectiva de las proteínas (Beyer *et al.*, 2004; Greenbaum *et al.*, 2003). Muchos de los cambios que atraviesa una célula para adaptarse a las condiciones externas se llevan a cabo desde el nivel genético. Por lo tanto, la determinación de los mRNA presentes en unas condiciones determinadas arroja información útil sobre el comportamiento celular, e

incluso se puede buscar correlaciones entre los mRNA (*transcriptoma*) y las cantidades de proteínas presentes (*proteoma*), o los mRNA y las cantidades de metabolitos (*metaboloma*) presentes en la célula (Kresnowati *et al.*, 2006).

Todos los procesos celulares se crean a partir de la codificación genética, pues las secuencias genómicas especifican los programas de expresión genética que producen las células. Sin embargo, en este proceso está involucrada una regulación de qué genes son expresados y en qué cantidad debe llevarse a cabo la transcripción de los mismos. Esto le permite a la célula adaptarse a unas determinadas condiciones ambientales, e incluso a diferentes etapas del ciclo celular o de desarrollo del organismo (Lee *et al.*, 2002). El mayor reto respecto al entendimiento de las funciones celulares a partir de las secuencias genómicas radica en dilucidar y explicar los principios y mecanismos que gobiernan las estructuras y comportamientos de los sistemas expresados (Stelling, 2004).

Se han obtenido muchos datos a gran escala sobre la abundancia de proteína y ARN mensajero (mRNA – *messenger ribonucleic acid*), el estado de la traducción (*translation*) y de la transcripción de *Saccharomyces cerevisiae*, los cuales han permitido explorar la relación existente entre estos aspectos celulares (Beyer *et al.*, 2004). Algunos estudios han mostrado una débil correlación entre los niveles de mRNA y la abundancia de proteína (ver por ejemplo Greenbaum *et al.*, 2003), pero se ha indicado posteriormente que la correlación mejora significativamente cuando se basa la densidad de ribosomas en la longitud del transcripto, en vez de basarla en la longitud del marco de lectura abierto (ORF – *Open Reading Frame*) (Beyer *et al.*, 2004). Algunas de las causas de que no exista una correlación fuerte directa entre estos parámetros de expresión genómica y proteómica, son el efecto de la existencia de múltiples enzimas y de múltiples genes, y los niveles de expresión de éstos a diferentes condiciones. Existen muchos casos en los cuales un gen codifica múltiples enzimas, o se requieren enzimas codificadas por múltiples genes para formar un complejo enzimático activo, lo que ocasiona que el análisis de la relación entre los datos de expresión genética y el metabolismo no sea tan sencillo (Åkesson *et al.*, 2004).

Entre la información cuya inclusión puede ser útil para el modelado de procesos celulares se encuentran restricciones en los valores de las variables, producto de diferentes factores fisicoquímicos, genéticos y cinéticos que afecta el comportamiento fenotípico de la célula. En la **Tabla 1** se mencionan algunos de los factores más comúnmente incluidos como restricciones en los modelos (Joyce y Palsson, 2007; Qian y Beard, 2005; Palsson, 2000; Terzer *et al.*, 2009).

Muchos de los efectos de estos factores se introducen habitualmente en los modelos, permitiendo determinar qué valores son posibles para las variables sin violar las restricciones que aquellos imponen, o hacer diversos tipos de análisis estructurales y funcionales sobre el sistema, entre otras cosas. La inclusión de estas restricciones en los

modelos es ya común en la representación de las redes metabólicas (Joyce y Palsson, 2007), pero también se han empezado a incluir en modelos de redes metabólicas que incluyen simultáneamente representación de la síntesis de RNA y proteínas (Allen y Palsson, 2003), de algunos procesos de señales celulares (Papin *et al.*, 2005), y de redes regulatorias transcripcionales (Covert *et al.*, 2004; Herrgård *et al.*, 2006).

**Tabla 1.** Factores que restringen la función metabólica. Ajustable: restricciones cuyo valor puede ser manipulado.

<b>Factor</b>	<b>Tipo de restricción</b>
<i>Capacidad</i> Máximo valor de los <i>fluxes</i>	No ajustable. Su valor depende de la estequiometría de la red, y de la cinética del <i>flux</i> .
<i>Conectividad</i> Estequiometría del sistema	En general, no ajustable. Si se tiene en cuenta la regulación transcripcional, puede considerarse ajustable.
<i>Velocidad</i> Acción de masas, cinética enzimática, regulación	Ajustables, como consecuencia de procesos evolutivos.
<i>Termodinámica</i> Entropía, potencial químico, energía de Gibbs, tasa de disipación de calor	Algunas ajustables (dependen de múltiples variables), otras no ajustables.
<i>Ambiente</i> Disponibilidad de sustrato, disponibilidad de aceptores electrónicos, pH externo, etc.	Ajustables.
<i>Regulación genética</i> Expresión genética	Ajustable, pero no independiente.
<i>Otros factores</i> Presión osmótica, electroneutralidad, capacidad de solvente, difusión molecular	No ajustables, restricciones fuertes.

Se puede obtener un número significativo de restricciones y ecuaciones celulares a partir de consideraciones termodinámicas, teniendo en cuenta la fuerza motriz química, la tasa de disipación de calor, la tasa de producción de entropía, entalpía y energía de Gibbs, entre otras, usando el enfoque de termodinámica de los sistemas en no-equilibrio (Qian y Beard, 2005). De esta manera, se ha comenzado a introducir conceptos termodinámicos en los análisis de los sistemas vivos, aunque ha habido inconvenientes relacionados con la escasez de datos termodinámicos sobre las reacciones metabólicas, la falta de modelos rigurosos sobre la química metabólica y la ausencia de bases de datos extensivas (Henry *et al.*, 2007). Debido a la dificultad que representa la inclusión manual de este tipo de restricciones en un modelo a escala genómica, ya se han presentado avances con miras a

sistematizar la anotación de restricciones termodinámicas en modelos metabólicos (Kümmel *et al.*, 2006). Se ha notado que la aplicación de conceptos termodinámicos en sistemas biológicos requiere un enfoque que difiere bastante del usado en otros tipos de sistemas, debido a que un sistema vivo no está aislado, sino que siempre está abierto al intercambio de energía y materia con sus alrededores. Se dice que un sistema vivo, tras suficiente tiempo, no tiende a un equilibrio termodinámico, sino a un estado estacionario en no-equilibrio (*NESS – nonequilibrium steady-state*) (Qian y Beard, 2005).

### 1.1. CONSTRUCCIÓN DE MODELOS CELULARES

En la presente era post-genómica, la investigación biológica desde una perspectiva de biología de sistemas desarrolla dos actividades fundamentales: por un lado, incrementar la cantidad de datos disponibles usando las técnicas “ómicas” (genómica, proteómica, etc.), y por otro lado, analizar a través de técnicas bioinformáticas una gran cantidad de datos y modelos que se generan a partir de aquellos usando sofisticadas herramientas computacionales (Ishii *et al.*, 2004).

Todas las fuentes de información, complementarias entre sí, permiten crear análisis holísticos de los sistemas vivos, en contraste con enfoques tradicionales que se centraban en componentes aislados o subsistemas concretos (Ishii *et al.*, 2004). Está claro que ese cambio de enfoque fue posible gracias a la aparición de técnicas que permitieron no sólo incrementar la precisión de las mediciones, sino también la variedad y simultaneidad de sustancias medidas, así como el desarrollo acelerado de las herramientas informáticas y de algoritmos computacionales. A su vez, la representación matemática de la información biológica facilita gran cantidad de estudios biológicos, incluyendo evaluación del contenido de la red, generación y prueba de hipótesis, análisis de características fenotípicas, e ingeniería metabólica (Thiele y Palsson, 2010).

La construcción de modelos puede efectuarse de dos maneras diferentes. Esto depende del enfoque con que se quiere construir el modelo. Si se considera una construcción holística de las funciones celulares sin presentar muchos detalles de éstas, se habla de un enfoque *top-down*. A su vez, si su base es información experimental biológica sobre los componentes celulares, la cual se conecta hasta obtener el modelo del sistema biológico, se tiene un enfoque *bottom-up* (Stelling, 2004). En el primer caso, se parte desde una descripción de la red bioquímica sin mucho detalle de los componentes individuales, incrementando posteriormente la información biológica de la red. En el segundo caso, se lleva a cabo el análisis de componentes individuales del proceso y su posterior ensamblaje de manera modular, permitiendo la construcción del modelo. Actualmente, la creación de modelos usa elementos complementarios de ambos enfoques de construcción de modelos (Müller *et al.*, 2005), pues se necesitan inevitablemente tanto las técnicas “ómicas” - para obtener gran cantidad de información sobre componentes celulares en un

determinado estado y condición ambiental - como las bioinformáticas, para manejar, analizar e interpretar esa elevada cantidad de información (Ishii *et al.*, 2005). Además, se necesita la fusión de ambas secuencias de modelado para la contextualización de conjuntos de datos de alto rendimiento, recurriendo por lo general a un modelo metabólico como marco de comparación (Oberhardt *et al.*, 2009).

El proceso de construcción de modelos matemáticos de procesos biológicos complejos es inevitablemente iterativo. Por lo general, se parte de un modelo inicial *top-down*, apoyado en el conocimiento existente sobre el sistema u organismo, recurriendo a datos genómicos, bioquímicos y fisiológicos establecidos. Esos modelos iniciales deben posteriormente ser ajustados usando otras fuentes de información y datos experimentales, pues el conocimiento incompleto de las restricciones y la presencia de anotaciones erróneas, por lo general, ocasionan que dichos modelos sólo representen correctamente algunas funciones del organismo (Oberhardt *et al.*, 2009; Palsson, 2000).

En la construcción de modelos, no se puede dejar de lado el pragmatismo; un modelo matemático siempre es una simplificación del fenómeno real, y diferentes modelos pueden plantearse de acuerdo con los objetivos deseados y la información disponible (Gombert y Nielsen, 2000). Producto de la gran cantidad de información que puede obtenerse a partir de las técnicas de alto rendimiento, puede caerse en la tendencia de creer que deben conocerse absolutamente todos los detalles celulares para poder utilizar y aplicar adecuadamente los modelos. Parte del arte de la construcción de modelos consiste en determinar qué es necesario para sintetizar un modelo que sea suficientemente detallado y profundo (modelo estructurado), pero también que sea manejable, aplicable e interpretable (Palsson, 2000).

Cuando se trata con el modelado de conglomerados de células, hay que decidir si cada célula se modela independientemente o no. Cualquier población de células es heterogénea, pues cada individuo presenta pequeñas diferencias en su metabolismo y en la dinámica de su ciclo de vida. Sin embargo, para muchas aplicaciones se pueden crear modelos representativos del comportamiento promedio de la población celular, considerando que ésta se compone de una mezcla homogénea de células idénticas. Con esta suposición, en realidad se está modelando el comportamiento promedio que tienen las células de la población, como si ésta estuviera compuesta por “células promedio” iguales entre sí. Incluso este enfoque de célula promedio se ha aplicado a cultivos compuestos por diferentes microorganismos (cultivos mixtos), representando el metabolismo combinado de las distintas especies con un solo modelo metabólico, y aplicando exitosamente este modelo para estimar la producción de diversos metabolitos secundarios en el fermentador (Rodríguez *et al.*, 2005).

Hay otras situaciones en que la heterogeneidad de las células tiene impacto significativo en la dinámica de la población y en la producción de metabolitos. Algunos de los enfoques

que tienen en cuenta la diferencia entre las células individuales son: los modelos de ecuación de balance de la población (PBE – *population balance equation*), los modelos de ensambles de células, los modelos probabilísticos de mezcla, y las cadenas de Markov de estado finito (Henson, 2003; Yin *et al.*, 2001). En el modelo PBE, las células individuales se diferencian usando variables que caractericen su estado intracelular, y se genera un balance dinámico que incluye descripciones de crecimiento celular y de división. El resultado es un conjunto de ecuaciones diferenciales parciales no lineales, cuya cantidad y complejidad dependen de la cantidad de variables para caracterizar el estado interno que se hayan seleccionado (para simplificar los cálculos, puede seleccionarse una sola, como edad celular o masa celular), así como de la cantidad de sustratos límites que se elijan (Henson, 2003).

El modelo de ensamble de células se genera a partir de un modelo dinámico de una sola célula, que es un sistema de ecuaciones diferenciales ordinarias no lineales. Posteriormente, se construye un ensamble de un número grande de células individuales, en las cuales los estados intracelulares son perturbados aleatoriamente de sus valores nominales para representar la heterogeneidad asociada con procesos celulares particulares. Generalmente, se asume que las perturbaciones tienen distribución normal con media cero y varianza consistente con datos disponibles. El modelo se termina al añadir ecuaciones diferenciales que representen la dinámica de la concentración de especies extracelulares relevantes (Henson, 2003).

Se han aplicado también algunos enfoques estadísticos al análisis de modelos celulares, tales como los modelos probabilísticos de mezcla. Estos son modelos para representar la presencia de sub-poblaciones dentro de una población global, sin necesidad de que un conjunto de datos experimentales identifique la sub-población a la cual pertenece una observación individual. De este modo, la técnica permite hallar la configuración más probable para la red, así como un amplio rango de características estructurales, entre ellas la conectividad (Newman y Leicht, 2007). Otro enfoque estadístico que se ha usado para modelar poblaciones celulares como conjuntos de individuos heterogéneos es la representación por cadenas de Markov, de modo que se simula estadísticamente la secuencia de estados que atraviesa un sistema (Yin *et al.*, 2001).

### **1.1.1. Tipos de modelos celulares**

Los modelos de los microorganismos pueden clasificarse en estructurados y no estructurados, de acuerdo con si toman en cuenta las reacciones intracelulares o no. Otra clasificación consiste en llamar al modelo sistema constante agrupado, si considera la masa celular como un sistema homogéneo, o sistema constante distribuido, si la considera en términos de tamaño y morfología (Ishii *et al.*, 2004). Una tercera clasificación importante depende de si los procesos intracelulares se consideran en estado estable o

no. En la **Tabla 2**, se presentan algunas clasificaciones importantes de los modelos celulares y sus características principales.

Los modelos no estructurados son modelos fenomenológicos que describen la evolución dinámica de unos cuantos componentes macroscópicos (Haag *et al.*, 2005). Este tipo de modelos ha resultado muy útil para representar el comportamiento de poblaciones celulares en ingeniería de procesos, en modelado de lodos activados y de bio-reactores, entre otros (Ishii *et al.*, 2005). Como los modelos no estructurados involucran sólo unos pocos componentes macroscópicos, tienen pocos parámetros y propiedades más fácilmente identificables, lo cual es útil para optimización y control de procesos (Haag *et al.*, 2005). La complejidad de este tipo de modelos varía de acuerdo con las condiciones del medio, el número de sustratos limitantes de crecimiento, la consideración de invariabilidad o no de las constantes del modelo, entre otras. Curiosamente, para el caso más simple, el de crecimiento celular de un cultivo puro con un solo sustrato limitante bajo condiciones ideales, el modelo propuesto originalmente por Monod o alguna ligera variación del mismo constituye una excelente representación (Kovárová-Kovar y Egli, 1998).

**Tabla 2.** Clasificación de modelos celulares.

<b>Tipo de modelo</b>	<b>Características</b>	<b>Ventajas</b>	<b>Desventajas</b>
<b>No estructurado</b>	No incluye reacciones intracelulares.	Fácil de construir. Bien establecido.	No puede reproducir el complejo comportamiento celular.
<b>Estructurado</b>	Incluye las reacciones intracelulares.	Puede expresar sistemas complejos.	Requiere información detallada y precisa.
<b>Sistema constante agrupado</b>	Considera la célula como un sistema homogéneo.	Fácil de construir, bien establecido.	Expresa sólo estados promedio.
<b>Sistema constante distribuido</b>	Considera la existencia de distribuciones de parámetros en la célula.	Buena aproximación para la verdadera masa celular.	Subjetividad en las reglas de división y la medición de parámetros.
<b>Dinámico (cinético)</b>	Las variables cambian con el tiempo.	Pueden calcularse estados transitorios.	Se requieren parámetros cinéticos y valores iniciales. Alta complejidad del modelo.
<b>Estequiométrico (estático)</b>	La célula se modela en estado estable. Las variables no dependen del tiempo.	Puede incorporar gran cantidad de reacciones. Permite análisis sistémicos sobre las redes.	No puede representar estados transitorios.

Fuente: Ishii *et al.*, 2004 (adaptada).

Cuando la intención es investigar la fisiología de las células, o mejorar la producción de sustancias por medio de ingeniería metabólica (Leonard *et al.*, 2007), se requiere el uso de modelos de tipo estructurado, que incluyan o se enfoquen en los mecanismos intracelulares (Ishii *et al.*, 2005).

En las ciencias biológicas, se pueden representar como redes, sistemas tan dispares como cadenas de alimentación en ecología e interacciones bioquímicas en biología molecular. En particular, la complejidad de las interacciones entre las moléculas intracelulares ocasiona que el concepto de redes describa muy bien sus características más sobresalientes (Almaas *et al.*, 2005). Los procesos intracelulares se modelan principalmente usando tres tipos de redes bioquímicas: la red metabólica, la red de señales (reacciones intracelulares que ocasionan respuestas a cambios en el ambiente extracelular) y la red regulatoria (conjunto de efectos regulatorios sobre la transcripción genética) (Min Lee *et al.*, 2008).

La red metabólica es un modelo de los procesos celulares de transformación de sustancias; en ella se incluyen las reacciones bioquímicas celulares, así como el transporte de metabolitos hacia y desde el interior de la célula y entre compartimentos celulares. En ocasiones, se introduce también información acerca de la cinética de las reacciones celulares, aunque este no es un requisito para que el modelo sea catalogado como red metabólica.

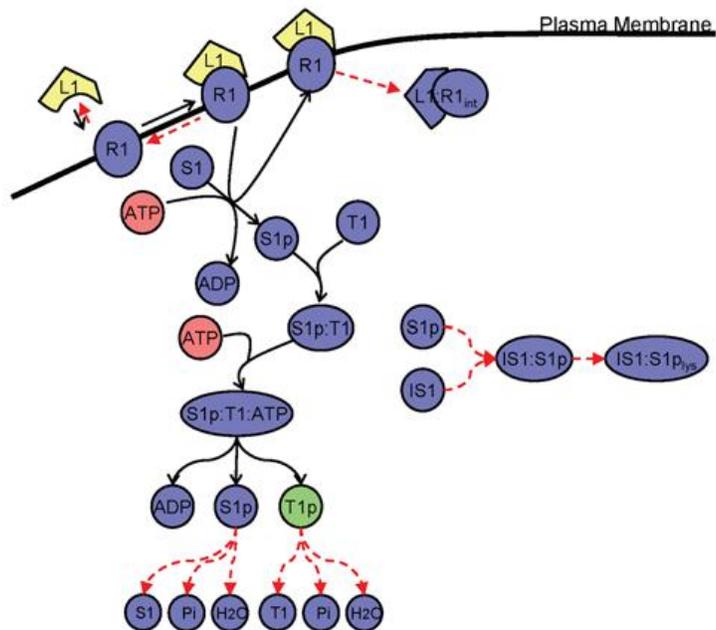
En la **Figura 1** se muestra un ejemplo de representación gráfica de la red metabólica, que corresponde a un modelo del metabolismo central de *S. cerevisiae* (Velagapudi *et al.*, 2007). Estas representaciones son útiles para redes simplificadas, como esquemas generales y para rutas particulares, pero por la cantidad de metabolitos y reacciones que incluyen los modelos a escala genómica, a ese nivel ya no resulta útil este tipo de figuras. En un gráfico de la red metabólica, los metabolitos se representan como nodos de la estructura (las cajas de texto en la figura 1), mientras que las reacciones son mostradas como flechas que unen los diferentes metabolitos, indicando de esta manera la conversión química que lleva a cabo la reacción. Muchas veces en las representaciones gráficas se omiten cofactores como NADH/NAD<sup>+</sup> y NADPH/NADP<sup>+</sup>, porque son metabolitos que participan en muchas reacciones (tienen elevada conectividad) y por consiguiente harían confuso el gráfico. Los metabolitos rotulados *Bio1*, *Bio2*, etc. del ejemplo, representan el uso de metabolitos para crecimiento celular y mantenimiento de sus estructuras, de modo que en conjunto las reacciones que llegan hasta esas cajas conforman la llamada reacción de biomasa.

Las redes metabólicas representan muchos años de acumulación de datos experimentales, y pueden ser confrontadas *in silico* para determinar sus estados funcionales y características (Price y Shmulevich, 2007). Su reconstrucción es un procedimiento ya bien establecido (Duarte *et al.*, 2007; Duarte *et al.*, 2004; Thiele y Palsson, 2010), mientras que



ser usadas por las células para regular sus programas globales de expresión genética. El elemento básico de la red regulatoria es el *motif*, que consiste en la representación de una capacidad regulatoria específica, como por ejemplo, asociación y disociación de complejos proteicos, o activación e inactivación de enzimas o procesos de transporte entre los límites de los compartimentos celulares, o un lazo de retroalimentación positivo o negativo. La descripción de una red regulatoria reside básicamente en la clasificación de los efectos regulatorios observados entre los posibles *motifs*, y el posterior ensamble de los *motifs* que forman la red regulatoria (Lee *et al.*, 2002). De esta manera, la conectividad de las redes regulatorias es inferior a la conectividad de las redes metabólicas (Müller *et al.*, 2005). No obstante, se han trasladado a las redes regulatorias algunas de las técnicas basadas en el análisis convexo usadas en las redes metabólicas, como por ejemplo, la determinación de rutas extremas (Xiong *et al.*, 2004).

**Figura 2.** Representación gráfica de una ruta de la red de señales. Los elementos en azul son componentes cuya principal función es transmitir señales, los elementos rojos son componentes con función en el metabolismo y los verdes son componentes regulatorios. Las líneas sólidas representan *fluxes* diferentes de cero, mientras que las líneas discontinuas rojas representan *fluxes* cuyo valor se hace cero al maximizar la producción del factor activado de transcripción T1p.

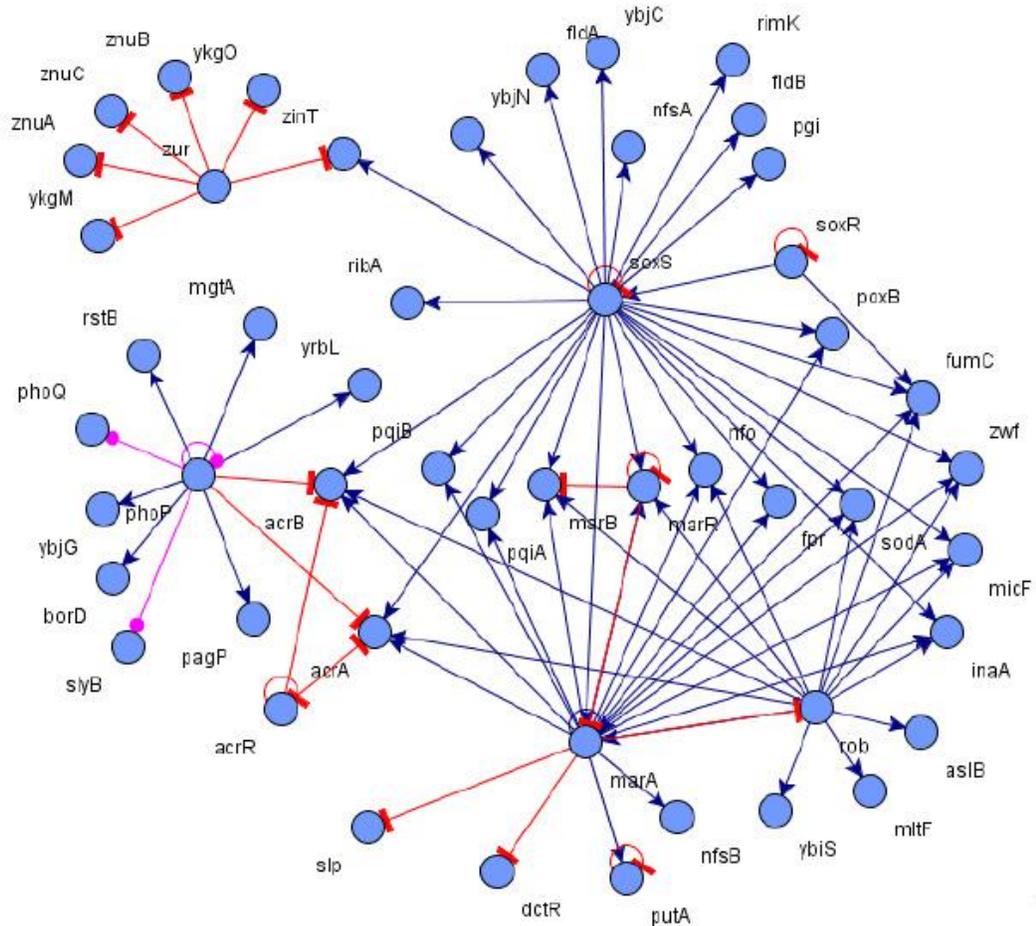


Fuente: Min Lee *et al.*, 2008.

En la **Figura 3** se presenta un ejemplo de una sub-red de interacciones entre genes de *Escherichia coli*, generada con la herramienta *GeneNetWeaver* (Schaffter *et al.*, 2011). En la figura, los nodos representan genes regulatorios, y los colores indican si el gen es

regulado o no por otros genes. Por otra parte, las flechas presentan las relaciones regulatorias presentes (Cerulo *et al.*, 2010).

**Figura 3.** Representación gráfica de una red regulatoria. Los nodos representan diferentes genes, y las flechas diferentes tipos de interacciones entre ellos.



Fuente: Cerulo *et al.*, 2010.

Una característica funcional que diferencia las redes metabólicas y las de señales, o regulatorias, es que los compuestos de las redes de señales y regulatorias pueden tener doble papel: un compuesto que hace parte de una de estas dos redes, puede también tener efecto como reactante o catalizador en la red metabólica (Müller *et al.*, 2005).

Un enfoque alternativo que se ha utilizado para generar redes metabólicas, regulatorias y de señales, consiste en la aplicación de técnicas estadísticas. La creciente cantidad y fidelidad de los datos obtenidos permiten ejecutar inferencias estadísticas al nivel de sistemas, tanto sobre la estructura como la dinámica de mecanismos de control

biomoleculares. La idea básica tras este enfoque es que un modelo que puede capturar de manera fidedigna este tipo de relaciones, tendrá capacidad predictiva y puede usarse también para estudiar propiedades del sistema global (Price y Shmulevich, 2007).

Dos de las técnicas estadísticas usadas para representar redes regulatorias y genéticas, e integrarlas con redes metabólicas, son las redes booleanas y las redes bayesianas. Las primeras se han aplicado en la representación de diversos efectos regulatorios transcripcionales (por ejemplo Albert y Othmer, 2003; Li *et al.*, 2004). Las redes booleanas son inherentemente determinísticas, lo cual representa una desventaja para el modelado de redes celulares, debido a la presencia de aleatoriedad en la expresión genética y la incertidumbre en las mediciones; para superar estos inconvenientes, se ha propuesto el uso de redes bayesianas (Price y Shmulevich, 2007).

Las redes bayesianas son modelos gráficos que permiten representar dependencias condicionales e independencias entre las variables correspondientes a mediciones de expresión genética (Friedman, 2004). Una limitación de las redes bayesianas para modelar redes genéticas es que requieren estar compuestas de grafos acíclicos y, por esta razón, no pueden representar mecanismo de control por retroalimentación; sin embargo, se han propuesto algunas alteraciones para superar ese inconveniente (Price y Shmulevich, 2007).

Se han llevado a cabo múltiples esfuerzos respecto al modelado simultáneo de múltiples redes celulares para algunos organismos. La integración de redes metabólicas y regulatorias ha servido para identificar mecanismos regulatorios en *S. cerevisiae* y *E. coli*; la integración en este caso se llevó a cabo usando modelado estequiométrico para la red metabólica y un formalismo booleano para representar las reglas regulatorias (Covert y Palsson, 2002; Covert *et al.*, 2001; Herrgård *et al.*, 2006). Se han utilizado también datos transcriptómicos para determinar qué genes son expresados en una situación particular, y de acuerdo con esa información, modificar la matriz estequiométrica del organismo (Åkesson *et al.*, 2004). El modelado simultáneo de las redes metabólica y regulatoria, más información sobre la transducción de señales (activación de receptores superficiales por parte de moléculas extracelulares), representadas en un modelo cinético, ha permitido estudiar la respuesta de *S. cerevisiae* al shock osmótico (Klipp *et al.*, 2005). Otro enfoque desarrollado simula en estado pseudo-estacionario los eventos 'rápidos' de la red metabólica, de la red de señales y de reacciones regulatorias, mientras que otros efectos 'lentos' se incorporan con un retraso en el tiempo, modificando las redes que se modelaron en estado pseudo-estacionario; de esta manera, se logró integrar esas tres redes sin recurrir al modelado cinético, a pesar de la gran limitación que representa la diferencia en escalas de tiempo entre los diferentes efectos incluidos en las tres redes (Min Lee *et al.*, 2008).

Las redes celulares, a medida que incorporan más elementos y pueden representar más fielmente el comportamiento celular, incrementan su complejidad, así que para estudiar sus características topológicas, las relaciones entre estructura y función, y las reglas de formación y reestructuración, se debe recurrir a diversos métodos desarrollados en teoría de grafos, estadística, física estadística, simulación computacional, optimización, etc. (Zhao *et al.*, 2006).

### 1.1.2. Técnicas de medición

A las técnicas que permiten obtener gran cantidad de mediciones sobre un tipo de estructura o componente celular, o identificar completamente la secuencia de alguna sustancia intracelular, se les ha denominado con la expresión “ómica” (*omics*), denotando que los análisis abarcan la totalidad de los componentes biológicos a los que apunta el análisis (Park *et al.*, 2008).

Gracias a las tecnologías “ómicas” se ha obtenido gran cantidad de información sobre los sistemas y sus interacciones. Por otra parte, se requiere del uso de modelos matemáticos para el análisis y la integración de datos. Por esta razón, la biología de sistemas tiene una naturaleza iterativa, integrando continuamente datos experimentales y nuevas características en los modelos (Borodina y Nielsen, 2005).

Las técnicas “ómicas” son las principales fuentes de información para el modelado de procesos celulares, su validación y su aplicación; en la **Tabla 3** se presentan los objetivos de las técnicas “ómicas”.

**Tabla 3.** Objetivo de las diferentes técnicas “ómicas”.

Técnica	Objetivo
Genómica	Determinación de las secuencias de ADN y ARN en las células.
Proteómica	Determinación de las diferentes proteínas presentes en la célula bajo una situación determinada.
Metabolómica	Medición de las cantidades de los metabolitos presentes en la célula en condiciones determinadas.
Transcriptómica	Medición y análisis de los niveles de expresión de los ARN mensajeros.
Metagenómica	Determinación de las secuencias de ADN presentes en un conglomerado celular compuesto por diferentes organismos, en un ambiente y condiciones determinadas.

A su vez, la integración de las diferentes técnicas “ómicas” contribuye a profundizar el conocimiento sobre un organismo (Park *et al.*, 2005). Adicionalmente a las técnicas

“ómicas” mencionadas, algunos autores también se refieren con el término “fluxómica” a la determinación de todos los *fluxes* metabólicos de las reacciones que ocurren dentro de una célula, y presentan como ejemplos de este tipo de análisis ala técnica de ingeniería metabólica FBA (ver capítulo 3) y a los experimentos de marcación isotópica (Park *et al.*, 2008).

### **1.1.2.1 Genómica**

Genómica engloba los esfuerzos dirigidos a la determinación de las secuencias completas en las que está codificada la información genética de los organismos (ADN para la mayoría de ellos; ARN para algunos virus), así como el análisis de diversas interacciones que se presentan entre los diferentes tipos de ADN y ARN en muchos organismos. El desarrollo de las técnicas genómicas ha conducido a una explosión de conjuntos de datos a escala genómica de cientos de organismos (Joyce y Palsson, 2007).

La generación de datos de expresión genética para múltiples organismos y condiciones experimentales permite obtener conjuntos de datos que pueden ser analizados usando mapas metabólicos (Borodina y Nielsen, 2005). Se ha utilizado también para dilucidar mecanismos de regulación alostérica por medio de la evaluación de la inhibición presentada tanto en células normales de *S. cerevisiae* como en mutantes carentes de sensibilidad por el inhibidor (Serre *et al.*, 2004).

### **1.1.2.2 Proteómica**

La proteómica busca examinar cuantitativamente la expresión global de proteínas en las células, tejido u organismo, por medio de la separación, identificación y cuantificación sistemática de las proteínas (Park *et al.*, 2005). Gran parte de las mediciones usadas en proteómica se basan en diferentes métodos de espectrometría de masas (MS), combinados con electroforesis en dos dimensiones (2D-SDS-PAGE) o cromatografía líquida de alto rendimiento (HPLC) (Ishii *et al.*, 2004).

Los datos proteómicos se han usado para monitorear cambios metabólicos y celulares en microorganismos durante la producción de diversos metabolitos como proteínas recombinantes y biopolímeros (Pham *et al.*, 2006). Una de las cuestiones importantes en la proteómica consiste en determinar si los niveles de expresión proteica se correlacionan con los valores exactos de los *fluxes* metabólicos y las actividades enzimáticas (Park *et al.*, 2005).

### 1.1.2.3 Metabolómica

Por la misma naturaleza de las redes metabólicas, se pueden usar las concentraciones de los metabolitos intracelulares para el análisis del metabolismo (Borodina y Nielsen, 2005). La metabolómica es cualquier técnica que intente identificar y cuantificar todos los metabolitos presentes en una muestra biológica de un organismo creciendo en unas condiciones determinadas; esto implica la cuantificación de gran número de intermediarios metabólicos, esto es, de metabolitos internos (Ishii *et al.*, 2004; Lee *et al.*, 2006). Una ventaja del metaboloma sobre el transcriptoma o el proteoma, es que es numéricamente más pequeño y por consiguiente el manejo de datos es más sencillo (Oliver *et al.*, 1998).

La interpretación de cambios en la abundancia de metabolitos no es tan sencilla, pues la cantidad presente de un metabolito no depende únicamente de los *fluxes* que lo producen, sino también de los *fluxes* que consumen el metabolito. Por esta razón, para analizar correctamente los cambios en las redes metabólicas, es necesario cuantificar los cambios en los *fluxes* (Baxter *et al.*, 2007).

Se puede considerar que la metabolómica amplifica los cambios en el proteoma, y representa de manera más cercana el fenotipo de un organismo (Kell, 2006). De esta manera, los análisis metabolómicos contribuyen en la interpretación del impacto real de cambios en las actividades enzimáticas sobre los *fluxes*, y son una herramienta poderosa para validar rutas *in silico* construidas a partir de secuencias genómicas (Park *et al.*, 2005). El desarrollo de plataformas tecnológicas de perfilamiento (*profiling*) de metabolitos permite analizar cambios metabólicos a un nivel de detalle sin precedentes (Fernie *et al.*, 2004). Analizando la abundancia de metabolitos en diferentes tiempos, es posible obtener información detallada acerca de la naturaleza y magnitud de las perturbaciones metabólicas tras un cambio de condiciones ambientales en una célula (Baxter *et al.*, 2007).

Si bien el objetivo último de la metabolómica sería obtener mediciones de la concentración de todos los metabolitos en un sistema particular, se mide normalmente un subconjunto más restringido, denominado el “perfil metabólico” (Kell, 2006). Normalmente, la medición de metabolitos incluye una etapa de separación, usando electroforesis o cromatografía de gases o líquida, acoplada con una etapa de identificación, usualmente llevada a cabo por espectrometría de masas.

Algunas aplicaciones directas de la medición del metaboloma incluyen: la detección de biomarcadores para detectar defectos metabólicos “de nacimiento” (Pitt *et al.*, 2002; Vaidyanathan *et al.*, 2006), estudios del modo de acción, con miras a determinar el sitio en que actúa un inhibidor o farmacéutico o sitio en que se encuentra una lesión genética (Allen *et al.*, 2004), y análisis de toxicidad y de eficiencia de medicamentos (Lindon *et al.*, 2005; Picard y Bergeron, 2002).

#### **1.1.2.4 Transcriptómica**

Transcriptómica consiste en la medición y análisis de los niveles de expresión de los ARN mensajero o ARNm (*mRNA – messenger ribonucleic acid*), usando chips de ADN (*DNA microarrays*) o de oligonucleótidos (Ishii *et al.*, 2004). Generalmente, se emplean dos tipos de chips, los cuales difieren en la longitud de cadena de oligonucleótidos. Posteriormente, se usa hibridización basada en fluorescencia entre los fragmentos y el ADNc (*cDNA – complementary deoxyribonucleic acid*), de modo que se determina el nivel de expresión de los genes contenidos en cada fragmento (Park *et al.*, 2005).

Actualmente, es factible monitorear y examinar la expresión genética global usando análisis con chips de ADN, dado que se conozca la secuencia genómica del organismo. Por otra parte, los patrones de expresión de ARNm no necesariamente son consistentes con los niveles de concentración de las proteínas, es decir, no existe una buena correlación entre el transcriptoma y el proteoma (Park *et al.*, 2005). Se han usado los chips de ADN con éxito para examinar cambios en los perfiles de expresión de ARNm de microorganismos bajo diversas condiciones ambientales (Oh *et al.*, 2002; Schoondermark-Stolk *et al.*, 2006; Schweder y Hecker, 2004), así como tras sufrir alteraciones genéticas para la producción de proteínas recombinantes (Choi *et al.*, 2003; Lee *et al.*, 2003).

La información transcriptómica puede integrarse en un modelo metabólico. Si no se ha expresado un gen determinado, la proteína correspondiente estará ausente, y por consiguiente, en el modelo metabólico la reacción catalizada por esa enzima debe eliminarse del modelo. Esta misma idea puede extenderse un poco más, teniendo en cuenta que muchas reacciones requieren de múltiples proteínas formando un complejo, o de proteínas chaperonas, entonces, si cualquiera de estos factores está ausente, el *flux* de la reacción correspondiente se restringe a cero (Åkesson *et al.*, 2004).

#### **1.1.2.5 Metagenómica**

La metagenómica consiste en el análisis simultáneo de los genomas de un conjunto de microorganismos que coexisten en un medio determinado. Este campo de estudio surgió debido a la imposibilidad de cultivar en laboratorio la gran mayoría de microorganismos presentes en ambientes naturales, y para poder describir especies desconocidas y explorar la diversidad de comunidades microbianas (Streit y Schmitz, 2004). Los principales objetivos de la metagenómica son la identificación de los microorganismos presentes en un ambiente, y la búsqueda y asignación funcional de genes de dichos organismos (Desai *et al.*, 2012).

Un estudio metagenómico se inicia con el aislamiento del DNA del nicho microbiano elegido. Posteriormente, se construyen bibliotecas de DNA, se realiza un minado mediante *screening* de clones orientado a la búsqueda de funciones o secuencias de interés, y opcionalmente se amplifican clones y secuencias de DNA seleccionadas (Streit y Schmitz, 2004). Las herramientas de secuenciación son las mismas que se usan en la determinación de genomas, pero el análisis de los resultados difiere significativamente, y de hecho, el principal cuello de botella actual en los estudios metagenómicos es el análisis de los datos (Desai *et al.*, 2012).

#### **1.1.2.6 Marcación isotópica**

La marcación isotópica puede ser usada para cuantificar *fluxes* metabólicos internos de las células. La técnica consiste en alimentar a la célula con una combinación de sustrato marcado junto con un sustrato no marcado. Los isótopos se distribuyen por toda la red cuando el sustrato es asimilado, y el patrón de marcación obtenido depende de la topología de la red y de los *fluxes* intracelulares. El sustrato puede ser uniformemente marcado o selectivamente marcado, y la selección es un componente fundamental del diseño experimental (Iyer *et al.*, 2007). La determinación de la distribución fraccional de los *fluxes* depende de un modelado del metabolismo, cuyos procedimientos de introducción y evaluación de restricciones y simplificaciones están relativamente bien establecidos, así como los procedimientos de ajuste de datos y solución del sistema (Mason y Rothman, 2004).

El cálculo de *fluxes* metabólicos a partir de medición de distribuciones isotópicas requiere que el sistema alcance estado estable tanto metabólicamente, esto es, que la tasa de cambio de las cantidades de metabolitos sea prácticamente cero, como isotópicamente, que implica que los patrones de marcación isotópica de los metabolitos no cambien con el tiempo (Iyer *et al.*, 2007). A continuación, se determina el enriquecimiento fraccional del isótopo de marcación usado en los metabolitos, lo que permite cuantificar una serie de *fluxes* metabólicos. El modelado metabólico necesario para analizar los datos de mediciones de transformaciones de sustancias que incluyen isótopos marcados requiere los siguientes elementos: fuentes metabólicas (*metabolic pools*), fuentes de sustrato incorporado al sistema, tasas de rutas metabólicas, los datos medidos experimentalmente, y a veces, fuentes combinadas (*combination pools*). Con estos cinco elementos es posible generar ecuaciones que representan los flujos isotópicos entre las fuentes metabólicas, y ajustarlas con la mejor aproximación posible a la realidad (Mason y Rothman, 2004).

La medición experimental de la distribución de metabolitos marcados isotópicamente se lleva a cabo tradicionalmente usando resonancia magnética nuclear (NMR – *Nuclear Magnetic Resonance*), aunque recientemente algunas otras técnicas se han logrado

aplicar para este fin, como por ejemplo la espectrometría de masas MALDI-ToF (*Matrix-Assisted Laser Desorption/Ionization-Time of Flight Mass Spectrometry*) (Velagapudi *et al.*, 2007), y la resonancia magnética nuclear en dos dimensiones (Zhang *et al.*, 2007).

Las fuentes metabólicas representan la abundancia de cada especie química, y son caracterizadas por cuatro componentes: concentración inicial, posición de la marcación isotópica debida a determinada secuencia de *fluxes*, las entradas de especies, y sus salidas (consumos de sustrato e intermediarios y producción de metabolitos). Algunos modelos requieren las llamadas fuentes combinadas, debido a que para algunos metabolitos no pueden ser distinguidas sus distintas fuentes metabólicas, a pesar de la elevada especificidad química de la resonancia magnética nuclear (Mason y Rothman, 2004).

En las fuentes de sustrato usadas como entrada al sistema, hay que tener en cuenta cuál es la fracción de sustrato marcado que se alimenta, y en dónde se encuentra la marcación en el sustrato. Es necesario también caracterizar esa alimentación como función del tiempo, pues su variación en el tiempo permite el planteamiento del sistema de ecuaciones que permitirá el ajuste de los *fluxes* (Iyer *et al.*, 2007).

El método más usado para ajustar el modelo a las mediciones experimentales obtenidas es el método de mínimos cuadrados (Mason y Rothman, 2004), en el cual se eligen como parámetros del modelo aquellos que minimicen el cuadrado de la distancia entre los datos obtenidos con el modelo y los datos experimentales (Montgomery y Runger, 2003). Se han empleado también exitosamente para este fin otros métodos numéricos, como por ejemplo los algoritmos genéticos hibridizados (Zhang *et al.*, 2007).

Mediante la obtención de los *fluxes* determinados por la marcación isotópica, además de otros que puedan medirse de forma experimental (por ejemplo, la ingesta de sustrato y la excreción de metabolitos secundarios), es posible encontrar los valores de otros *fluxes* en la red mediante el uso de alguna técnica de balance (Sauer, 2006; Schwerder *et al.*, 2004; Wiechert, 2002).

## **1.2. MODELADO DE LA RED METABÓLICA**

Tradicionalmente, el estudio de las rutas metabólicas ha sido un paradigma central en la biología. En un principio, estos estudios se enfocaron en el descubrimiento paso a paso de las reacciones que conformaban una vía metabólica específica (Papin *et al.*, 2003). De esta manera, los primeros modelos metabólicos publicados consistían de rutas metabólicas separadas, tales como la glicólisis, el Ciclo de Krebs, el transporte electrónico, la fotosíntesis, etc. (Gevorgyan *et al.*, 2008).

La base para la creación de un modelo metabólico con un alto nivel de detalle, lo constituye actualmente la descripción del genoma de un organismo a partir del cual se crea un modelo que se ajusta con una combinación de procedimientos automáticos y manuales (Kell, 2006). A un modelo metabólico creado de esa manera se le conoce como modelo metabólico a escala genómica, y este tipo de modelo es el elemento central de la biología de sistemas, puesto que permite incorporar diferentes tipos de datos en una estructura que puede analizarse matemáticamente (Borodina y Nielsen, 2005). La primera reconstrucción metabólica a escala genómica de un microorganismo apareció en 1999, y correspondió a un modelo de *Haemophilus influenzae* (Edwards y Palsson, 1999). En la década posterior, fueron creados más de 50 modelos metabólicos a escala genómica de organismos diferentes (Oberhardt *et al.*, 2009).

La aparición de modelos a escala genómica requiere del análisis de las capacidades y funciones de los modelos metabólicos con un enfoque de red. Esto ha conducido al desarrollo de nuevos análisis matemáticos que han permitido la utilización de los modelos genómicos para la contextualización de datos provenientes de técnicas de alto rendimiento, el diseño ingenieril de sistemas biológicos, la generación y prueba de hipótesis respecto a estructura y función de las redes, el estudio de relaciones entre múltiples especies, y la dilucidación de propiedades sistémicas que no pueden describirse por simple descripción de los componentes individuales, como rendimiento de producto, robustez de la red, reacciones correlacionadas y predicciones de medio de cultivo mínimo (Oberhardt *et al.*, 2009; Papin *et al.*, 2003). Redes metabólicas de alta calidad o con incorporación de información regulatoria han permitido el análisis de fenotipos tras procesos de delección genética, y también análisis de expresión genética para identificar genes regulados en un conjunto determinado de genes (Borodina y Nielsen, 2005).

Los modelos metabólicos son indispensables para estudiar la biología de sistemas del metabolismo, pues son los modelos matemáticos que actualmente incluyen una mayor fracción de los componentes celulares. Esto se debe principalmente a que el metabolismo de muchos organismos está bien establecido, permite la integración de información de muchas fuentes, y porque el concepto de balance de *fluxes* es sólido, y en general, los periodos transitorios del promedio de las células son cortos (Borodina y Nielsen, 2005; Thiele y Palsson, 2010). Otra característica atractiva, en contraste con, por ejemplo, las redes de señales celulares, es que el modelado metabólico está sujeto directamente a restricciones termodinámicas y estequiométricas (Palsson, 2006).

Las redes metabólicas operan en estado cuasi-estacionario, manteniéndose en dicho estado debido a un complejo arreglo de sistemas regulatorios (Baxter *et al.*, 2007). Por otra parte, las redes metabólicas deben pasar de un estado estacionario a otro para adaptarse a cambios en el ambiente externo, así como para responder a diferentes demandas en distintas fases del desarrollo del organismo (Bohnert y Sheveleva, 1998). Adicionalmente, hay sistemas biológicos en los cuales un verdadero estado estable nunca

es alcanzado. Estas dos situaciones requieren extender el alcance de los análisis de los *fluxes* metabólicos, de modo que permitan la estimación de *fluxes* durante estado no estable; es decir, recurrir al modelado dinámico de los procesos metabólicos (Baxter *et al.*, 2007).

Los modelos usados en análisis dinámicos del metabolismo se construyen típicamente por medio de un conjunto de ecuaciones diferenciales ordinarias describiendo el balance másico de cada especie. A pesar de su generalidad, la aplicación de este tipo de modelado mecanístico a escala genómica no es práctica, debido a que requiere gran cantidad de reacciones detalladas y parámetros cinéticos que aún no son conocidos (Min Lee *et al.*, 2008). Por otra parte, un modelo estequiométrico de la red metabólica posee muy pocos parámetros que necesariamente deben determinarse a partir de datos experimentales, y sin embargo, es suficiente para describir la funcionalidad en estado estable del organismo (Herrgård *et al.*, 2006).

### **1.2.1. Características estructurales de las redes metabólicas**

A partir del modelo metabólico, es posible evaluar una serie de propiedades topológicas y estructurales de las redes, tales como el grado de conectividad promedio de la red (Almaas *et al.*, 2005) o el tamaño del espacio de solución de la red, denotando de esta manera el conjunto completo de *fluxes* estables compatibles con las restricciones metabólicas. El espacio de solución puede encontrarse de forma determinística para redes pequeñas (Avis y Fukuda, 1992); para redes de mayor tamaño, hay que recurrir a técnicas estadísticas, como muestreo Monte Carlo (Wiback *et al.*, 2004), o estimaciones de su tamaño usando discretización y un algoritmo de transferencia de mensajes llamado propagación de creencia (Braunstein *et al.*, 2008). Igualmente, puede explorarse qué *fluxes* de la red son calculables a partir de cierto conjunto de valores de *fluxes*; si bien el sistema es subdeterminado, por la estructura de la red y su espacio nulo, y bajo la suposición de estado pseudo-estable, se pueden calcular algunos valores de *fluxes* internos, así como se puede encontrar qué información es necesaria para calcular algún *flux* en particular (Klamt *et al.*, 2002).

Algunas características estructurales de las redes metabólicas son las siguientes:

1. Son robustas. Esto quiere decir que la red puede conservar sus funciones bajo ciertas perturbaciones internas o externas. Esta característica suele presentarse en muchos sistemas biológicos, y en el caso de la red metabólica, es fácil de determinar observando que los organismos conservan su capacidad de crecer a pesar de haber sufrido algunas deleciones genéticas (Almaas *et al.*, 2005; Zhao *et al.*, 2006).

2. Son modulares. Esta característica se refiere a que el funcionamiento de la red puede considerarse hasta cierto punto como el producto de la interacción de diferentes módulos. Esta propiedad tiende a confinar algunas perturbaciones y daños en ciertas regiones de la red, previniendo que ciertos efectos indeseables lesionen toda la célula (Almaas *et al.*, 2005; Kitano, 2004).

3. Poseen una estructura de tipo corbatín (*bow-tie*). Desde el punto de vista estructural, hay un pequeño conjunto de metabolitos fuertemente interconectados en la red porque participan en gran número de reacciones metabólicas (ATP, NADH, NADPH, entre otros), que por consiguiente generan el núcleo del sistema desde el punto de vista de conectividad. En contraste, la mayoría de metabolitos tienen una baja conectividad, poseyendo mayor variabilidad y menos restricciones (Zhao *et al.*, 2006). Ese patrón de conectividad provee una ventaja en la generación de respuestas coordinadas a diversos estímulos (Kitano, 2004).

4. Tienen una topología libre de escala (*scale-free*). Esta es una propiedad que tiene que ver con la distribución de la conectividad de los nodos respecto a la fracción de los nodos. La característica clave de este tipo de redes es que presentan un alto nivel de tolerancia al error; la habilidad de sus nodos de comunicarse no es muy afectada por la falla de algunos nodos seleccionados aleatoriamente (Albert *et al.*, 2000). Esto puede explicar por qué algunas disfunciones enzimáticas no tienen efecto fenotípico sustancial (Jeong *et al.*, 2000). Sin embargo, se ha mostrado que nodos de bajo grado tienen casi la misma probabilidad de resultar esenciales para la célula que los nodos de alto grado de conectividad (Mahadevan y Palsson, 2005). Esto demuestra que la interacción entre estructura y función en las redes reales tiene una elevada complejidad, y por consiguiente, es arriesgado intentar obtener conclusiones a partir de un determinado aspecto de la topología (Zhao *et al.*, 2006).

### **1.2.2. Construcción de modelos metabólicos a escala genómica**

La dimensión creciente de los modelos metabólicos puede conducir a serios problemas en su construcción y validación (Gevorgyan *et al.*, 2008). Además, también es un hecho que la mayoría de modelos metabólicos que han sido creados a escala genómica no han sido sometidos a un proceso apropiado de validación y comprobación experimental (Terzer *et al.*, 2009). Por las razones anteriores, se ha hecho evidente la necesidad de estandarizar la creación de modelos a gran escala, así como técnicas para determinar posibles errores en la construcción, y criterios de validación más unificados. El procedimiento general de creación de un modelo de la red metabólica a escala genómica se puede describir por cuatro etapas (Thiele y Palsson, 2010):

1. Creación de un borrador de la reconstrucción: La primera etapa consiste en crear un borrador de la red, basándose en la versión más reciente del genoma del organismo y en bases de datos bioquímicas. Esta etapa puede llevarse a cabo hasta cierto punto de forma automática, y hay que tener en cuenta que los errores que puede incluir el borrador pueden provenir tanto del genoma (funciones falsamente incluidas, o anotaciones erróneas, incompletas o faltantes) como de las bases de datos (asignación errónea de una función metabólica por homología, por ejemplo) (Oberhardt *et al.*, 2009).

2. Refinación manual de la reconstrucción: Esta es la etapa de creación del modelo que consume más tiempo. Para cada gen y reacción, debe cuestionarse la presencia de cada elemento, y considerar la introducción de nuevos elementos para conectar el gen o reacción con el resto de la red. Esta revisión manual es importante, porque no todas las fuentes de información sobre el metabolismo de un organismo tienen el mismo nivel de confiabilidad (Thiele y Palsson, 2010). En cualquier caso, es evidente que la tarea de modelado se simplificará a medida que aparezcan más métodos automáticos de detección de error y validación del modelo (Gevorgyan *et al.*, 2008).

El primer paso es ensamblar la reconstrucción, preferiblemente ruta por ruta. De esa forma, es más fácil identificar la existencia de vacíos en las rutas y algunos otros efectos estructurales. Posteriormente, se revisan las reacciones, en cuanto a la existencia de evidencia real de la función metabólica asignada, el uso de sustratos y cofactores específicos, el balance de carga, la estequiometría y la dirección de la reacción, la información acerca de la ubicación del gen y de la reacción (en caso de que haya diversos compartimentos celulares), y la asociación gen-proteína-reacción (Österlund *et al.*, 2011). Posteriormente, se incluyen en la reconstrucción las reacciones de transporte intracelular, de ser necesarias, así como algunas reacciones que ocurren espontáneamente en el interior de las células; no hacen parte del metabolismo propiamente, pero ocurren y alteran las especies químicas presentes, por consiguiente lo ideal es incluirlas. A continuación, se define la reacción de biomasa, que representa los metabolitos que usa el organismo para crear biomasa, así como su proporción. Se determina también la magnitud de las reacciones de mantenimiento asociadas y no asociadas al crecimiento, que representan el consumo energético de la célula debido al mantenimiento de sus funciones. La primera de ellas se incluye como costo energético en la reacción de biomasa, la segunda se representa como una reacción de hidrólisis de ATP. Para terminar con la revisión manual, se determinan cuáles reacciones de intercambio se consideran reversibles, teniendo en cuenta también qué productos de desecho deben salir del sistema y cuáles son los requisitos mínimos de nutrientes del medio de cultivo para el organismo en cuestión (Thiele y Palsson, 2010).

Se han presentado algunos algoritmos que permiten la identificación de varios problemas de consistencia en la red, cuando provienen de la incorrecta definición de la

estequiometría de la reacción, y pueden resultar útiles en esta etapa de refinación manual del modelo (Gevorgyan *et al.*, 2008).

3. Conversión de la reconstrucción en un modelo matemático: La tercera etapa puede automatizarse casi por completo, y consiste en transformar la reconstrucción en un formato matemático, describiendo la estequiometría y las restricciones de la red. Aunque estas últimas pueden variar de acuerdo con la aplicación en que se usará la red metabólica y las condiciones ambientales en que se estudiará al organismo (Thiele y Palsson, 2010).

4. Depuración de la red: La cuarta fase consiste en la verificación, evaluación y validación del modelo. Algunos errores comunes en las reconstrucciones son la presencia de restricciones erróneas, la desaparición de reacciones de intercambio o de transporte, la imposibilidad de producir o consumir algún cofactor, y el lanzamiento de compuestos a través de un compartimento. Se debe también verificar si la red presenta rupturas (vacíos, reacciones faltantes) en las rutas metabólicas, si hay reacciones bloqueadas (reacciones que no pueden ocurrir en ninguna condición simulada) y si es posible la producción de todos los precursores descritos de la biomasa (Thiele y Palsson, 2010). La validación del modelo suele hacerse de forma cualitativa, evaluando el modelado de fenotipos de deleciones individuales, de incapacidades conocidas del organismo y de algunas otras propiedades fisiológicas conocidas, aunque también se usa la comparación cuantitativa de fenotipos de crecimiento, y está empezando a aplicarse la evaluación cualitativa de la predicción del crecimiento celular para determinar mejoras posibles a los modelos metabólicos a escala genómica (Österlund *et al.*, 2011).

### **1.2.3. Modelado cinético**

Cuando el objetivo del modelado es entender y predecir el efecto de cambios en actividades y propiedades cinéticas enzimáticas, o evaluar el efecto de cambios en las condiciones ambientales o de diversos eventos regulatorios, e incluso la modulación de pasos catalizados por enzimas y procesos de transporte, debe recurrirse al modelado cinético del metabolismo, en el cual se tiene en cuenta el comportamiento dinámico intracelular (Heijnen, 2005; Polisetty *et al.*, 2006).

Un modelado matemático guiado por una detallada descripción de los mecanismos de operación del metabolismo, el procesamiento de señales y la regulación genética, conduciría en principio, a la obtención de predicciones cuantitativas sobre la dinámica celular (Stelling, 2004). Para ello, se han hecho esfuerzos para aplicar la nueva e incrementada disponibilidad de datos, producto de las técnicas de alta fidelidad, en el modelado dinámico (Polisetty *et al.*, 2006). Sin embargo, estos modelos han presentado

inconvenientes para integrar información proveniente de estudios “ómicos”, pues la complejidad de los sistemas biológicos, así como la falta de conocimiento acerca de mecanismos y parámetros asociados, ha limitado la aplicación en las representaciones dinámicas (Terzer *et al.*, 2009). Las principales complicaciones pueden resumirse en tres categorías. La primera, el modelo que se use debe estar lo suficientemente detallado para capturar la dinámica del sistema, pero también requiere una estructura que permita la asignación de significado biológico a los parámetros. La segunda, la preparación de los datos y el pre-procesamiento de la tarea misma son complejos computacionalmente. Y la tercera es la estimación de los parámetros del modelo; debido a las inherentes no-linealidades, la búsqueda corre el riesgo de atascarse en óptimos locales y de sufrir problemas de convergencia, ya sea por falta de ésta o muy baja velocidad de la misma (Borodina y Nielsen, 2005; Polisetty *et al.*, 2006; Terzer *et al.*, 2009).

Respecto a la preparación de los datos, el modelado cinético presenta inconvenientes como consecuencia de la insuficiencia de mediciones experimentales intracelulares (Llaneras y Picó, 2008), lo que ha ocasionado que en muchos casos se use información cinética obtenida de estudios hechos con enzimas purificadas, realizadas por lo general en condiciones no fisiológicas y que por consiguiente tienen dudosa aplicación en el modelado celular (Teusink *et al.*, 2000; Nikerel *et al.*, 2006). Además, la cinética obtenida suele imposibilitar la obtención de una solución analítica para la red (Heijnen, 2005). Para una buena determinación de los parámetros del modelo se requiere que se lleven a cabo experimentos de perturbación de la red metabólica *in vivo*, sea cual sea el formato cinético usado (Nikerel *et al.*, 2006). Se presentan también complicaciones matemáticas debido a que el ruido en los datos experimentales tiene un efecto muy grande en el modelo, y porque se requiere estimar gran cantidad de parámetros debido a la insuficiencia de información presente, lo que se convierte en un problema computacional muy pesado (Polisetty *et al.*, 2006).

Para incrementar el número de mediciones experimentales celulares, una de las técnicas más valiosas ha sido la marcación isotópica de sustratos, aunque su uso en estado dinámico no deja de ser un reto considerable. El ajuste global de los parámetros a los datos no tiende a producir una solución única, excepto para redes metabólicas muy sencillas, esencialmente lineales (Morgan y Rhodes, 2002). Al alimentar a un sistema celular un precursor marcado isotópicamente, es posible escribir ecuaciones diferenciales que describan el balance de masa entre la entrada y la salida de carbono marcado isotópicamente a un *pool* metabólico dado. Estas ecuaciones pueden integrarse para encontrar los valores de *fluxes* que generan los patrones de marcación de metabolitos que mejor se ajusten a los datos experimentales. Sin embargo, debido a que en estado no estacionario no hay restricciones sobre los posibles valores de los *fluxes*, se hace excesivo el tiempo de cómputo para generar un modelo con mínimos errores cuadrados (Baxter *et al.*, 2007).

Una de las técnicas desarrolladas para estimar *fluxes* en estado no estacionario por medio de marcación isotópica se basa en la escritura de múltiples balances de masa para cada isotópomo másico, obteniéndose un sistema sobredeterminado (esto es, más ecuaciones que incógnitas). La medición de las distribuciones de isótopos se hace periódicamente por medio de cromatografía de gases-espectrometría de masas, obteniéndose una serie de tiempo de los cambios de los tamaños de los *pools* metabólicos y en el porcentaje de marcación de distintos metabolitos. Posteriormente, se aplican métodos de regresión lineal para obtener estimaciones de los *fluxes* en estado no estacionario. Esta técnica se usó para determinar los *fluxes* en estado no estable de la síntesis de treonina en una suspensión de células de *Arabidopsis* (Baxter *et al.*, 2007).

El enfoque inicial que se pensó aplicar para el modelado cinético fue la representación tradicional de Michaelis-Menten (Michaelis y Menten, 1913). Sin embargo, muy temprano se hizo evidente que esta representación no es apta para redes grandes metabólicas (Heinrich y Rapoport, 1974), aunque vale la pena mencionar que esto no ha impedido que se construyan modelos útiles usando esta representación (Shinto *et al.*, 2007). Los modelos cinéticos son típicamente construidos a partir de la descripción de los balances de masa por medio de ecuaciones diferenciales, y representando la cinética de las diferentes reacciones y procesos usando ecuaciones empíricas (Borodina y Nielsen, 2005), o información detallada de la cinética de procesos celulares específicos (Gombert y Nielsen, 2000).

Las técnicas más frecuentes de creación de modelos cinéticos se basan en la representación de la cinética de las reacciones y procesos intracelulares por medio de expresiones aproximadas. Hay una serie de características deseables en estas expresiones: la tasa de reacción debe ser proporcional al nivel enzimático, la relación entre tasa y concentración de metabolito debe mostrar una tendencia cóncava hacia abajo en la región de alta concentración de metabolito; el número de parámetros cinéticos debe ser tan bajo como se pueda, y es preferible que al ensamblar las aproximaciones cinéticas en una red puedan lograrse soluciones analíticas en estado estable (Heijnen, 2005).

Las técnicas más usadas para representar la cinética de los organismos son el análisis de control metabólico (MCA – *Metabolic Control Analysis*), las representaciones provenientes de la llamada teoría de sistemas bioquímicos de Savageau (BST – *Biochemical System Theory*) (Savageau, 1976), que incluye los sistemas-S (*S-systems*) y los modelos de acción de masas generalizadas (GMA – *Generalized Mass Action*), y el enfoque log-lineal (Guebel, 2004; Klipp, 2007; Polisetty *et al.*, 2006; Visser y Heijnen, 2003). A diferencia de las otras técnicas mencionadas, con el MCA se obtienen algunos parámetros descriptores de la cinética de ciertos procesos metabólicos. Sin embargo, no se obtiene un modelo completo. Además, todas estas representaciones son aproximadas de la cinética de las reacciones, y por esta razón, tienen rangos de validez limitados (Heijnen, 2005).

Aparte del uso de estos enfoques comunes, se han creado también modelos cinéticos mediante: la alteración de algunas técnicas de modelado estequiométrico, como es el caso del FBA dinámico (sección 3.2.1); la representación de la regulación celular por medio de expresiones cinéticas (Leduc *et al.*, 2006); el uso del enfoque de modelado cibernético, en el cual se considera a la regulación metabólica como un intento de la célula para efectuar ajustes óptimos como respuesta a cambios ambientales controlando la síntesis y actividades de las enzimas (Young *et al.*, 2008); el uso de una linealización de las expresiones cinéticas para representarlas de manera aproximada en cercanías de un punto de referencia (Steuer *et al.*, 2006); y la integración de unas cuantas expresiones cinéticas macroscópicas en un modelo estacionario, lo cual permite obtener predicciones de la relación entre las concentraciones de metabolitos extracelulares y de la ingesta de sustratos respecto del tiempo. Tres ejemplos de esta integración son el modelado dinámico de células de ovario de hámster chino (células CHO), usando un modelo metabólico pequeño (Provost y Bastin, 2004), el modelado cinético de fermentación *batch* de *S. cerevisiae* en producción de vino (Sainz *et al.*, 2002), y el modelado dinámico de la producción de etanol en un cultivo *fed-batch* usando *S. cerevisiae*, estudio en el cual se usó un modelo estequiométrico a escala genómica (Hjersted *et al.*, 2006).

### 1.2.3.1 Análisis de Control Metabólico (MCA)

Una técnica que permite determinar la importancia relativa de diferentes enzimas o agentes externos en el control de los valores de los *fluxes* metabólicos es el análisis de control metabólico (MCA – *Metabolic Control Analysis*). No es un enfoque con el que se pretenda modelar completamente la célula o su metabolismo, sino que se determinan una serie de coeficientes que cuantifican la relación entre un cambio en la concentración de una enzima o un ‘efector’ extracelular, y el cambio correspondiente en la concentración de un metabolito o *flux* intracelular (Visser y Heijnen, 2002). De este modo, es posible obtener cierta información sobre un mecanismo de control celular, sin necesidad de caracterizar todos los componentes del sistema (Leonard *et al.*, 2007).

Los coeficientes básicos del MCA se denominan coeficientes de control, coeficientes de respuesta y elasticidad, y se definen con relación a un estado estable de referencia, por lo cual son adimensionales (Visser y Heijnen, 2003).

Los coeficientes de control cuantifican el efecto del cambio en el nivel de una enzima, sobre un *flux* o concentración. Los coeficientes de control de *fluxes* se definen como se muestra en la ecuación [1].

$$C_{ij}^{v0} = \frac{e_j^0}{v_i^0} \frac{dv_i}{de_j} \quad [1]$$

Y los coeficientes de control de concentración como se presenta en la ecuación [2].

$$C_{ij}^{x0} = \frac{e_j^0}{x_i^0} \frac{dx_i}{de_j} \quad [2]$$

En estas ecuaciones, el superíndice 0 denota el estado de referencia,  $e_j$  es el nivel de la  $j$ -ésima enzima,  $v_i$  es el valor del  $i$ -ésimo flux, y  $x_i$  es la concentración del  $i$ -ésimo metabolito.

Los coeficientes de respuesta describen el efecto del cambio de un parámetro externo sobre la concentración de los metabolitos y los *fluxes* intracelulares. Los coeficientes de respuesta de los *fluxes* se definen según la ecuación [3].

$$R_{ij}^{v0} = \frac{c_j^0}{v_i^0} \frac{dv_i}{dc_j} \quad [3]$$

Y los coeficientes de respuesta de concentración como se indica en la ecuación [4].

$$R_{ij}^{x0} = \frac{c_j^0}{x_i^0} \frac{dx_i}{dc_j} \quad [4]$$

Donde  $c_j$  es la concentración del  $j$ -ésimo componente extracelular en el modelo.

La elasticidad es una propiedad local que cuantifica el cambio relativo en la velocidad de una reacción tras un cambio en el nivel metabólico, manteniendo todas las demás variables constantes, como se muestra en la ecuación [5].

$$\varepsilon_{ij}^{x0} = \frac{x}{v_i^0} \frac{dv_i}{dx_j} \quad [5]$$

El MCA permite examinar cómo se distribuyen las funciones de control en la red metabólica, lo cual resulta útil como guía para la ingeniería metabólica, y también ilustra aspectos importantes del control metabólico. Sin embargo, la aplicación del MCA es necesariamente limitada, porque la teoría es válida únicamente para cambios pequeños, en los cuales la linealización es una buena aproximación (Visser y Heijnen, 2003).

### **1.2.3.2 Modelos de acción de masas generalizados (GMA)**

Tanto el sistema-S como la representación GMA pueden servir para modelar múltiples tipos de redes bioquímicas, y se derivan de un formalismo denotado por las siglas BST

(*Biochemical System Theory*) que se basa en un enfoque de ley de potencias (Guebel, 2004; Cho *et al.*, 2006).

En la representación por GMA, el cambio en cada fuente dependiente se describe como la diferencia entre las sumas de todos los *fluxes* de entrada y salida de la fuente. Cada *flux* se linealiza usando coordenadas logarítmicas, que en coordenadas cartesianas corresponde al producto de funciones potenciales que contienen exclusivamente las variables que afectan directamente al *flux*, elevados a un exponente que se denomina el orden cinético. El producto contiene también una constante de velocidad que determina la magnitud de la velocidad del *flux* del proceso. De este modo, un modelo GMA se formula matemáticamente según la ecuación [6].

$$\frac{dx_i}{dt} = \pm \gamma_{i1} \prod_{j=1}^{n+m} x_j^{\vartheta_{ij1}} \pm \gamma_{i2} \prod_{j=1}^{n+m} x_j^{\vartheta_{ij2}} \pm \dots \pm \gamma_{ik} \prod_{j=1}^{n+m} x_j^{\vartheta_{ijk}} \quad [6]$$

para  $i = 1, 2, \dots, n$

Donde  $\gamma_{ik}$  son las constantes de velocidad correspondientes a las  $k$  reacciones de producción/consumo, y  $\vartheta_{ijk}$  son los órdenes cinéticos para la especie  $i$  en la reacción  $k$  que involucra las especies  $j$ . Las  $x$  representan las concentraciones de las especies. El signo de  $\vartheta_{ijk}$  indica si la especie es un activador o un inhibidor; los sustratos y los activadores tienen coeficientes  $\vartheta_{ijk}$  positivos, mientras que los inhibidores los tienen negativos. El exponente  $n$  es el número de ecuaciones diferenciales resultante, pero el modelo también contiene  $m$  variables independientes del tiempo que afectan al sistema pero no son alteradas por éste, siendo típicamente constantes entre experimentos. Los términos potenciales de la anterior ecuación son resultado de una aproximación por serie de Taylor, la cual puede aplicarse a una variedad esencialmente ilimitada de procesos, incluyendo interacciones, activaciones, inhibiciones, y procesos asociados con dilución y crecimiento (Polisetty *et al.*, 2006).

Los modelos GMA representan una gran mejora respecto al modelado de la cinética de una reacción por medio de un modelo lineal; sin embargo, cuanto más se aleje el nivel enzimático del nivel de referencia, peor será la aproximación. Las representaciones de rutas metabólicas usando modelos GMA pueden resolverse analíticamente, pero sólo para rutas metabólicas lineales, sin bifurcaciones; esto es un gran inconveniente para su aplicación en modelos a gran escala (Heijnen, 2005).

### 1.2.3.3 Sistema-S

El sistema-S es un caso particular de la representación GMA, en el cual existe máximo un término positivo y máximo uno negativo (Polisetty *et al.*, 2006). Esta representación surge

para superar el inconveniente del modelado GMA respecto a las rutas con bifurcaciones; todas las reacciones en que un metabolito es creado se agrupan en un solo término potencial, y lo mismo se hace con todas las reacciones en que el metabolito se consume, obteniéndose sólo los dos términos mencionados (Heijnen, 2005).

En el sistema-S la razón de cambio de la concentración de cada metabolito  $x_i$  se representa por una productoria, tal como se muestra en la ecuación [7].

$$\frac{dx_i}{dt} = \alpha_i \prod_{j=1}^{n+m} x_j^{g_{ij}} - \beta_i \prod_{j=1}^{n+m} x_j^{h_{ij}} \quad [7]$$

para  $i = 1, 2, \dots, n$

Donde  $n$  es el número de componentes de la red bioquímica (si es una red metabólica,  $n$  sería el número de metabolitos),  $\alpha$  y  $\beta$  son constantes de velocidad de reacción, y  $g_{ij}$  y  $h_{ij}$  son los órdenes cinéticos correspondientes (Voit, 2000). La cantidad de parámetros que debe estimarse es de por lo menos  $2n(n+1)$  (Cho *et al.*, 2006), por lo cual este enfoque ha podido aplicarse exitosamente a redes bioquímicas de tamaño moderado, como por ejemplo una ruta metabólica particular o un sistema regulatorio específico, y recurriendo generalmente a algoritmos evolutivos para conseguir estimar los parámetros que se requieren (Cho *et al.*, 2006; Kikuchi *et al.*, 2003; Tominaga *et al.*, 2000).

Si bien el desempeño de los sistemas-S es similar al del enfoque GMA para cambios pequeños en los niveles metabólicos, cuando se presentan cambios significativos en un subconjunto de actividades enzimáticas el procedimiento de agregación usado para crear el sistema-S conduce a grandes errores (Heijnen, 2005).

#### 1.2.3.4 Cinética lin-log

La representación lin-log de la cinética de las reacciones intracelulares es una representación semilogarítmica, aproximada y no mecanística, que sin embargo presenta ventajas cuando se ensambla toda la red de reacciones respecto a otras representaciones (Visser y Heijnen, 2003).

La ecuación [8] es el modelado de la cinética de las reacciones en la representación lin-log.

$$v = e(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k) \quad [8]$$

En la anterior ecuación,  $v$  es la velocidad de la reacción,  $e$  es el nivel enzimático (es decir, la cantidad de enzima que cataliza la reacción),  $x_1 \dots x_k$  son las concentraciones de los diferentes metabolitos que tienen influencia en la velocidad de la reacción, ya sean

activadores o inhibidores, y  $\beta_0 \dots \beta_k$  son parámetros. Es evidente que  $\beta_0$  cuantificará el efecto del nivel enzimático sobre la velocidad de reacción y debe tener signo positivo, mientras que los demás  $\beta_i$  representan los efectos regulatorios. Estos efectos tendrán signo negativo o positivo si corresponden a una inhibición o una activación, respectivamente.

A su vez, es conveniente introducir un estado de referencia que, como ya se mencionó, consiste en un estado estable adecuado para el análisis, y relacionar la cinética con éste, en analogía con el análisis que se lleva a cabo con el MCA. El estado de referencia sería entonces el presentado en la ecuación [9] (Heijnen, 2005).

$$v^0 = e^0 (\beta_0 + \beta_1 x_1^0 + \dots + \beta_k x_k^0) \quad [9]$$

Y a continuación se define la  $i$ -ésima elasticidad de referencia con la ecuación [10].

$$\varepsilon_{xi}^0 = \frac{\beta_i e^0}{v^0} \quad \text{para } i = 1, 2, \dots, k \quad [10]$$

De modo que la descripción de la cinética de la reacción respecto al estado de referencia queda tal como se muestra con la ecuación [11].

$$\frac{v}{v^0} = \frac{e}{e^0} \left( 1 + \varepsilon_{x1}^0 \ln \frac{x_1}{x_1^0} + \dots + \varepsilon_{xk}^0 \ln \frac{x_k}{x_k^0} \right) \quad [11]$$

De esta manera, la tasa de reacción queda descrita en función de las elasticidades y de los parámetros del estado de referencia, siguiendo el formato lin-log (Visser y Heijnen, 2003).

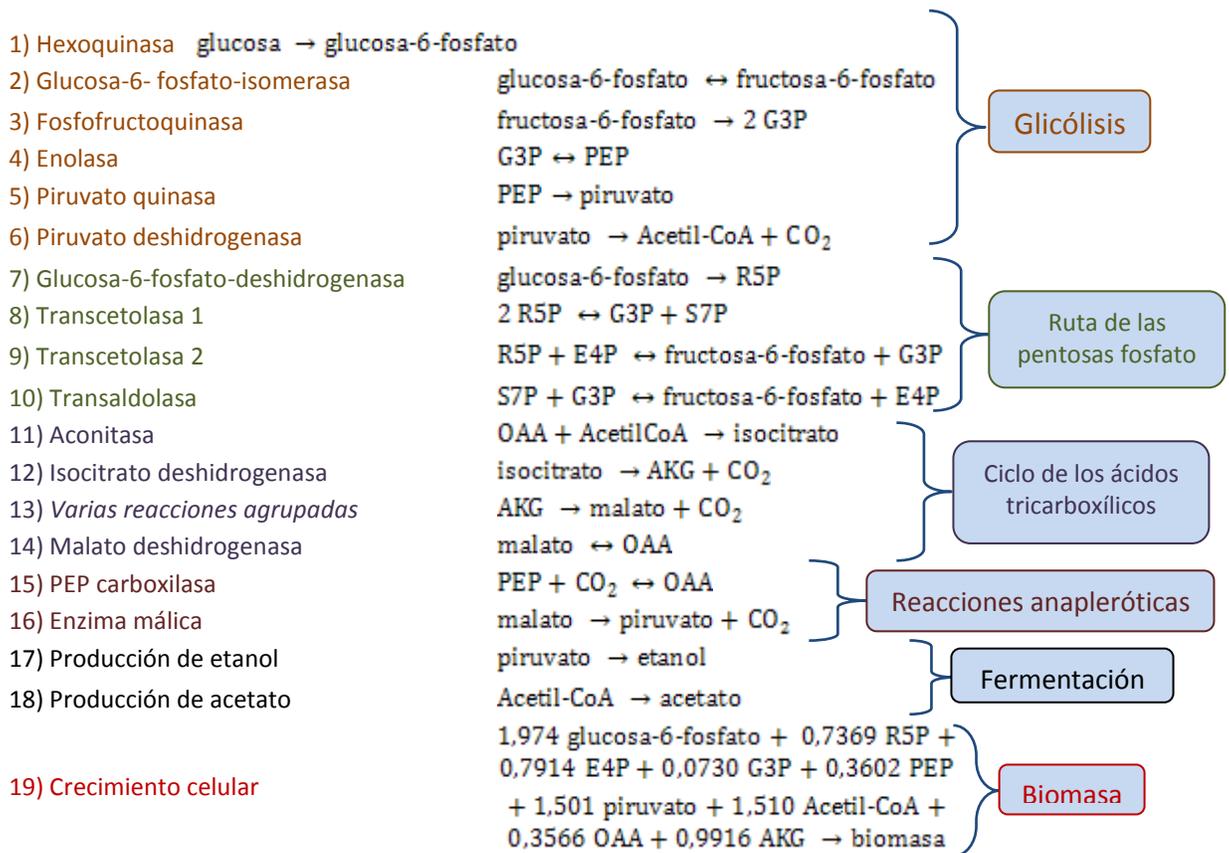
Un inconveniente del formato lin-log es que los logaritmos no están definidos en cero, lo cual implica que no se pueden representar estados en los cuales la concentración de los metabolitos es cero. El argumento en contra de este inconveniente es que se cree que existen mecanismos homeostáticos que mantienen las concentraciones de metabolitos en un rango limitado alrededor del estado estable (Heijnen, 2005). Por otra parte, tiene la gran ventaja de que este formato permite una solución analítica de los balances de masa en estado estable para la red metabólica, permitiendo determinar los niveles enzimáticos óptimos que maximizan un cierto *flux* (Nikerel *et al.*, 2006).

#### 1.2.4. Modelado estequiométrico

En general, el modelado estequiométrico permite analizar múltiples propiedades sistémicas de la red (Papin *et al.*, 2003), pero también se puede conseguir gran cantidad de objetivos diferentes de acuerdo con la información extra que se incluya en forma de

restricciones al modelo (Oberhardt *et al.*, 2009; Terzer *et al.*, 2009). Este modelado presenta múltiples ventajas sobre el cinético, tales como: los modelos estequiométricos están basados en coeficientes estequiométricos que son bien conocidos en la mayoría de los casos (Oberhardt *et al.*, 2009), permiten describir el comportamiento en estado estable del organismo (Herrgård *et al.*, 2006), y casi no requieren determinación o estimación de parámetros (Åkesson *et al.*, 2004). Independientemente de la variedad de alcances y objetivos, todas las técnicas de modelado estequiométrico tienen dos elementos en común: el primero es una representación de la red metabólica llamada matriz estequiométrica, y el segundo es la suposición de estado pseudo-estable, que consiste en modelar a la “célula promedio” con que se representa el comportamiento promedio de un conjunto de células como si estuviera en estado estable, de modo que la concentración de metabolitos intracelulares se mantiene constante; esta suposición es una simplificación matemática que resulta muy útil para múltiples objetivos (Gombert y Nielsen, 2000; Papin *et al.*, 2003; Llaneras y Picó, 2008).

**Figura 4.** Modelo del metabolismo central de *S. cerevisiae*. Siglas usadas: Acetil-CoA: Acetil-Coenzima-A; AKG: alfa-cetoglutarato; CO<sub>2</sub>: dióxido de carbono; E4P: eritrosa-4-fosfato; G3P: gliceraldehído-3-fosfato; OAA: oxaloacetato; PEP: fosfoenolpiruvato; R5P: ribulosa-5-fosfato; S7P: seduheptulosa-7-fosfato.



Uno de los elementos comunes de los modelos estequiométricos es la matriz estequiométrica. Esta es una matriz en la cual cada columna representa una reacción química o de transporte, con elementos diferentes de cero que corresponden al coeficiente estequiométrico de cada metabolito involucrado en la reacción, de modo que todos los metabolitos del sistema quedan representados en las filas de la matriz. La matriz estequiométrica también incluye información acerca de la dirección de las reacciones, pues los coeficientes de los sustratos de cada reacción se denotan con coeficientes estequiométricos negativos, mientras que los coeficientes de los productos de la reacción aparecen con signo positivo (Llaneras y Picó, 2008; Terzer *et al.*, 2009).

A manera de ejemplo, se presenta un modelo sencillo del metabolismo central de *S. cerevisiae* (adaptado de Zhang y Yao, 2007) que no incluye cofactores, y la construcción de la matriz estequiométrica correspondiente. Las reacciones incluidas en el modelo se presentan en la **Figura 4**.

**Tabla 4.** Matriz estequiométrica **S** correspondiente al modelo del metabolismo central de *S. cerevisiae* presentado en la Figura 4.

	Prod. biomasa	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Trans. glucosa	Trans. etanol	Trans. acetato	Trans. CO <sub>2</sub>
glucosa	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0
etanol	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	-1	0	0
acetato	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	-1	0
CO <sub>2</sub>	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	-1	1	0	0	0	0	0	-1
glucosa-6-fosfato	-1,974	1	-1	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
fructosa-6-fosfato	0	0	1	-1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
G3P	-0,073	0	0	2	-1	0	0	0	1	1	-1	0	0	0	0	0	0	0	0	0	0	0	0
PEP	-0,3602	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0
piruvato	-1,501	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0
Acetil-CoA	-1,510	0	0	0	0	0	1	0	0	0	0	-1	0	0	0	0	0	0	-1	0	0	0	0
R5P	-0,7369	0	0	0	0	0	0	1	-2	-1	0	0	0	0	0	0	0	0	0	0	0	0	0
S7P	0	0	0	0	0	0	0	0	1	0	-1	0	0	0	0	0	0	0	0	0	0	0	0
E4P	-0,7914	0	0	0	0	0	0	0	0	-1	1	0	0	0	0	0	0	0	0	0	0	0	0
OAA	-0,3566	0	0	0	0	0	0	0	0	0	0	-1	0	0	1	1	0	0	0	0	0	0	0
isocitrato	0	0	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	0
AKG	-0,9916	0	0	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0
malato	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	0	-1	0	0	0	0	0	0

La matriz estequiométrica correspondiente al modelo presentado en la Figura 4 se muestra en la **Tabla 4**. La escritura de la matriz estequiométrica comienza por decidir la posición de cada metabolito y reacción en la matriz. Posteriormente, se llena la información de cada reacción en la columna correspondiente, indicando qué metabolitos y en qué cantidad se producen y se consumen en esa reacción. El convenio habitualmente usado consiste en indicar el consumo de sustancias con signo negativo y la producción con signo positivo; por otra parte, la magnitud de la cantidad indica el coeficiente estequiométrico de la sustancia en la reacción. También es común, al menos en los modelos a gran escala, que la primera columna sea destinada a la escritura de la reacción de biomasa.

Adicionalmente a las reacciones del modelo, se deben incluir reacciones de intercambio con el medio extracelular, representando la entrada y salida de metabolitos. Es habitual escribir todas estas reacciones en la misma dirección y permitir que sean irreversibles, de modo que la dirección del *flux* (si el metabolito está entrando o saliendo de la célula) quede indicada por el signo del *flux* cuando se haya cuantificado o estimado. Por esta razón, las últimas cuatro columnas de la Tabla 4, que representan las reacciones de transporte extracelular, fueron escritas como si los cuatro metabolitos se produjeran en la célula y fueran excretados, a pesar de que es evidente que la glucosa es el sustrato y por consiguiente deberá entrar a la célula. En los modelos que incluyen múltiples compartimentos celulares, es necesario también agregar una serie de reacciones que representen el transporte de metabolitos entre compartimentos. Los mecanismos de transporte entre compartimentos son variados, desde transporte por difusión hasta intrincados mecanismos de lanzaderas de metabolitos (Cox y Nelson, 2005), así que para un buen modelado celular se necesita idealmente que el mecanismo de transporte esté bien identificado y se represente correctamente en la matriz; desafortunadamente, no siempre es el caso.

Por las características de las redes metabólicas, las matrices estequiométricas son matrices dispersas (es decir, que tienen gran cantidad de elementos iguales a cero), y poseen más columnas que filas. Esto es consecuencia de que las redes metabólicas presenten un nivel significativo de conectividad, lo cual les confiere características interesantes de resiliencia y versatilidad, así como una gran cantidad de grados de libertad que conducen a que el modelado del metabolismo suela ser un problema subdeterminado.

Una cualidad de las redes es la reversibilidad de las reacciones, la cual pareciera no estar incluida explícitamente en la matriz estequiométrica. Sin embargo, esa información se incluye en los modelos estequiométricos por medio de las restricciones sobre los valores de los *fluxes*, de las cuales se hablará más adelante en esta sección.

De acuerdo con lo anterior, cada fila de la matriz estequiométrica contiene toda la información acerca de las reacciones de la red en que un metabolito particular participa, así como en cuáles de ellas es producto y en cuáles es reactivo. Por consiguiente, cada fila de la matriz contiene el balance molar de cada uno de los metabolitos del sistema, de modo que al multiplicar la matriz estequiométrica por los valores de cada una de las reacciones del sistema (o *fluxes*), como función del tiempo en el caso general, se obtiene la razón de cambio de la cantidad o concentración (depende de las unidades del vector de *fluxes*) de cada metabolito con respecto al tiempo (Gombert y Nielsen, 2000; Papin *et al.*, 2003; Llaneras y Picó; 2008). Matemáticamente, se representa por medio de la ecuación [12].

$$S \cdot \vec{v}(t) = \frac{d\vec{c}(t)}{dt} \quad [12]$$

Donde  $S$  es la matriz estequiométrica,  $\vec{v}(t)$  es el vector de fluxes metabólicos, y  $\vec{c}(t)$  es el vector de concentraciones de los metabolitos intracelulares.

La inclusión de una expresión que represente la dinámica de la concentración de los metabolitos respecto al tiempo, conduce a que el modelo resultante sea dinámico. Pero como ya se mencionó, el segundo elemento común de los modelos estequiométricos es la suposición de estado estable (o *cuasi-* o *pseudo-*estable, como es llamado por muchos autores) para la célula, que ocasiona que los *fluxes* metabólicos no sean función del tiempo en el modelo, y por consiguiente que la razón de cambio de la concentración de los metabolitos respecto al tiempo sea cero (Stephanopoulos *et al.*, 1998). Por consiguiente, la ecuación [13] es la ecuación básica del modelado estequiométrico del metabolismo.

$$S \cdot \vec{v} = \vec{0} \quad [13]$$

La ecuación [13] contiene simultáneamente la información sobre la red metabólica de la célula y la suposición de estado estable que usan todas las técnicas y aplicaciones de este tipo de modelado. La ecuación fundamental es un sistema homogéneo de ecuaciones lineales, que requiere que cada metabolito intracelular sea consumido a la misma velocidad a que es producido. Los metabolitos que intercambia la célula con su entorno no están sujetos a esta restricción.

Las razones fundamentales para justificar la utilidad, en ciertas situaciones, de la suposición de que el sistema se encuentra en estado estable, son las siguientes:

1. El metabolismo opera en una escala de tiempo mucho más rápida que muchos eventos regulatorios o vinculados con la división celular, por lo cual se puede considerar que la célula se mantiene en un estado pseudo-estable, hasta que

ocurra alguna perturbación que lo mueva a un estado pseudo-estable diferente (Terzer *et al.*, 2009).

2. El análisis del sistema celular en estado pseudo-estable permite aplicar técnicas con las cuales se obtiene información importante sobre características y propiedades sistémicas de la red, que difícilmente podrían conseguirse considerando la dinámica celular (Llaneras y Picó, 2008; Oberhardt *et al.*, 2009).
3. La mayoría de procesos industriales son operados en condiciones cuasi-estacionarias, en las cuales los principales parámetros del proceso (concentración de sustrato, concentración de oxígeno, tasa de dilución, etc.) se mantienen constantes en lapsos de tiempo considerables (Palsson, 2000; Park *et al.*, 2005).
4. Un sistema biológico en el cual las propiedades macroscópicas permanecen constantes puede modelarse como si fuera producto del promedio temporal de una población celular estructurada, de modo que el elemento que se modela es una “célula promedio” (Henson, 2003; Wiechert, 2002).

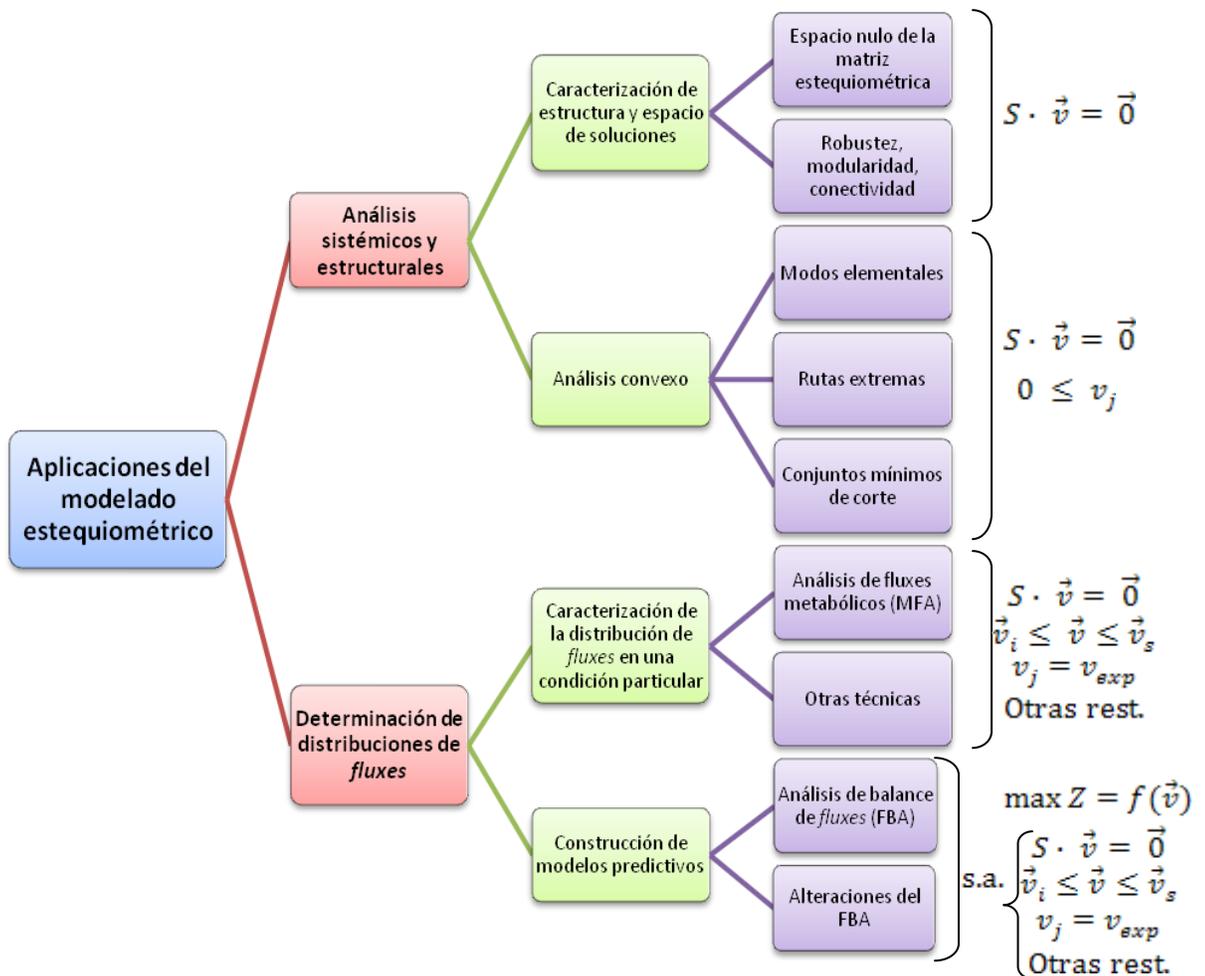
La información contenida en un modelo estequiométrico resulta en un sistema de ecuaciones lineales subdeterminado, y por consiguiente, para poder hallar una distribución de *fluxes* particular, se necesita combinar el modelo con datos experimentales adicionales o restricciones extra (Åkesson *et al.*, 2004). Se han creado gran cantidad de técnicas matemáticas basadas en el modelado estequiométrico (Gombert y Nielsen, 2000; Llaneras y Picó, 2008; Oberhardt *et al.*, 2009; Schilling *et al.*, 2000a; Wiechert *et al.*, 2001), con objetivos y alcances muy variados, entre los que se encuentran la búsqueda de comprensión estructural (Lee *et al.*, 2005), la simulación exploratoria (Heinrich y Schuster, 1996), la interpretación y evaluación de los datos medidos (Mason y Rothman, 2004; Takors *et al.*, 1997; von Lieres *et al.*, 2004; Wiechert y de Graaf, 1996), el análisis de sistemas (Albe y Wright, 1992; Edwards y Palsson, 1998; Goldbeter, 1996; Llaneras y Picó, 2007; Shiraishi y Savageau, 1992; Steuer *et al.*, 2006; Torres, 1994; Young *et al.*, 2008), la predicción de producción de metabolitos y biomasa (Famili *et al.*, 2003; Haag *et al.*, 2005; Segrè *et al.*, 2002; Shlomi *et al.*, 2005), el análisis de la interacción entre el metabolismo y la cinética de los procesos industriales (Hjersted *et al.*, 2006; Rodríguez *et al.*, 2005; Sainz *et al.*, 2002) y la optimización (Duarte *et al.*, 2004; Hatzimanikatis *et al.*, 1996).

Las células están sujetas a una serie de restricciones que pueden incluirse en los modelos estequiométricos; tras haber reconstruido la red y compilado la matriz estequiométrica, el siguiente paso de la creación de un modelo metabólico estequiométrico es identificar y asignar las restricciones apropiadas (Joyce y Palsson, 2007). De acuerdo con los objetivos del análisis, se requieren distintas cantidades y tipo de restricciones. En la **Figura 5** se presenta una clasificación de las técnicas de aplicación más comunes basadas en el modelado estequiométrico, junto con las expresiones de las restricciones, adicionales a la

ecuación básica del modelado estequiométrico, que se usan en la aplicación en cuestión. Los alcances y el funcionamiento de las diferentes técnicas se explican en las secciones 1.2.4.1 y 1.2.4.2.

Las restricciones pueden provenir de diversas consideraciones: ambientales, fisicoquímicas, evolutivas, cinéticas, regulatorias, etc. (Covert *et al.*, 2003; Price *et al.*, 2004); algunas restricciones pueden considerarse invariables y otras ajustables. Las primeras sirven para limitar el rango de comportamiento posible, y si bien no varían entre integrantes de una especie, puede esperarse que cambien como producto de la evolución. El segundo tipo de restricciones, como por ejemplo las constantes cinéticas, pueden variar entre diferentes individuos (Palsson, 2000).

**Figura 5.** Aplicaciones del modelado estequiométrico del metabolismo, junto con el modelo matemático básico de cada técnica.  $\vec{v}_i$  y  $\vec{v}_s$  son los vectores de los límites inferior y superior de los valores de los *fluxes*, considerando irreversibilidades, capacidades enzimáticas, etc.  $v_{exp}$  representa los valores medidos experimentalmente de algunos *fluxes*.



La misma matriz estequiométrica es implícitamente un conjunto de restricciones que proviene de los balances de masa y de carga de las posibles reacciones metabólicas de la célula, de modo que define un espacio geométrico de solución que incluye todos los posibles comportamientos metabólicos. Agregar más restricciones válidas al problema permite disminuir aún más el espacio de solución para una situación particular (Joyce y Palsson, 2007). De acuerdo con la cantidad y tipo de las restricciones que se inserten en el modelo, los objetivos y alcances de la aplicación de esa representación estequiométrica serán diferentes (Llaneras y Picó, 2008). Existen variadas aplicaciones del modelado estequiométrico, pero se clasificarán acá, de acuerdo con el objetivo, en análisis sistémicos y estructurales (sección 1.2.4.1), y determinación cuantitativa de distribuciones de *fluxes* particulares (sección 1.2.4.2).

#### **1.2.4.1 Análisis sistémicos y estructurales**

En contraste con los modelos cinéticos, que requieren gran cantidad de parámetros, o los modelos predictivos, que necesitan de múltiples datos de entrada y suposiciones, los análisis orientados a la estructura de las redes sólo requieren el conocimiento de la topología de la red, que es relativamente bien conocida en muchos casos (Stelling *et al.*, 2002). Muchos de los análisis sobre la estructura y las propiedades sistémicas de las redes metabólicas han surgido de la aplicación de técnicas de álgebra lineal y análisis convexo a modelos estequiométricos del metabolismo (Llaneras y Picó, 2008; Papin *et al.*, 2003).

Partiendo exclusivamente de la matriz estequiométrica, pueden llevarse a cabo algunos análisis sobre la consistencia y estructura de la red metabólica. La detección de las filas linealmente dependientes de la matriz estequiométrica conduce al hallazgo de los metabolitos que pertenecen a un grupo conservado, es decir, aquellos cuya concentración global permanece constante, como por ejemplo ATP, ADP y AMP (Terzer *et al.*, 2009). En el caso de modelos estequiométricos a escala genómica, se han realizado diversos estudios sobre propiedades sistémicas, tales como: la determinación de existencia de lazos cerrados (Kun *et al.*, 2008; Wright y Wagner, 2008), la evaluación de la conectividad de los metabolitos (Becker *et al.*, 2006; Samal *et al.*, 2006; Guimera *et al.*, 2007), y estudios sobre la redundancia de las rutas (Papin *et al.*, 2002; Mahadevan y Lovley, 2008).

La ecuación básica del modelado estequiométrico restringe el espacio de posibles soluciones, es decir, de posibles vectores de los *fluxes* intracelulares, a un subespacio de  $R^n$ , donde  $n$  es el número de *fluxes* en la red. En este modelo, las reacciones irreversibles se indican por medio de restricciones como se muestra en la ecuación [14].

$$0 \leq v_j \quad [14]$$

Con la inclusión de la información sobre reacciones irreversibles, el subespacio de posibles soluciones es un cono polihedral convexo en  $R^n$  (Llaneras y Picó, 2008).

La aplicación de análisis convexo al subespacio de posibles soluciones ha conducido a la determinación de los modos elementales (EMs – *elementary modes*), que consisten en las rutas más simples que son factibles estequiométrica y termodinámicamente para convertir sustratos en productos (Schuster *et al.*, 1999). Los modos elementales son un subconjunto de los vectores base del espacio nulo de la matriz estequiométrica, que cumplen dos condiciones adicionales: los modos elementales deben ser vectores factibles, y su combinación no-lineal debe generar todo el espacio de vectores de *fluxes* factibles (Terzer *et al.*, 2009).

Otro tipo de vectores con un objetivo similar a los modos elementales son las llamadas rutas extremas (EPs – *extreme pathways*) (Schilling *et al.*, 2000b). Éstas consisten en un subconjunto de los modos elementales, que resulta de un manejo ligeramente diferente de las reacciones reversibles de transporte (Klamt y Stelling, 2003), y satisfacen una condición adicional, que consiste en que ninguna ruta extrema pueda construirse como una combinación lineal no negativa de otras rutas extremas (Llaneras y Picó, 2008). La aplicación de las rutas extremas y los modos elementales es similar: son un conjunto de *fluxes* que pueden representar todas las capacidades metabólicas del microorganismo por medio de combinación convexa (Schilling *et al.*, 2000b).

Los modos elementales y las rutas extremas han sido usadas para llevar a cabo distintos análisis, como por ejemplo, el análisis de la contribución de rutas particulares sobre una determinada distribución de *fluxes* (Wiback *et al.*, 2003); la identificación de procesos metabólicos dominantes y su redistribución como respuesta a cambios ambientales, cinéticos o de nivel metabólico (Schwartz y Kanehisa, 2006); el grado de regulación transcripcional de los *fluxes* de *S. cerevisiae*, por medio del estudio de la relación entre los modos elementales y la regulación transcripcional (Çakır *et al.*, 2007); la comprensión de efectos inesperados en microorganismos recombinantes, como la ingesta de xilosa en medio aerobio, pero no anaerobio, por parte de *S. cerevisiae* modificada genéticamente para ingerir ese azúcar (Jin y Jeffries, 2004); la comparación de las capacidades metabólicas de *Helicobacter pylori* y *Haemophilus influenzae*, usando descomposición de valor singular sobre las matrices de rutas extremas para más fácil manejo (Price *et al.*, 2003); la comparación de la importancia relativa de diferentes modos elementales, por medio de la asignación de pesos relativos de importancia a los modos elementales involucrados en una distribución de *fluxes* particular (Wang *et al.*, 2007) y el estudio de los modos elementales involucrados en la fabricación de un producto de interés (Xu *et al.*, 2008).

Los modos elementales y las rutas tienen el inconveniente de sufrir una explosión combinatorial en su número a medida que la red se hace más grande (Zhao *et al.*, 2006).

Por esta razón, aún no es posible computar modos elementales o rutas extremas para un modelo metabólico a escala genómica (Terzer *et al.*, 2009). Para superar este problema, una de las alternativas sugeridas consiste en descomponer la red metabólica en módulos más pequeños, ya sean escogidos de una manera relativamente intuitiva (Schilling y Palsson, 2000), o de acuerdo con características de conectividad o modularidad de la red (Guimera y Amaral, 2005; Schuster *et al.*, 2002).

Otro concepto que se puede derivar de la matriz estequiométrica son los llamados conjuntos mínimos de corte (*minimal cut sets*), los cuales se definan como el conjunto más pequeño posible de reacciones que conduce a que la red falle respecto a una función específica, como por ejemplo, producción de biomasa (Klamt y Gilles, 2004). Los conjuntos de corte mínimos más simples son reacciones esenciales, puesto que tienen tamaño mínimo y su desactivación anula la reacción objetivo (Terzer *et al.*, 2009). Usando conjuntos mínimos de corte, se puede analizar la fragilidad interna estructural de la red, identificar objetivos potenciales para fármacos, conducir nuevas hipótesis y disminuir el número de posibles candidatos de prueba para experimentos costosos (Terzer *et al.*, 2009), así como determinar estrategias de intervención mínimas para producir en gran cantidad un bioquímico en una cepa bacteriana, como lo muestra el método OptKnock (Burgard y Maranas, 2003).

Aunque es posible crear modelos de redes metabólicas que contienen mucha información basada en anotación genómica e información de la literatura, hay casos en los cuales las predicciones obtenidas usando estos modelos no coinciden con los datos experimentales, y una de las posibles causas es que las reacciones del modelo no coincidan con las que realmente están activas. Una técnica sugerida para ajustar más fielmente las reacciones de un modelo metabólico se basa en determinar el conjunto de reacciones que presenta mayor concordancia entre los valores predichos *in silico* y distribuciones de *fluxes* medidas experimentalmente; esta técnica ha sido llamada identificación de red óptima metabólica (OMNI – *optimal metabolic network identification*) (Herrgård *et al.*, 2006).

#### **1.2.4.2 Determinación cuantitativa de distribuciones de fluxes**

En muchas metodologías dentro del modelado estequiométrico, se usan mediciones experimentales para fijar los valores de ciertos *fluxes* de la red, pudiendo reducir e incluso anular la indeterminación de la expresión básica del modelado estequiométrico (ver ecuación [13]). Esto se ha usado principalmente para estimar la distribución de *fluxes* presente en determinadas condiciones, y para desarrollar modelos capaces de predecir el comportamiento de los *fluxes* metabólicos en diferentes contextos (Llaneras y Picó, 2008).

Entre las técnicas para estimar la distribución de *fluxes* por medio de la obtención de suficientes mediciones para que el sistema sea determinado, la más destacada es la llamada análisis de *flux* metabólico (MFA – *Metabolic Flux Analysis*). El MFA consiste en la determinación de los valores de los *fluxes* metabólicos internos de la célula en un estado particular, a partir de la matriz estequiométrica y los valores de *fluxes* que puedan ser medidos (Palsson, 2006). De esta manera, adicional a la ecuación básica del modelado estequiométrico y a la información sobre irreversibilidad de las ecuaciones, se agregan al modelo restricciones igualdad, como la mostrada en la ecuación [15].

$$v_j = v_{exp} \quad [15]$$

La cantidad de mediciones necesarias y de *fluxes* estimados depende del tamaño y características del modelo metabólico usado, pues la determinación de valores de *fluxes* intracelulares depende del número de grados de libertad que posea el modelo metabólico usado, lo cual indican cuántas *fluxes* diferentes deben ser medidos para que el sistema esté determinado y se obtenga una solución única (Iyer *et al.*, 2007).

Tradicionalmente, el MFA se llevaba a cabo únicamente a partir de la información de *fluxes* de intercambio, es decir, ingesta de sustratos y producción de metabolitos (Vallino y Stephanopoulos, 1993). Algunos problemas con este enfoque tradicional son el pequeño tamaño de las redes que pueden quedar determinadas con tan pocas mediciones, la imposibilidad de obtener información confiable sobre reacciones paralelas o bidireccionales, y la dependencia sobre balances de NADH y/o NADPH, que pueden no ser precisos (Iwatani *et al.*, 2008).

Para superar los inconvenientes del MFA tradicional, se ha recurrido a agregar datos provenientes de marcación isotópica, y así cuantificar más *fluxes*. Se alimenta al microorganismo con sustratos marcados isotópicamente, y cuando se ha alcanzado un estado estable se mide la distribución de aminoácidos proteínogénicos marcados usando resonancia magnética nuclear, cromatografía de gases acoplada a espectrometría de masas (GC-MS – *gas chromatography-mass spectrometry*) o espectrometría de masas MALDI-ToF (Fischer y Sauer, 2003; Grotkjaer *et al.*, 2004; Velagapudi *et al.*, 2007). La determinación de la distribución interna de los *fluxes* requiere métodos analíticos y software de bioinformática (ver por ejemplo Zamboni *et al.*, 2005). Sin embargo, la necesidad de haber alcanzado estados estables isotópicos y metabólicos es una limitación del método, porque se requiere mucho tiempo para alcanzar la estabilidad (Iwatani *et al.*, 2008).

Otra alternativa que se ha desarrollado para mejorar el MFA consiste en cuantificar los metabolitos intermediarios marcados isotópicamente, a través de un muestreo rápido de microorganismos creciendo en un quimiostato, seguido por extracción y detección usando

cromatografía líquida acoplada a espectrometría de masas de los metabolitos intracelulares libres marcados (van Winden *et al.*, 2005).

Se han logrado también mejoras en la aplicación del MFA por medio de la inclusión de restricciones termodinámicas en el proceso de estimación de los *fluxes*. A través del cálculo de las energías de Gibbs de las reacciones de la red, usando el método de contribución de grupos y ajustando los valores según la fuerza iónica estimada en la célula, se obtiene un conjunto de restricciones mixtas enteras lineales, las cuales se tienen en cuenta junto con la matriz estequiométrica para la determinación de los *fluxes* metabólicos. A esta técnica se le denominó análisis de *flux* metabólico termodinámico (TMFA – *thermodynamics-based metabolic flux analysis*) (Henry *et al.*, 2007).

Además de las restricciones sobre irreversibilidad de las reacciones, se puede incluir en el modelo la información sobre el máximo valor que pueden tomar los *fluxes*, partiendo de limitaciones sobre capacidades enzimáticas o de transporte. Esta información se representa por restricciones como se indica en la ecuación [16].

$$v_j \leq v_{max} \quad [16]$$

Con la inclusión de estas restricciones, el subespacio de posibles soluciones queda limitado, convirtiéndose en un polítopo convexo en  $R^n$  (Llaneras y Picó, 2008).

Otra alternativa al MFA, que incorpora las restricciones sobre valor máximo de los *fluxes* e incluso la incertidumbre de las mediciones experimentales, es el enfoque de espectro de *fluxes* (FSA – *flux spectrum approach*). El objetivo de la técnica es utilizar toda la información disponible para limitar los valores que pueden tomar los *fluxes* en la red, aunque no haya suficientes mediciones para que el sistema no tenga grados de libertad. Lo primero que se hace es tomar las mediciones experimentales y calcular la mayor cantidad de *fluxes* intracelulares que sea posible. Posteriormente, para todos los *fluxes* que no pudieron determinarse, se encuentran los valores mínimo y máximo que pueden tomar, teniendo en cuenta las restricciones existentes, los valores de los *fluxes* intracelulares determinados y la incertidumbre de las mediciones, por medio de programación lineal. De esta manera, se cuantifican unos *fluxes*, y para los demás se halla su intervalo de posibles valores para la condición particular explorada (Llaneras y Picó, 2007).

Para lograr que los modelos estequiométricos puedan ser usados para predecir las distribuciones intracelulares de *fluxes* en diferentes condiciones superando el inconveniente de los grados de libertad positivos, actualmente la mejor solución es convertir el problema de estimación de los *fluxes* en un problema de optimización. Esta estrategia es la base del FBA, la técnica más exitosa hasta el momento para generar

predicciones, aunque su aplicación no se limita a esta finalidad (Oberhardt *et al.*, 2009; Österlund *et al.*, 2011; Terzer, 2009).

Convirtiendo el sistema en un problema de optimización, los grados de libertad positivos se usan en las variables de decisión, y se puede encontrar la distribución completa de *fluxes* metabólicos para una condición en particular (por ejemplo, cierta velocidad de ingesta de sustrato, o la ausencia de algún nutriente). Pero el problema de optimización requiere la definición de una función objetivo, que debe representar matemáticamente un criterio que explica el funcionamiento del sistema. De acuerdo con la función objetivo que se use, la distribución de *fluxes* que se obtendrá como respuesta será diferente. La técnica de FBA, las alteraciones propuestas al FBA y técnicas afines se tratan en más detalle en el capítulo 3.

La aplicación de otros criterios de optimización (extensiones del tipo de problema planteado en el FBA) a modelos estequiométricos ha permitido evaluar las productividades de un microorganismo en función de dos variables, por medio de la creación del plano de fase fenotípico, que es una proyección de la superficie de máximo crecimiento posible para todos las parejas de valores estudiadas de dos variables como tasa de ingesta de glucosa y tasa de ingesta de oxígeno (Schilling *et al.*, 2000a). Los planos de fase fenotípico se han usado para caracterizar fenotipos de *E. coli* y *S. cerevisiae*, así como para identificar genes esenciales para crecimiento en medio mínimo de estos organismos (Duarte *et al.*, 2004; Edwards y Palsson, 2000; Ibarra *et al.*, 2003). También vale la pena mencionar estudios que se ha utilizado la optimización para encontrar estrategias óptimas de borrado de genes para maximizar un producto deseado; esta estrategia se ha aplicado con modelos a escala genómica de *S. cerevisiae* (Patil *et al.*, 2005).

## 2. MODELOS METABÓLICOS DE *Saccharomyces cerevisiae*

*Saccharomyces cerevisiae* es ampliamente usado como microorganismo de producción biotecnológica, y se ha transformado en el ser vivo eucariótico más profundamente estudiado, lo cual lo ha posicionado como el organismo eucariótico modelo (Pham *et al.*, 2006). Resulta ser un organismo muy atractivo con el cual trabajar, por no ser patógeno y por su larga historia de aplicación en diversos procesos de producción, entre ellos comidas y bebidas, y recientemente incluso en procesos de la industria farmacéutica (Ostergaard *et al.*, 2000). La conjunción de las dos anteriores afirmaciones conduce a la existencia de una gran cantidad de conocimiento respecto a la genética, la bioquímica, la fisiología y los procesos de fermentación a gran escala de *S. cerevisiae*, que puede capitalizarse para lograr un incremento sustancial de la aplicación de esta levadura (Nielsen y Jewett, 2007).

Debido a que tradicionalmente las rutas metabólicas debían ensamblarse reacción por reacción en la medida en que se estudiaban las diferentes transformaciones bioquímicas (Cox y Nelson, 2005), en la era pre-genómica los modelos metabólicos de *S. cerevisiae* solían ser modelos muy limitados; aún los modelos más detallados no alcanzaban las 90 reacciones en total para representar el sistema metabólico (Österlund *et al.*, 2011).

El cromosoma III de *S. cerevisiae* fue el primer cromosoma completo de cualquier organismo en ser secuenciado (Oliver *et al.*, 1992), y el genoma de esta levadura fue el primero en ser secuenciado completamente entre todos los organismos eucarióticos; contiene alrededor de 12 millones de pares de bases y 6275 marcos de lectura abiertos, que se encuentran organizados en 16 cromosomas (Goffeau *et al.*, 1996, Dolinski y Botstein, 2005). La disponibilidad del genoma completo no solo generó un cambio en el nivel de detalle y complejidad de los modelos metabólicos, sino que abrió el camino a la formulación y solución de nuevos interrogantes acerca del funcionamiento sistémico del metabolismo de los organismos eucarióticos (Dolinski y Botstein, 2005).

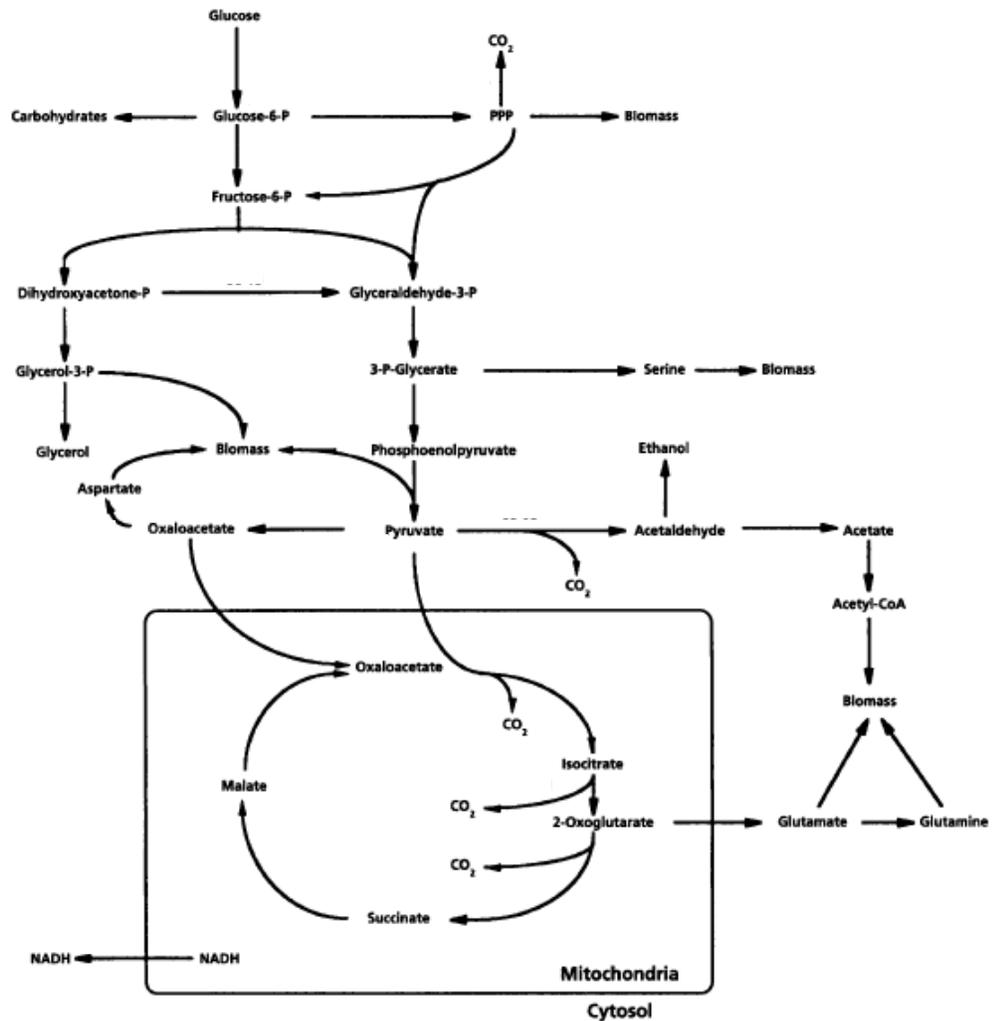
A continuación se presenta un resumen de los modelos metabólicos más importantes de *S. cerevisiae*. En la sección 2.1 se muestran algunos modelos de tamaño moderado, mientras que la sección 2.2 trata sobre los modelos metabólicos a escala genómica (se incluyen las reconstrucciones a escala genómica de modelos metabólicos; los intentos de integrar otros tipos de informaciones y redes en las redes metabólicas no son tratados acá).

### 2.1. MODELOS DEL METABOLISMO CENTRAL

Finalizando el siglo XX, no eran comunes los modelos metabólicos específicos de *S. cerevisiae*. Uno de los modelos interesantes desarrollados en la época data de 1995, en

un estudio en el cual fueron desarrollados simultáneamente modelos del metabolismo central para *S. cerevisiae* y para *Candida utilis* (Van Gulik y Heijnen, 1995). Las redes se construyeron recurriendo a la evidencia existente en la literatura de la época, y esto condujo a un modelo para *S. cerevisiae* con 88 reacciones y 81 metabolitos, que permitía simular el crecimiento aeróbico en glucosa, y anaeróbico en glucosa, produciendo etanol y acetato. En el modelo se incorporaron un par de parámetros, la proporción fósforo/oxígeno y el coeficiente de mantenimiento asociado al crecimiento, para aplicaciones en simulación.

Figura 6. Modelo del metabolismo central de *S. cerevisiae*.



Fuente: Nissen *et al.*, 1997.

Otro modelo de *S. cerevisiae* que vale la pena destacar es uno creado en 1997, cuyo objetivo principal era simular el crecimiento en cultivo continuo, con limitación de

sustrato, anaeróbico y creciendo en glucosa (Nissen *et al.*, 1997). Este modelo está compuesto por 37 reacciones y 27 metabolitos intracelulares, y su aplicación inicial fue la realización de un MFA para estimar la canalización de carbono en las reacciones intracelulares. En la **Figura 6** se presenta un gráfico representativo de este modelo; en él se muestran las rutas metabólicas centrales y se incluyen las reacciones que producen la mayoría de metabolitos excretados que se miden experimentalmente, como etanol, acetato, glicerol y piruvato, y se muestran algunas reacciones que producen dióxido de carbono. La ingesta de oxígeno no se muestra explícitamente, pues incluir este metabolito en la representación gráfica es inconveniente debido a que tiene una alta conectividad (es decir, participa en muchas reacciones).

Varios estudios posteriores se basaron en estos modelos, con los que ya quedaba compuesto un prototipo de los modelos metabólicos de *S. cerevisiae*. El desarrollo de estos modelos y su aplicabilidad proveyeron una base sólida para el desarrollo de modelos metabólicos exhaustivos tras la disponibilidad de la secuencia genómica de la levadura (Österlund *et al.*, 2011).

## 2.2. MODELOS A ESCALA GENÓMICA

El primer modelo metabólico a escala genómica de *S. cerevisiae* data de 2003 (Förster *et al.*, 2003). Desde aquel entonces, se han ido construyendo múltiples modelos, con variaciones principalmente en la cantidad de genes incluidos (y por consiguiente, la cantidad de metabolitos y reacciones representadas) y el número de compartimentos en que se clasifican los metabolitos. Otras diferencias consisten en el nivel de detalle de algunas rutas metabólicas específicas, la nomenclatura usada, el balance de la carga en las reacciones y la expresión de la reacción de biomasa.

Respecto a los nombres de los modelos a escala genómica, la convención habitual consiste en usar una sigla con la estructura *iXXyyy*, donde *i* representa la expresión *in silico* (construido con ayuda computacional; es una necesidad en los modelos a escala genómica, dado el tamaño de éstos), *XX* son las iniciales del o los investigadores encargados de la reconstrucción, y *yyy* es el número de genes metabólicos usados para crear el modelo.

### 2.2.1. iFF708

Esta primera reconstrucción no sólo fue pionera para *S. cerevisiae*, sino también fue la primera construcción a escala genómica de un modelo metabólico para un organismo eucariota.

El modelo iFF708 representa la información contenida en 708 genes, la cual está asociada con 1035 reacciones metabólicas. Tras incluir otras transformaciones con base en evidencia bioquímica, el modelo completó 1175 reacciones (842 únicas, es decir, sin considerar las isoenzimas como reacciones diferentes) y 584 metabolitos. Éstos se distribuyen en tres compartimentos: citosol, mitocondria y espacio extracelular (Förster *et al.*, 2003).

Para validar este modelo, se compararon las predicciones de viabilidad de la levadura tras distintas deleciones, respecto a datos experimentales de esencialidad de los genes. Los resultados mostraron un nivel de correcta predicción ligeramente superior al 80% (Famili *et al.*, 2003). En estos estudios usaron también el modelo para evaluar cuantitativamente las predicciones que se obtenían aplicando un FBA con maximización de biomasa como objetivo. La comparación la hicieron contra los datos experimentales presentados previamente en dos artículos, uno de ellos correspondiente a cultivos anaerobios con cuatro tasas de dilución diferentes (Nissen *et al.*, 1997) y el otro usando un experimento aerobio (Overkamp *et al.*, 2000). Por la manera en que fue ajustado el modelo a las diferentes situaciones, esta prueba no permite sacar conclusiones sobre el potencial predictivo del modelo, ni sobre la técnica o la función objetivo, pero sí indicó que el modelo abarcaba correctamente las regiones de valores en donde estaban los valores experimentales; en otras palabras, dentro de sus capacidades de modelado, se incluyó el espacio en que se encuentran los *fluxes* reales del organismo (Famili *et al.*, 2003).

### **2.2.2. iND750**

El modelo iND750 fue publicado en 2004. Como su nombre lo indica, el modelo incluye información sobre 750 genes, que junto con la evidencia en la literatura sobre reacciones, condujeron a un total de 1489 reacciones (1149 únicas), con 646 metabolitos diferentes. Este modelo distribuye los metabolitos en ocho compartimentos distintos: espacio extracelular, citosol, mitocondria, peroxisoma, núcleo, retículo endoplasmático, aparato de Golgi y vacuola (Duarte *et al.*, 2004).

La mayoría de genes 'nuevos' (que no están en el modelo iFF708) corresponden a tRNA sintetasas y ATPasas encontradas en la vacuola y el aparato de Golgi. La mayor parte de los metabolitos agregados están involucrados en reacciones que estaban agrupadas en el iFF708 y que fueron incluidas como pasos individuales en el iND750, sobre todo respecto a la degradación de ácidos grasos. Esto también influye en el incremento del número de reacciones, pero en este aspecto también fue de gran importancia la mayor compartimentalización del modelo, pues requiere un mayor número de reacciones de transporte intracelular. Otros cambios del iND750 respecto al iFF708 incluyen: introducción de información acerca de las asociaciones gen-proteína-reacción; remoción de abreviaturas de compuestos redundantes y reacciones duplicadas, y actualización de

nombres de genes y de nomenclatura de enzimas, es decir, de los números EC (*Enzyme Commission number*) correspondientes.

La validación de este modelo fue realizada comprobando las predicciones de viabilidad del organismo en 4154 casos de deleciones diferentes, cuya información experimental provenía de dos estudios diferentes (Giaever *et al.*, 2002; Steinmetz *et al.*, 2002); la viabilidad fue correctamente predicha en un 83% de los casos. Las predicciones fueron obtenidas aplicando un FBA al modelo tras representar la deleción, usando la maximización de producción de biomasa como función objetivo (Duarte *et al.*, 2004).

### **2.2.3. iLL672**

Este modelo es una revisión del iFF708. Se reformuló la ecuación de biomasa, y se ajustaron una serie de reacciones, tras identificar que 151 metabolitos no podían ser producidos o consumidos en estado estable. Esos metabolitos estaban involucrados en 143 reacciones, de las cuales 110 fueron removidas y 33 fueron conectadas basándose en nuevo conocimiento bioquímico (Kuepfer *et al.*, 2005).

El modelo resultante contiene 672 genes, y consta de 1038 reacciones en las cuales participan 636 metabolitos. La validación del modelo se hizo por evaluación de la predicción de letalidad de deleciones individuales usando tanto FBA, con maximización de biomasa como función objetivo, como MOMA, técnica llamada así por sus siglas en inglés (*Minimization of Metabolic Adjustment* - sección 3.2.5). La media geométrica del nivel de precisión del modelo para esencialidad de genes estuvo entre 81% y 89% para el FBA, y mejoró en un 3% al usar MOMA (Kuepfer *et al.*, 2005).

### **2.2.4. iIN800**

Este modelo fue presentado en el año 2008, y su principal objetivo fue incrementar el nivel de detalle sobre el metabolismo de lípidos. El modelo se construyó usando al iFF708 como marco, así que distribuye los metabolitos entre las mismas tres ubicaciones (espacio extracelular, citosol y mitocondria). iIN800 incluye información de 800 genes, y contiene 1446 reacciones en las que participan 1013 metabolitos (Nookaew *et al.*, 2008).

La mayoría de reacciones nuevas en el modelo pertenecen al metabolismo lipídico, principalmente a síntesis de ácidos grasos en la mitocondria, elongación, activación y beta-oxidación de ácidos grasos, síntesis de triacilglicéridos y fosfolípidos, y degradación de lípidos. Otro cambio con respecto a los modelos anteriores está en la reacción de biomasa, sobre la cual los autores presentaron diferentes expresiones de acuerdo a la existencia de limitación de carbono o de nitrógeno en el medio (Nookaew *et al.*, 2008).

Para validar el modelo, se compararon las predicciones de crecimiento al aplicarle un FBA al modelo, usando como restricciones de igualdad toda la información disponible sobre los *fluxes* de intercambio, respecto del crecimiento para los datos experimentales reportados en diez artículos diferentes (Aon *et al.*, 2001; Bakker *et al.*, 2000; Gombert *et al.*, 2001; Lidén *et al.*, 1995; Nissen *et al.*, 1997; Overkamp *et al.*, 2000; Tai *et al.*, 2005; Tai *et al.*, 2007; Usaite *et al.*, 2007; Vemuri *et al.*, 2007). Adicionalmente, se evaluó la predicción de esencialidad de 3392 casos de deleciones genéticas (incluyendo varios sustratos), obteniéndose una tasa de predicción del 89%. Los autores atribuyeron la mejora respecto a los modelos anteriores, a la revisión de la reacción de biomasa que se llevó a cabo para este modelo (Nookaew *et al.*, 2008).

#### **2.2.5. iMM904**

Este modelo fue publicado en el 2009, y es una extensión considerable del modelo iND750. iMM904 incluye información sobre 904 genes, 1228 metabolitos independientes y 1412 reacciones, de las cuales 395 son reacciones de transporte (Mo *et al.*, 2009).

Los genes y reacciones agregados están vinculados principalmente a los subsistemas de lípidos y carbohidratos, y a reacciones de transporte incluidas para llenar vacíos que impedían completar funciones fisiológicas conocidas y excretar metabolitos que han sido observados experimentalmente. Otros cambios importantes incluyen la modificación de asociaciones gen-proteína-reacción, y la curación manual de las reacciones de biosíntesis de cofactores como quinona, beta-alanina y riboflavina; la información presente en este aspecto en modelos anteriores había sido inferida a partir de los mecanismos presentes en modelos de *E. coli* (Mo *et al.*, 2009).

La validación del modelo se llevó a cabo comparando las predicciones *in silico* de deleciones de genes individuales con los resultados *in vivo*, tomados del estudio realizado para evaluar el modelo iLL672 (Kuepfer *et al.*, 2005). Para las predicciones *in silico*, se usaron dos formas del modelo, una completa y una reducida al número de genes de iLL672, y se recurrieron a tres técnicas para calcular los *fluxes*: el FBA, el MOMA (sección 3.2.5) y una versión lineal del MOMA. Los datos experimentales se clasificaron de acuerdo con el sustrato, y los autores concluyeron que las predicciones de esencialidad mejoraron con respecto al modelo iLL672 para los casos en que el sustrato era glucosa, galactosa o glicerol, pero empeoraron cuando el sustrato era etanol (Mo *et al.*, 2009).

#### **2.2.6. Red consensuada v1**

La principal motivación de este modelo fue la unificación de los términos usados en los modelos respecto a nombres de metabolitos y enzimas, así como empezar a trabajar

comunitariamente con el fin de integrar conocimientos y esfuerzos por el objetivo de conseguir el modelo metabólico de *S. cerevisiae* más real posible. Para esto, treinta y cuatro investigadores, incluyendo varios integrantes de algunos de los grupos más influyentes en la biología de sistemas y el modelado de la levadura (entre ellos, los grupos de Bernhard Palsson, Douglas Kell, Uwe Sauer y Jens Nielsen), reunieron esfuerzos con este objetivo, obteniendo un modelo al que simplemente se le llamó “red metabólica consensuada de la levadura” (*consensus yeast metabolic network*) (Herrgård *et al.*, 2008).

La estandarización de los nombres de los metabolitos se buscó referenciándolos con bases de datos persistentes, o usando formas independientes, como las representaciones con cadenas de caracteres denominadas SMILE (*Simplified molecular-input line-entry*) e InChi (*International Chemical Identifier*). Como modelos base para empezar la reconstrucción fueron usados los modelos iMM904 (antes de su publicación) e iLL672 (Kuepfer *et al.*, 2005). Las formas moleculares fueron incluidas en su estado de protonación más común en un medio a pH 7,2. Las proteínas y genes fueron anotadas con referencias a la base de datos de *Saccharomyces* (SGD – *Saccharomyces Genome Database* (Nash *et al.*, 2007)) y se usó el catálogo de información sobre proteínas UniProt (*Universal Protein Resource* (The UniProt Consortium, 2008)) (Herrgård *et al.*, 2008).

El modelo resultante contiene 1168 metabolitos, 832 genes y 1857 reacciones, y fue presentado en formato SBML (siglas de *systems biology markup language*, el cual consiste en un formato libre y abierto cuyo objetivo es estandarizar los datos para modelos computacionales de procesos biológicos). Los metabolitos se encuentran compartimentalizados en ocho ubicaciones diferentes, a saber: citoplasma, espacio extracelular, aparato de Golgi, mitocondria, núcleo, retículo endoplasmático, vacuola y peroxisoma, aunque en algunos de ellos el número de reacciones es muy bajo (en el aparato de Golgi y la vacuola sólo se incluyen dos reacciones). Muchas de las reacciones de los lípidos fueron removidas, en espera de que experimentos futuros mejoraran el conocimiento sobre el metabolismo lipídico en *S. cerevisiae*, aunque vale la pena mencionar que el modelo iIN800 fue presentado mientras el artículo de la red consensuada estaba en proceso de publicación. Para esta red, no se llevaron a cabo experimentos de validación (Herrgård *et al.*, 2008).

Esta red consensuada, sin embargo, no es apta para análisis computacional debido a la falta de información acerca de reversibilidad de algunas reacciones, así como la existencia de regiones no conectadas en la red (Österlund *et al.*, 2011).

### **2.2.7. iAZ900**

El modelo iAZ900 es un ajuste automatizado del modelo iMM904, buscando incrementar un tipo particular de predicciones de esencialidad de genes, conocido como especificidad,

que corresponden a las delecciones letales predichas correctamente (Zomorodi y Maranas, 2010).

Las correcciones propuestas por el procedimiento computacional fueron 120, y pertenecen a una amplia variedad de casos, incluyendo relajación de restricciones de irreversibilidad, adición de reacciones, compuestos o genes, modificación de la ecuación de biomasa, alteración de asociaciones gen-proteína-reacción y de restricciones específicas de medio. Para el cálculo, mantuvieron algunas restricciones que el modelo iMM904 trae por defecto, como restricciones en el valor máximo de ingesta de glucosa e ingesta de oxígeno, limitándolos a valores de 10 y 2 mmol/(g biomasa seca\*h). Igualmente, las predicciones fueron hechas aplicando un FBA con maximización de biomasa como función objetivo (Zomorodi y Maranas, 2010).

#### **2.2.8. Red consensuada v4**

Este es un modelo que refleja el avance en el trabajo realizado sobre la red consensuada publicada en el año 2008. El cambio en el modelo proviene principalmente de la incorporación de reacciones de la literatura, pero también de la exclusión de treinta y dos enzimas por insuficiente evidencia de su existencia y relocalización de metabolitos en compartimentos diferentes (Dobson *et al.*, 2010).

La cuarta versión de la red consensuada contiene 1102 reacciones metabólicas, con 924 genes e igual número de metabolitos. Esta versión tiene una conectividad superior respecto a la primera versión, aunque aún un porcentaje significativo de metabolitos permanece inalcanzable desde los constituyentes del medio de crecimiento (sustratos). Las predicciones de viabilidad al aplicar FBA con maximización de biomasa como función objetivo son comparables a las obtenidas con los otros modelos metabólicos a escala genómica (Dobson *et al.*, 2010).

### 3. ANÁLISIS DE BALANCE DE FLUX (FBA)

El análisis de balance de *flux* o FBA (*Flux Balance Analysis*) es una técnica de modelado estequiométrico que permite encontrar una distribución específica de *fluxes* metabólicos sin necesidad de gran cantidad de mediciones experimentales, por medio de la conversión del problema de estimar los valores de las reacciones metabólicas en un problema de optimización, generalmente uno de programación lineal (Raman y Chandra, 2009).

La historia del FBA se remonta a inicios de la década de 1990, cuando se empezó a usar la optimización de objetivos celulares para lograr estimar las capacidades de producción de diferentes posibles metabolitos secundarios por parte de *Escherichia coli* (Varma *et al.*, 1993). Debido a la inexistencia de bases de datos que compilaran las rutas metabólicas activas para determinado organismo en ese momento, los primeros modelos en los que se aplicó FBA eran modelos pequeños, involucrando generalmente el metabolismo central del carbono (Leonard *et al.*, 2007).

El FBA, al igual que las demás técnicas de modelado estequiométrico, está basado en la ecuación básica del modelado estequiométrico (ecuación [13]). Como ya se mencionó antes, esta ecuación representa un problema subdeterminado, debido a que siempre hay más reacciones que metabolitos en las redes metabólicas (Kauffman *et al.*, 2003), pero incluyendo restricciones sobre los valores de los *fluxes* (usando las ecuaciones [14] y [16], por ejemplo), se puede identificar el espacio de posibles valores de los *fluxes*, el cual es acotado (Schilling *et al.*, 1999). Estos valores mínimo y máximo pueden provenir de diversas fuentes: restricciones fisicoquímicas, restricciones espaciales o topológicas, restricciones dependientes de las condiciones ambientales y restricciones regulatorias (Raman y Chandra, 2009). Las reacciones irreversibles se representan fijando el límite inferior del valor del *flux* en cero. Para reacciones reversibles sobre las que no se tiene información acerca de los límites de su velocidad, en los valores mínimo y máximo se escriben valores grandes en magnitud, negativo el primero y positivo el segundo (Llaneras y Picó, 2008).

Adicional a estas restricciones, se deben incluir en el modelo algunos valores de *fluxes* de intercambio y/o la tasa de crecimiento de la célula en las condiciones ambientales y/o genéticas en que se desea modelar la célula. Estos valores se incluyen como restricciones de igualdad, y son los que permiten ajustar el comportamiento celular a la situación particular que se desea analizar. La introducción puede hacerse por medio de expresiones del tipo  $v_j = v_{exp}$ , aunque en la práctica la manera más habitual de hacerlo es igualando los valores inferior y superior del *flux* de intercambio deseado al valor experimental (Segrè *et al.*, 2002), tal como se presenta en la ecuación [17].

$$v_{inf,j} \leq v_j \leq v_{sup,j} \quad \therefore \quad v_{inf,j} = v_{sup,j} = v_{exp} \quad [17]$$

Todas esas restricciones, las provenientes de la ecuación básica del modelado estequiométrico más las relacionadas con los límites superior e inferior de los *fluxes*, son las restricciones en el problema de optimización que conforma el FBA. De esta manera, la determinación de una distribución de *fluxes* particular se consigue por medio de la suposición de que la célula funciona optimizando cierta meta, la cual se describe matemáticamente como la función objetivo del problema (Kauffman *et al.*, 2003), y el FBA puede describirse por medio de la ecuación [18].

$$\begin{aligned} \max Z &= f(\vec{v}) \\ \text{s.a.} \quad &\begin{cases} S \cdot \vec{v} = \vec{0} \\ \vec{v}_{inf} \leq \vec{v} \leq \vec{v}_{sup} \\ v_j = v_{exp} \end{cases} \end{aligned} \quad [18]$$

La suposición de comportamiento óptimo celular tiene ciertas implicaciones: (i) se está suponiendo que la célula ha evolucionado para alcanzar un comportamiento óptimo respecto a su objetivo, (ii) se asume que el objetivo supuesto de la célula es correcto, y (iii) se supone que dicho objetivo puede ser establecido matemáticamente en una forma apropiada (Llaneras y Picó, 2008).

Generalmente, el objetivo supuesto de la célula se representa por medio de una función lineal, tal y como se muestra en la ecuación [19].

$$\max Z = \vec{w}' \cdot \vec{v} \quad [19]$$

El vector  $\vec{w}$  representa los pesos que se le asignan a cada *flux* en la función objetivo. La función objetivo más comúnmente usada es la maximización del *flux* de la reacción de biomasa, lo cual implica que la célula funciona tratando de maximizar su crecimiento y reproducción. No obstante, se han estudiado muchas otras funciones objetivo (Raman y Chandra, 2009). En la sección 3.1 se aborda este tema.

Una complicación que se presenta al aplicar el FBA es la existencia de múltiples soluciones óptimas, en los cuales el mismo valor máximo de la función objetivo puede alcanzarse por medio de distintas distribuciones de *fluxes* (Lee *et al.*, 2000). Sin embargo, no puede descartarse que existan múltiples posibles distribuciones de *fluxes* para un estado ambiental particular, teniendo en cuenta que las redes metabólicas son robustas. Igualmente, no en todas las situaciones la variabilidad es significativa, y afecta casi exclusivamente a los *fluxes* internos (Mahadevan y Schilling, 2003).

Las aplicaciones del FBA han sido muy variadas. Entre ellas se encuentran la definición de diferentes fenotipos metabólicos (por ejemplo Edwards y Palsson, 1999); el cálculo de propiedades de las redes, incluyendo el requerimiento de energía asociada al crecimiento y el número de moléculas de ATP producidas por par de electrones donados al sistema de transporte electrónico (Famili *et al.*, 2003); el hallazgo de estrategias de ingeniería metabólica, buscando incrementar la producción de biomasa (por ejemplo Gheshlaghiet *al.*, 2007) o de diversos metabolitos secundarios (por ejemplo Förster *et al.*, 2003); la predicción de productos genéticos co-regulados y co-expresados (Schilling y Palsson, 2000); el balance de protones en redes a escala genómica (Reed y Palsson, 2003); la identificación de medios mínimos de cultivo (por ejemplo Becker y Palsson, 2005); el análisis de vacíos en la red metabólica (por ejemplo Oh *et al.*, 2007; Puchałka *et al.*, 2008); la identificación de conjuntos de reacciones acopladas y potenciales objetivos de medicamentos en el ser humano (Duarte *et al.*, 2007); el análisis *in silico* de deleciones genéticas (por ejemplo Beste *et al.*, 2007; Jamshidi y Palsson, 2007); y la evaluación de las predicciones de esencialidad de genes *in silico*, para validar las redes metabólicas (por ejemplo Dobson *et al.*, 2010; Mo *et al.*, 2009; Oberhardt *et al.*, 2008).

### 3.1. FUNCIONES OBJETIVO

El FBA se basa en la suposición de que las células han evolucionado para alcanzar cierto comportamiento óptimo, es decir, que las células redirigen sus *fluxes* metabólicos generándose una distribución óptima de *fluxes* debida a la presión evolutiva (Llaneras y Picó, 2008). Esto se representa matemáticamente por medio de la función objetivo usada en el problema de optimización que define al FBA, y la distribución particular de *fluxes* que se encuentre depende críticamente de la función objetivo usada para el análisis (Knorr *et al.*, 2007). Es por esto que una de las líneas investigativas que se ha desarrollado para mejorar el desempeño predictivo del FBA es la exploración de funciones objetivo (Nielsen y Jewett, 2007).

Diversos estudios han abordado el problema de evaluar cuál es la mejor función objetivo para el FBA, la mayoría de los cuales pertenecen a una de las siguientes clases: estudios de comparación entre funciones objetivo supuestas (sección 3.1.1), y estudios que aplican técnicas de optimización para generar funciones objetivo biológicas (sección 3.1.2). En ambos casos, la selección se hace por medio de comparación con datos experimentales (Feist y Palsson, 2010).

Un enfoque alternativo de exploración de funciones objetivo consiste en suponer que el comportamiento celular puede representarse convirtiendo la estimación de *fluxes* celulares en un problema de optimización con varios objetivos globales simultáneamente, con lo cual el problema pertenece a la categoría de optimización multi-objetivo, y no es posible encontrar una solución única; de hecho, lo que se puede definir es un conjunto de

soluciones para las cuales al mejorar uno de los objetivos al menos uno de los otros empeora. Este conjunto, que se denomina conjunto o frente de Pareto, es infinito. Esta idea se ha aplicado al metabolismo de *E. coli*, usando tres objetivos celulares: maximización de biomasa, maximización de ATP y minimización de *fluxes* intracelulares. Usando varios conjuntos de datos, se comparó el desempeño del integrante del frente de Pareto que se ajustara mejor a los datos experimentales en comparación con predicciones hechas previamente, obteniéndose distribuciones de *fluxes* similares (Sendín *et al.*, 2009). De igual manera, se ha aplicado la optimización multi-objetivo para integrar el FBA con el EBA (sección 3.2.3) para el análisis de las funciones metabólicas de hepatocitos (Nagrath *et al.*, 2007).

### 3.1.1. Comparación de funciones objetivo

En la mayoría de aplicaciones del FBA se ha asumido que la maximización del crecimiento, esto es, de producción de biomasa, es la función objetivo más apropiada para describir el comportamiento metabólico de una célula (por ejemplo Almaas *et al.*, 2004, Edwards *et al.*, 2001; Edwards y Palsson, 2000; Förster *et al.*, 2003; Ibarra *et al.*, 2002; Schilling *et al.*, 2002; Varma y Palsson, 1994), y en ciertas situaciones, la maximización de biomasa ha mostrado ser una función objetivo adecuada.

Entre los estudios que comparan el desempeño de diferentes funciones objetivo predeterminadas, los primeros estudios indicaban que la maximización de biomasa permitía modelar adecuadamente el comportamiento observado de las células (Pramanik y Keasling, 1997; Savinell y Palsson, 1992). Posteriormente, comenzaron a aparecer evidencias que indicaban que la maximización de biomasa podía no ser la mejor función objetivo en todos los casos (Ibarra *et al.*, 2002; Ramakrishna *et al.*, 2001). Un estudio posterior sobre el crecimiento aeróbico de *E. coli* en cultivo por lotes (*batch*) concluyó que la mejor función objetivo para esas condiciones era la minimización de tasa de producción de poder reductor en la célula (Knorr *et al.*, 2007). Otro trabajo publicado el mismo año, usando *E. coli* en un amplio rango de distintas condiciones de cultivo, llegó a la conclusión de que diferentes funciones objetivo son las más apropiadas dependiendo de la condición de crecimiento de la célula; incluso mostró que si bien la maximización de biomasa explica bien el comportamiento metabólico de *E. coli* en condiciones limitantes de carbono y energía, al no existir estas limitaciones dicho objetivo celular arroja unas predicciones muy pobres (Schuetz *et al.*, 2007). Otro trabajo en el que se evaluaron tres diferentes funciones objetivo para crecimiento aeróbico con limitación de carbono de *E. coli* tanto salvaje como modificada con plásmidos, mostró que para el segundo caso la mejor función objetivo es la maximización de gasto metabólico de moléculas de ATP (Ow *et al.*, 2009).

### 3.1.2. Generación de funciones objetivo a partir de datos

La primera propuesta de generación de funciones objetivo a partir de datos experimentales consistió en medir la distancia euclidiana entre el vector de *fluxes* arrojado por el *FBA*, con un vector experimental de *fluxes* proveniente, por ejemplo, de un experimento de marcación isotópica. Al comparar las distancias obtenidas por las diferentes funciones objetivo, se buscó determinar cuál representa mejor el verdadero comportamiento metabólico del microorganismo. Esta es la base del método ObjFind, que consiste en la búsqueda de los llamados coeficientes de importancia (*CoIs*), que representan la consistencia de la hipótesis de que un *flux* determinado es maximizado por el organismo como parte de su objetivo celular; esto se consiguió representando el objetivo celular como la maximización de una combinación lineal positiva de los *fluxes* que consumen recursos en la red, y hallando los pesos relativos que mejor ajustaban el desempeño de la función objetivo con los datos experimentales (dichos pesos son los *CoIs*). Al aplicar esta técnica a dos casos de estudio con *E. coli*, uno de crecimiento aeróbico y otro anaeróbico, se encontró que los *CoIs* eran similares en ambos casos, que la biomasa tenía el *CoI* más grande -sugiriendo que la biomasa era un descriptor del comportamiento celular, pero no el único-, y que las distribuciones de *fluxes* en la red no podían definirse de manera única basándose solamente en los *CoIs* identificados (Burgard y Maranas, 2003).

Usando la misma expresión de comparación entre funciones objetivo, que se mencionó en el párrafo anterior, se ha creado un método de generación de funciones objetivo a partir de la red estequiométrica y las restricciones disponibles, provenientes de los valores conocidos experimentalmente, de la irreversibilidad de las reacciones, de los valores máximos de velocidad que puedan tener las reacciones, etc. Este método, titulado *BOSS* (*Biological Objective Solution Search*), genera una función objetivo con la misma apariencia de una reacción metabólica; esto quiere decir que, a diferencia de la técnica usada para identificar los *CoIs*, se genera una reacción que actúa como función objetivo y en la que se permite que algunos *fluxes* tengan signo negativo, dando pasos importantes en la búsqueda de conseguir una asignación de significado biológico a la función objetivo. Al aplicar *BOSS* a un caso de estudio de *S. cerevisiae* con glucosa como sustrato limitante, el método arrojó una función objetivo muy similar a la maximización de biomasa, apoyando la idea de que uso de la maximización de biomasa es un objetivo razonable para muchos casos. No obstante, si el mejor objetivo hace parte de la matriz estequiométrica, es necesario hacer cálculos extra para obtener una respuesta con *BOSS* (Gianchandani *et al.*, 2008).

## 3.2. VARIACIONES DEL FBA

Con el paso de los años, se han sugerido múltiples estrategias de análisis, estimación y predicción de redes metabólicas, todas ellas basadas en el FBA. Entre ellas se incluyen: enfoques tendientes al uso del FBA para modelar procesos dinámicos, la incorporación de más restricciones sobre los valores de los *fluxes* a partir de consideraciones regulatorias y termodinámicas, el análisis de rangos de valores posibles de los *fluxes* en una condición determinada (en contraste con la distribución única que se obtiene en el FBA tradicional) y sugerencias alternativas de funciones objetivo para el análisis de células que han sufrido deleciones genéticas. A continuación se presentan estas variaciones sugeridas del FBA.

### 3.2.1. FBA dinámico

Una de las suposiciones básicas del modelado estequiométrico, es decir, que la red metabólica se encuentra en estado estable, evita que el FBA pueda ser usado directamente para modelar procesos dinámicos. Sin embargo, discretizando el tiempo es posible modelar un proceso dinámico apoyándose en el FBA, el cual se usa entonces para averiguar la distribución de *fluxes* en cada paso del tiempo, mientras que por medio de expresiones dinámicas se representan variaciones en diversos valores de *fluxes* de intercambio. Esta es la base del llamado FBA dinámico (dFBA – *dynamic flux balance analysis*) (Mahadevan *et al.*, 2002).

Otra alternativa propuesta para aplicar el FBA en el modelado de procesos metabólicos dinámicos involucra la integración de expresiones cinéticas aproximadas, basadas en la representación lin-log (sección 1.2.3.4), junto con la predicción de los *fluxes* obtenida por el FBA. Este enfoque fue aplicado a un modelo de metabolismo central de *S. cerevisiae*, obteniéndose aproximaciones interesantes al comportamiento, aunque no con un nivel de precisión alto (Smallbone *et al.*, 2007).

Se han llevado a cabo gran cantidad de trabajos de aplicación del dFBA; igualmente, han aparecido técnicas que combinan el dFBA con información regulatoria (sección 3.2.2).

### 3.2.2 Incorporación de información regulatoria

Algunas de las alteraciones propuestas del FBA tienen como objetivo la integración de información regulatoria en el modelo metabólico al cual se le aplica el FBA, con miras a incrementar la fidelidad del modelado y la cantidad de procesos celulares representados.

Los primeros intentos de integración de información regulatoria datan de principios del siglo XXI, y consistieron en la introducción de restricciones regulatorias en los modelos de

FBA usando operadores lógicos booleanos, siendo denominada la técnica FBA regulatorio (rFBA – *regulatory flux balance analysis*) (Covert *et al.*, 2001). Estas restricciones regulatorias representan limitaciones sobre los *fluxes* que surgen como producto de un entorno específico, en contraste con las restricciones fisicoquímicas que representan limitaciones más fundamentales (Raman y Chandra, 2009). El estado metabólico en un momento particular se representa por los *fluxes* hallados por un FBA, mientras que el estado regulatorio se representa por una función booleana que indica la expresión de cada gen. El rFBA involucra la predicción en instantes sucesivos de los estados metabólicos y regulatorios, los cuales pueden alterarse mutuamente, modelando de esa manera la evolución del proceso (Covert *et al.*, 2001).

Otra manera de unificar el FBA e información regulatoria es el llamado SR-FBA (*steady state regulatory flux balance analysis*) (Shlomi *et al.*, 2007). En este enfoque, la determinación de los estados metabólico y regulatorio involucra la solución de un problema de programación mixta-entera-lineal (MILP), formulado por la traducción de las restricciones regulatorias booleanas y el mapeo entre genes y reacciones a ecuaciones lineales (Raman y Chandra, 2009). Aplicando la técnica para estudiar el metabolismo de *E. coli*, se encontró que si bien la mayoría de las restricciones metabólicas determinaban el estado de actividad de los *fluxes* de la mayoría de los genes, el papel de la regulación transcripcional era apreciable, determinando los estados de casi el 20% de los genes (Shlomi *et al.*, 2007).

Una estrategia que combina la integración de información regulatoria con el dFBA es el llamado FBA integrado dinámico (*idFBA – integrated dynamic Flux Balance Analysis*), en el cual se crean modelos de la red metabólica, la red regulatoria y la red de señales, así como la interacción entre ellas, por medio de las sustancias que ocasionan cambios o efectos en las diferentes redes. Para manejar la diferencia de magnitudes de la escala de tiempo de múltiples procesos modelados por estas redes, la dinámica de los procesos se clasifica en ‘rápido’ y ‘lento’, de acuerdo con el tiempo que tarda el proceso en alcanzar un nuevo estado estable tras una perturbación. El modelado del sistema se hace por medio de una secuencia de estados a medida que el tiempo cambia; para un tiempo particular, se calcula la distribución de *fluxes* de acuerdo con el FBA aplicado a la red metabólica existente en el momento, y entre diferentes instantes de tiempo, pueden presentarse alteraciones en la red metabólica actual producto de respuestas regulatorias o de señales, lo cual cambia la estimación de las distribuciones de *fluxes* (Min Lee *et al.*, 2008).

### **3.2.3. Incorporación de restricciones termodinámicas**

El FBA tiene en cuenta la conservación de masa, pero no considera el potencial químico asociado con los *fluxes*, los cuales se encuentran en estado estacionario pero no en equilibrio (Nagrath *et al.*, 2007). Para incluir información sobre las restricciones

energéticas sobre los fluxes, surgió una alteración del FBA conocida como análisis de balance de energía (EBA – *Energy Balance Analysis*), que consiste en incluir restricciones adicionales al sistema que provienen de consideraciones termodinámicas. Dichas restricciones son análogas a las que se presentan al aplicar la ley de voltaje de Kirchoff a un circuito eléctrico (Beard *et al.*, 2002).

En una red metabólica, los lazos para balance de energía no pueden obtenerse a partir de un simple análisis de los lazos topológicos de la red, como se hace cuando se analizan circuitos, debido a que las redes metabólicas están compuestas por reacciones que involucran múltiples especies (Balabanian y Bickart, 1981).

En el EBA, las ecuaciones de balance de energía se obtienen a partir de un análisis de la estequiometría de la red. Primero, se reduce la matriz estequiométrica  $S$  removiendo los fluxes de intercambio con el entorno y combinando los fluxes redundantes, obteniéndose una matriz  $S_{rd}$ . Posteriormente, se determina el espacio nulo de la matriz  $S_{rd}$ , el cual puede hallarse por medio de una descomposición de valor singular (ecuación [20]).

$$S_{rd} = AVB^T \quad [20]$$

Donde  $V$  tiene la forma mostrada en la ecuación [21].

$$V = \begin{bmatrix} l_1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & l_m & 0 & \cdots & 0 \end{bmatrix} \quad [21]$$

Los valores diagonales de las primeras  $m$  columnas de la matriz  $V$  son los valores singulares, y las demás columnas están compuestas de ceros. Si  $n_{Srd}$  es el número de columnas de  $S_{rd}$  y  $r$  es el rango de  $S_{rd}$ , entonces las columnas desde  $(r + 1)$  hasta  $n_{Srd}$  de la matriz  $B$  conforman el espacio nulo  $(n_{Srd} - r)$ -dimensional de  $S_{rd}$  (Qian y Beard, 2005).

Almacenando el espacio nulo de  $S_{rd}$  en una matriz  $K$  se obtiene la ecuación [22].

$$K = \begin{bmatrix} b_{1,r+1} & \cdots & b_{1,nSrd} \\ \vdots & \ddots & \vdots \\ b_{nSrd,r+1} & \cdots & b_{nSrd,nSrd} \end{bmatrix} \quad [22]$$

Cada columna de  $K$  puede multiplicarse por las  $n_{Srd}$  reacciones, y se obtiene una ecuación de reacción perfectamente balanceada. A continuación, se determina el vector  $\Delta\mu$  que contiene las  $n_{Srd}$  diferencias de potencial químico asociadas a cada reacción. Como cada

columna de la matriz  $K$  provee pesos para balancear globalmente las reacciones químicas, entonces las soluciones a la ecuación [23]:

$$K\Delta\mu = 0 \quad [23]$$

Balancen la energía potencial global de la red (Qian y Beard, 2005). Por otra parte, la segunda ley de la termodinámica toma la forma de una restricción desigualdad para cada *flux*, como se muestra en la ecuación [24] (Qian, 2001a, b).

$$v_i \cdot \Delta\mu_i < 0 \quad [24]$$

Las restricciones de las anteriores dos ecuaciones se agregan a las restricciones del FBA, y el problema de optimización no lineal que se obtiene es el problema de EBA. Este análisis se aplicó con buenos resultados a un modelo metabólico de *Escherichia coli*, compuesto de 953 *fluxes* y 536 metabolitos (Beard *et al.*, 2002).

#### 3.2.4. Análisis de variabilidad de *fluxes* (FVA)

Un inconveniente que presenta el FBA es que las soluciones pueden no ser únicas. La existencia de múltiples óptimos alternativos representa situaciones en que diferentes distribuciones de *fluxes* pueden ser usadas por el sistema para alcanzar exactamente el mismo valor de la función objetivo. Para analizar el efecto de la multiplicidad de soluciones, se desarrolló un algoritmo de programación mixta entera lineal para encontrar las soluciones múltiples y se aplicó en un modelo metabólico pequeño (Lee *et al.*, 2000), pero después se hizo evidente que usar esta técnica para hallar los múltiples óptimos podía volverse computacionalmente inmanejable para ciertas condiciones en un modelo a escala genómica (Reed *et al.*, 2003).

Para tratar con el problema de la posible existencia de múltiples óptimos sin tener que determinarlos a todos, de forma exacta, se creó una técnica bautizada análisis de variabilidad de *fluxes* (FVA – *flux variability analysis*), cuyo objetivo es determinar cuál es el rango de valores que puede tomar cada uno de los *fluxes* de la red metabólica en los múltiples óptimos que puedan existir al resolver el FBA con una función objetivo particular (Mahadevan y Schilling, 2003). Esto se logra resolviendo primero el problema FBA, mostrado en la ecuación [25].

$$\begin{aligned} \max Z &= f(\vec{v}) \\ \text{sujeto a } &\begin{cases} S \cdot \vec{v} = \vec{0} \\ \vec{v}_i \leq \vec{v} \leq \vec{v}_s \\ v_j = v_{exp} \end{cases} \end{aligned} \quad [25]$$

Denotando al mejor valor posible de la función objetivo por  $Z^*$ , con lo que el problema de encontrar los valores máximos y mínimo de cada uno de los *fluxes* consiste en resolver para cada *flux* dos problemas de programación lineal, que se muestran en la ecuación [26] (Mahadevan y Schilling, 2003):

$$\begin{array}{l} \max/\min v_j \\ \text{sujeto a } \left\{ \begin{array}{l} Z = Z^* \\ S \cdot \vec{v} = \vec{0} \\ \vec{v}_i \leq \vec{v} \leq \vec{v}_s \\ v_j = v_{exp} \end{array} \right. \end{array} \quad [26]$$

Un análisis que también resulta interesante es evaluar la variabilidad de los *fluxes* por encima de cierto umbral de cercanía con el óptimo de la función objetivo usada en el FBA; para esta finalidad, se alteran los problemas de programación lineal como se indica en la ecuación [27].

$$\begin{array}{l} \max/\min v_j \\ \text{sujeto a } \left\{ \begin{array}{l} Z \geq \gamma Z^* \\ S \cdot \vec{v} = \vec{0} \\ \vec{v}_i \leq \vec{v} \leq \vec{v}_s \\ v_j = v_{exp} \end{array} \right. \end{array} \quad [27]$$

Donde  $\gamma$  es un parámetro positivo, entre cero y uno, que representa el margen de cercanía deseado con el valor óptimo. De este modo, se relaja la restricción de valor óptimo obtenido en el FBA y se puede explorar la variación de los *fluxes* para distintos cambios en el valor óptimo del objetivo supuesto (Gudmundsson y Thiele, 2010).

Si existen  $n$  reacciones a las cuales se les quiere evaluar la variabilidad, el FVA requiere la solución de  $2n$  problemas de programación lineal. Esto ha conducido a la proposición de mejoras en el algoritmo para efectuar el FVA, con resultados positivos en cuanto a tiempo de cómputo (Gudmundsson y Thiele, 2010).

### 3.2.5. MOMA - *Minimization of metabolic adjustment*

La minimización de ajuste metabólico (MOMA – *minimization of metabolic adjustment*) es una técnica propuesta para determinar las distribuciones de *fluxes* intracelulares en microorganismos que han sido alterados genéticamente. La premisa tras la técnica es que las células que han sufrido delecciones genéticas no han tenido tiempo, en términos

evolutivos, de adaptar sus *fluxes* de una manera óptima respecto a algún objetivo (Segrè *et al.*, 2002).

En la técnica MOMA, se asume que la distribución de *fluxes* de la célula modificada genéticamente corresponde a aquella que haya sufrido una alteración mínima respecto a la distribución que tenía la célula en estado salvaje (es decir, sin la alteración genética). De modo que hallar el vector de *fluxes* intracelulares para una célula modificada usando MOMA implica resolver dos problemas de optimización: primero se lleva a cabo un FBA, usando la matriz estequiométrica de la célula en estado salvaje, y luego se encuentra el vector que presenta una distancia euclidiana mínima hasta la distribución de *fluxes* en estado salvaje, lo que requiere resolver el problema de programación cuadrática mostrado en la ecuación [28].

$$\begin{aligned} \min E &= \|\vec{v} - \vec{v}_{sal}\| \\ \text{s.a.} &\begin{cases} S_{alt} \cdot \vec{v} = \vec{0} \\ \vec{v}_{inf} \leq \vec{v} \leq \vec{v}_{sup} \\ v_j = v_{exp} \end{cases} \end{aligned} \quad [28]$$

En la ecuación [28],  $S_{alt}$  es la matriz estequiométrica de la célula alterada genéticamente, y  $\vec{v}_{sal}$  es la distribución de *fluxes* de la célula hallada en estado salvaje mediante la aplicación de FBA. Es evidente que la distribución de *fluxes* depende también de la función objetivo usada para determinar el vector de *fluxes* en estado salvaje. Usando la técnica MOMA para predecir las distribuciones para distintos casos de deleciones genéticas de *E. coli* (utilizando maximización de crecimiento como objetivo del FBA previo), se mostró que con el MOMA se obtenía una correlación mucho mayor entre datos experimentales y valores estimados, respecto a las predicciones arrojadas por el FBA (usando maximización de biomasa como función objetivo) (Segrè *et al.*, 2002).

### 3.2.6. ROOM - *Regulatory on/off minimization*

La minimización regulatoria *on/off* (ROOM - *Regulatory on/off minimization*) es otra técnica diseñada específicamente para modelar el metabolismo de células que han sufrido una alteración genética. Parte de una premisa similar a la que condujo a la creación de la técnica MOMA, usando una función objetivo diferente para representar el ajuste del metabolismo tras la deleción genética (Shlomi *et al.*, 2005).

En la ROOM, se asume que la célula intenta minimizar el número de *fluxes* que sufren cambios significativos, tras haber sufrido una alteración en su red metabólica, respecto a la distribución que tenía en estado salvaje (Shlomi *et al.*, 2005). El resultado de la técnica, además de depender de la distribución de *fluxes* supuesta para el estado salvaje (que a su

vez, depende de la función objetivo supuesta en el FBA realizado), también es función de la tolerancia elegida para definir el “cambio significativo” en el valor de un *flux*.

La aplicación de la ROOM demostró ser mejor que la MOMA y el FBA en la predicción de distribuciones de *fluxes* poco tiempo después de la aparición de la alteración genética, mientras que las otras dos técnicas funcionan mejor para la estimación de los *fluxes* cuando ha transcurrido más tiempo y ya empieza a estabilizarse el metabolismo (Shlomi *et al.*, 2005).

Recientemente se aplicó la idea de minimizar el número de cambios significativos en los valores de los *fluxes* para modelar dinámicamente una red, conduciendo a una integración de las técnicas ROOM y dFBA que fue llamada R-DFBA. La técnica fue aplicada a un ciclo metabólico pequeño con buenos resultados respecto a otras técnicas existentes (Kleessen y Nikoloski, 2012).

## 4. MÉTODOS, MODELOS Y DATOS

### 4.1. HIPÓTESIS DE TRABAJO

De acuerdo con el marco referencial presentado, es posible ver que se requiere continuar avanzando en la mejora de la calidad de los modelos metabólicos y en la representación del objetivo celular dependiendo de las condiciones metabólicas y ambientales, de manera que se pueda lograr un mejor modelado predictivo y mayor comprensión del metabolismo celular a través del FBA.

El presente trabajo de doctorado se enfocó al estudio de una nueva propuesta de representación matemática del objetivo de la célula; teniendo en cuenta la multiplicidad de organelos celulares que existe en las células eucariotas, se plantea que aquéllos, representados en los modelos metabólicos como compartimentos, pueden tener objetivos distintos, y que su combinación puede conformar la mejor representación del objetivo celular. Por consiguiente, la siguiente es la hipótesis de trabajo:

El objetivo metabólico global de las células eucariotas puede modelarse adecuadamente como una combinación lineal de diferentes objetivos de los compartimentos celulares.

El tipo de funciones objetivo propuestas y evaluadas en el presente trabajo (esto es, provenientes de una combinación de objetivos de los compartimentos) es válido para modelar células eucariotas, pues las procariotas no poseen múltiples compartimentos. Por consiguiente, se eligió usar un modelo metabólico y datos experimentales sobre *S. cerevisiae*, pues se ha posicionado como organismo modelo eucariótico (ver capítulo 2). Aunque la mayoría de evaluaciones del desempeño de la técnica FBA, y otras afines, se han llevado a cabo usando modelos metabólicos de tamaño moderado, en el presente trabajo se decidió utilizar un modelo a escala genómica, utilizando de esta manera gran cantidad del conocimiento que actualmente se tiene sobre *S. cerevisiae*.

Por esta razón, el objetivo del trabajo fue determinar cuáles funciones objetivo, provenientes de una combinación lineal de objetivos de los compartimentos celulares, permiten obtener mejores predicciones de los *fluxes* al ser usadas en el FBA. Para cada uno de los compartimentos considerados del modelo elegido (sección 4.2), se propusieron una serie de objetivos (secciones 4.5 y 4.6), y se compararon las predicciones obtenidas usando cada una de las combinaciones lineales posibles entre los objetivos de los distintos compartimentos por medio de comparación con valores experimentales (sección 4.3).

Para evaluar el error cometido por las predicciones se cuantificó el error en la predicción de la tasa de crecimiento, y el error en la predicción de los *fluxes* de intercambio conocidos que no fueron usados como datos de entrada (sección 4.4). De acuerdo con esos errores, se identificó qué combinaciones de objetivos compartimentales eran las más adecuadas para actuar como función objetivo en el FBA para cada conjunto de condiciones experimentales.

El nivel de adecuación de las funciones objetivo se definió de acuerdo con la predicción de producción de biomasa; sin embargo, como ya se mencionó, se calculó también el error en la predicción de los *fluxes* de intercambio, y se tuvieron en cuenta ambos tipos de error en el análisis del desempeño de las diferentes funciones objetivo.

## **4.2. OBJETIVOS**

### **4.2.1. OBJETIVO GENERAL**

Estudiar el desempeño del Análisis de Balance de *Fluxes* (FBA) aplicado a células eucariotas, utilizando como organismo modelo a *Saccharomyces cerevisiae* y como objetivo metabólico celular una función que sea una combinación lineal de diferentes objetivos de los compartimentos celulares.

### **4.2.2. OBJETIVOS ESPECÍFICOS**

- Diseñar la representación del objetivo metabólico celular usado en el Análisis de Balance de *Fluxes* (FBA), de manera que sea una combinación lineal de objetivos de los compartimentos celulares.
- Crear un algoritmo matemático que permita comparar los errores cometidos en las estimaciones del Análisis de Balance de *Fluxes* (FBA), al utilizar las funciones objetivo generadas a partir de las combinaciones lineales de los objetivos elegidos de los compartimentos celulares existentes.
- Analizar, para diferentes condiciones ambientales y de crecimiento de las células, las funciones objetivo (conformadas por combinaciones de objetivos de los compartimentos celulares) que permiten obtener estimaciones de los *fluxes* metabólicos más ajustados a sus valores reales, utilizando datos experimentales para este fin.

### 4.3. SELECCIÓN DEL MODELO METABÓLICO

Debido al tipo de funciones objetivo propuestas, el modelo metabólico debía ser uno que representara múltiples compartimentos celulares, por lo cual los modelos iFF708, iLL672 e iIN800 quedaban descartados. El modelo iND750 y la primera versión de la red consensuada tampoco se tuvieron en cuenta pues consistían en versiones primitivas de otros modelos disponibles.

Entre los tres modelos genómicos compartimentalizados restantes, se eligió al iMM904 como modelo metabólico de *S. cerevisiae* para el presente trabajo. La razón principal es que se encontró que dicho modelo presenta una mayor estabilidad para ser aplicado en el FBA que los otros dos modelos posibles, es decir, el modelo iAZ900 y la cuarta versión de la red consensuada.

Como ya se mencionó en la sección 2.2.5, el modelo iMM904 contiene 1577 reacciones (incluyendo las de transporte) y 1228 metabolitos, distribuidos en ocho localizaciones (7 de ellas compartimentos celulares): espacio extracelular, citosol, mitocondria, peroxisoma, núcleo, retículo endoplasmático, aparato de Golgi y vacuola (Mo *et al.*, 2009).

### 4.4. SELECCIÓN DE DATOS EXPERIMENTALES

Se tomaron de la literatura múltiples conjuntos de datos sobre experimentos de crecimiento de *S. cerevisiae*, conformando un amplio rango de condiciones para las estimaciones, de modo que las conclusiones obtenidas resultaran tan versátiles como fuera posible. Posteriormente, a los conjuntos de datos, se les realizó un análisis del ajuste del balance de masa (carbono), para evitar considerar datos experimentales con mucho error o mucha incertidumbre en la medición. Los conjuntos de datos que presentaban un error superior al 10% en el balance másico fueron desechados.

La **Tabla 5** presenta las fuentes y los valores de los datos experimentales usados para el presente trabajo. Se muestra también las condiciones de cultivo respecto a tipo de sustrato y disponibilidad de oxígeno, así como los *fluxes* de intercambio medidos en cada una de las referencias. Todos los conjuntos de datos provienen de cultivos continuos, excepto el mencionado en la décima fila, que proviene de un cultivo por lotes en fase de crecimiento exponencial (Heyland *et al.*, 2009). Los procesos metabólicos se encuentran balanceados en células que crecen exponencialmente en ausencia de perturbaciones, conduciendo a la homeostasis (Klipp, 2007), así que estos datos también fueron usados para la evaluación del desempeño de las diferentes funciones objetivo en el FBA.

**Tabla 5.** Conjuntos de datos experimentales usados para evaluar la precisión de las predicciones de las diferentes funciones objetivo.

Sustrato	Condición respiratoria	Fluxes de intercambio medidos	Cantidad de datos experimentales	Referencia
Glucosa	Respiro-fermentativa	Etanol, glicerol, acetato, succinato, acetaldehído, piruvato	1	Çakır <i>et al.</i> , 2007
Glucosa	Anaeróbica	Etanol, glicerol, acetato, succinato, piruvato, dióxido de carbono	4	Nissen <i>et al.</i> , 1997
Glucosa	Aeróbica	Etanol, glicerol, acetato, succinato, piruvato	1	Gombert <i>et al.</i> , 2001
Glucosa, maltosa, etanol, acetato	Aeróbica	Oxígeno, dióxido de carbono	4	Daran-Lapujade <i>et al.</i> , 2004
Glucosa	Aeróbica, respiro-fermentativa	Etanol, glicerol, acetato, dióxido de carbono	3	Frick y Wittmann, 2005
Glucosa más etanol	Aeróbica	Oxígeno, dióxido de carbono	1	van Winden <i>et al.</i> , 2005
Galactosa	Aeróbica	Etanol, glicerol, acetato, piruvato, dióxido de carbono	1	Ostergaard <i>et al.</i> , 2000
Glucosa	Respiro-fermentativa, anaeróbica	Etanol, glicerol, acetato, oxígeno	8	Jouhten <i>et al.</i> , 2008
Glucosa	Respiro-fermentativa, Anaeróbica	Etanol, glicerol, acetato, oxígeno	5	Wiebe <i>et al.</i> , 2008
Glucosa	Respiro-fermentativa	Etanol, glicerol, acetato, dióxido de carbono	8	Heyland <i>et al.</i> , 2009
Glucosa	Aeróbica	Etanol, glicerol, acetato, piruvato, oxígeno, dióxido de carbono	10	van Hoek <i>et al.</i> , 1998

En la sección Anexos, el **Anexo A** muestra una descripción más detallada de los diferentes conjuntos de datos experimentales usados en el presente trabajo.

#### 4.5. EVALUACIÓN DEL ERROR EN LAS PREDICCIONES

Los conjuntos de datos experimentales fueron clasificados en distintas categorías, para explorar la posible relación entre las condiciones ambientales y las funciones objetivo que mejor modelan el comportamiento celular en el FBA, así como la posible relación entre la tasa de crecimiento celular y las combinaciones de objetivos que se desempeñan mejor, en cuanto a la exactitud de las predicciones obtenidas con estas funciones al ser usadas

como objetivos del FBA. Como se indicó en la sección 4.1, la evaluación del potencial predictivo de cada posible combinación de objetivos fue hecha calculando, por una parte, el error en la predicción de la tasa específica de crecimiento, y por otra parte, el error en la predicción de los *fluxes* de intercambio para los cuales se conocía el valor experimental.

El error en la predicción de la tasa específica de crecimiento, o **Error biomasa**, fue calculado como porcentaje de error relativo, como se muestra en la ecuación [29].

$$\text{Error biomasa} = \frac{(\mu_{\text{estimado}} - \mu_{\text{medido}})}{\mu_{\text{medido}}} * 100 \quad [29]$$

En el modelo iMM904, como es habitual en los modelos metabólicos a escala genómica, los coeficientes de la ecuación de biomasa se ajustan para que las unidades obtenidas para esta reacción en particular coincidan con las de tasa específica de crecimiento, es decir,  $h^{-1}$ . De esta manera, no fue necesario hacer ninguna conversión de unidades a las estimaciones obtenidas del FBA para calcular  $\mu_{\text{estimado}}$ .

Como se verá en el capítulo 5, en la primera versión del trabajo se calculó este error en valor absoluto, pero para la versión final se decidió conservar el signo de la diferencia  $\mu_{\text{estimado}} - \mu_{\text{medido}}$ , de modo que éste indicara si el FBA había sobreestimado o subestimado el crecimiento celular con la función objetivo probada.

El error en la predicción de los *fluxes* de intercambio, o **Error fluxes**, consiste en la distancia euclidiana entre los valores experimentales de los *fluxes* de intercambio, y los valores estimados por el FBA de estos *fluxes*, como se indica en la ecuación [30].

$$\text{Error fluxes} = \|\vec{v}_{im,experimental} - \vec{v}_{im,medido}\| \quad [30]$$

Donde  $\vec{v}_{im,experimental}$  es el vector de valores experimentales de los *fluxes* de intercambio medidos, y  $\vec{v}_{im,medido}$  es el vector de valores estimados para los mismos *fluxes* de intercambio.

Los dos tipos de errores no fueron integrados en la misma medición fundamentalmente porque las unidades de las magnitudes comparadas en cada uno de ellos son diferentes (la tasa específica de crecimiento se expresó en  $h^{-1}$ , mientras que los *fluxes* fueron representados en  $mmol/(g \text{ biomasa seca} * h)$ ), y también para poder analizar independientemente el efecto sobre la predicción del crecimiento y otros metabolitos secundarios, teniendo presente que no en todos los conjuntos de datos experimentales los *fluxes* de intercambio medidos eran los mismos. En cualquier caso, se tuvieron en cuenta ambos tipos de errores para el análisis de las diferentes funciones objetivo exploradas.

Como el objetivo era comparar el potencial predictivo de las distintas funciones objetivo, como datos de entrada se usaron únicamente los valores de ingesta de sustrato y, si estaba disponible, de intercambio de oxígeno. De esta manera, los demás *fluxes* de intercambio se podían usar como medida del error, y se probaba el desempeño de cada función sin exceso de información, enfocándose de esta manera en la obtención de una buena representación matemática del objetivo celular.

Las simulaciones y comparaciones fueron programadas en **Matlab**<sup>®</sup> (The Mathworks, Inc.), utilizando el paquete **COBRA toolbox** (Schellenberger *et al.*, 2011) para leer el modelo metabólico y para resolver el problema de optimización interno (esto es, aplicar los FBAs), y una búsqueda exhaustiva para resolver el problema de optimización externo. Como *solver* de programación lineal se usó el software libre **glpk**, usando **glpkmex** como enlace con Matlab. Una descripción más detallada de la ejecución del algoritmo se encuentra en la tesis de Maestría en Ingeniería de Sistemas de César Augusto Vargas García, pues dicha tesis consiste en la implementación computacional del problema de estudio del presente trabajo, descrito en las secciones 4.6. y 4.7. (Vargas García, 2012).

#### **4.6. COMPARTIMENTOS Y OBJETIVOS – PRIMERA VERSIÓN**

Inicialmente, se evaluó el desempeño de las combinaciones lineales de los objetivos de los compartimentos, pero sin usar pesos relativos en la combinación. Es decir, se probó el desempeño de las funciones objetivo surgidas de combinar linealmente los objetivos de los compartimentos, teniendo todos un uno como peso relativo, obteniéndose el problema de optimización binivel que se muestra en la ecuación [31], pero con una diferencia: en vez de resolver el problema de optimización externo (es decir, seleccionar únicamente la función  $Z$  que ocasiona un menor error en la estimación de la producción de biomasa), se tomaron las tres a cinco mejores funciones  $Z$ ; es decir, las funciones que daban mejores predicciones de la tasa específica de crecimiento al ser usadas como objetivos del FBA. Es decir, sí se determinaron cuáles eran las funciones objetivo cuyas estimaciones arrojaban menor error relativo porcentual, pero no se dio como respuesta la función con menor error (que sería la solución normal del problema de optimización externo), sino que se presenta el conjunto de (tres a cinco) funciones que menor error arrojaron. El número específico de funciones (tres, cuatro o cinco) dependió de los resultados de cada categoría particular. En algunas categorías, la diferencia de potencial predictivo entre las funciones no justificó mostrar las cinco mejores funciones, y por consiguiente se muestra una cantidad menor.

$$\begin{aligned}
& \min \frac{\sum_{k=1}^m \left| \frac{\mu_{k,estimado} - \mu_{k,medido}}{\mu_{k,medido}} * 100 \right|}{\sum_{k=1}^m \left| \frac{\mu_{k,estimado} - \mu_{k,medido}}{\mu_{k,medido}} * 100 \right|} \\
& \text{sujeto a } \begin{cases} \max Z = \sum_{l=1}^n F_{j,l}(\vec{v}) \\ \text{sujeto a } \begin{cases} S * \vec{v} = 0 \\ \vec{v}_l \leq \vec{v} \leq \vec{v}_u \end{cases} \end{cases} \quad [31]
\end{aligned}$$

En la ecuación [31],  $\mu_{k,medido}$  es la tasa de crecimiento experimental,  $\mu_{k,estimado}$  es la tasa de crecimiento predicha por el FBA,  $n$  es el número de compartimentos considerados y  $F_{j,l}(\vec{v})$  es uno de los objetivos propuestos para el  $l$ -ésimo compartimento.

Dos de los siete compartimentos del modelo iMM904, a saber, aparato de Golgi y vacuola, presentan un número muy bajo de reacciones (seis y tres reacciones, respectivamente) y presentan varios problemas de conectividad (y esto sucede con los 3 modelos totalmente compartimentalizados; el conocimiento respecto de estos compartimentos aún es insuficiente), y por lo anterior no fueron considerados como posibles contribuyentes al objetivo celular.

Evaluaciones preliminares sobre los demás compartimentos celulares en el modelo iMM904 condujeron a descartar otros dos compartimentos (retículo endoplasmático y, de manera inesperada, núcleo), pues se hizo evidente que los objetivos que se les asignaban no alteraban las predicciones obtenidas en el FBA. La **Tabla 6** muestra los tres compartimentos considerados en la primera versión, junto con los posibles objetivos considerados para cada uno de ellos.

Los posibles objetivos seleccionados para cada compartimento están relacionados principalmente con el consumo y la producción de poder reductor y ATP, debido a la enorme importancia de estos cofactores en el comportamiento celular. Otros objetivos considerados están asociados con la producción de biomasa (debido a que la relevancia de este objetivo en el modelado con FBA está bien establecida), y con la producción de diversos metabolitos secundarios.

**Tabla 6.** Posibles objetivos de los compartimentos celulares evaluados con la primera versión de las combinaciones de objetivos compartimentales.

Citosol		Mitocondria	Peroxisoma
max (producción biomasa)	max (producción NADH+NADPH)	max (producción ATP)	max (producción de ácidos grasos)
min (producción biomasa)	min (producción NADH+NADPH)	max (producción NADH+NADPH)	max (consumo ATP)
max (producción acetato)	max (consumo NADH+NADPH)	max (consumo NADH+NADPH)	min (consumo ATP)
max (producción CO <sub>2</sub> )	min (consumo NADH+NADPH)	max (transporte de ATP hacia citosol)	max (producción NADH+NADPH)
max (producción etanol)	max (producción ATP)	Ninguno	min (consumo NADH+NADPH)
max (producción glicerol)	max (consumo ATP)		Ninguno
max (producción succinato)	min (consumo ATP)		
min (producción NADH)	Ninguno		

Cada objetivo celular presentado está compuesto por las reacciones en que participa el metabolito correspondiente. Por ejemplo, para representar la maximización de producción de ATP en el citosol, se determinaron todas las reacciones del citosol en que se genera ATP en el modelo iMM904, y se sumaron teniendo en cuenta los coeficientes estequiométricos de dichas reacciones; el algoritmo que ejecuta el FBA en el COBRA toolbox maximiza por defecto la función objetivo (Schellenberger *et al.*, 2011). Para ejecutar las funciones que representaban minimizaciones, se realizaba el mismo procedimiento de identificación y suma de reacciones, pero se le asignaba un signo negativo a cada sumando en el objetivo, lográndose minimizar dicho objetivo dado que el algoritmo global está escrito para maximizar. Y adicionalmente, usando de esta manera los signos es sencillo integrar en la misma función objetivos que se deseaba maximizar junto con objetivos que se buscaba minimizar.

Vale la pena destacar que en cada uno de los compartimentos, uno de los posibles objetivos es 'Ninguno', permitiendo de esta manera probar también combinaciones de

dos objetivos de compartimentos, e incluso cada uno de los objetivos propuestos independientemente, quedando incluidas en el análisis gran parte de las funciones objetivo propuestas hasta el momento, como maximización de crecimiento y maximización de producción de ATP.

#### 4.7. COMPARTIMENTOS Y OBJETIVOS – VERSIÓN FINAL

Los resultados obtenidos en la primera versión de la comparación de funciones objetivo (sección 5.1) condujeron a la asignación de un compartimento virtual, llamado ‘global’, a las funciones objetivo relacionadas con la producción de biomasa, debido a la evidente influencia del objetivo ‘maximización de producción de biomasa’ en las combinaciones de funciones con las que se obtuvieron mejores resultados. Igualmente, para esta versión final se incluyeron como posibles objetivos en el citosol las minimizaciones de producción de los diferentes metabolitos secundarios, pues algunas pruebas indicaron que estos objetivos podían mejorar las predicciones en ciertos casos.

Además de lo mencionado en el párrafo anterior, se incluyeron además pesos relativos en la combinación lineal de objetivos compartimentales, de modo que la comparación de las diferentes funciones objetivo evaluadas se puede expresar matemáticamente según la ecuación [32] (de nuevo, no se determinó únicamente la mejor función objetivo, sino que se comparó el desempeño de las mejores funciones).

$$\min \frac{\sum_{k=1}^m \left| \frac{\mu_{k,estimado} - \mu_{k,medido}}{\mu_{k,medido}} * 100 \right|}{m}$$

$$\text{sujeto a } \begin{cases} \max Z_{i,j} = \vec{w}_i \vec{F}_j^T \\ \text{sujeto a } \begin{cases} S * \vec{v} = 0 \\ \vec{v}_l \leq \vec{v} \leq \vec{v}_u \end{cases} \end{cases} \quad [32]$$

En la ecuación [32],  $\mu_{k,medido}$  es la tasa específica de crecimiento medida en el  $k$ -ésimo conjunto de datos experimentales de la categoría;  $\mu_{k,estimado}$  es la tasa específica de crecimiento predicha por el problema de optimización interno (es decir, por el FBA realizado usando  $Z_{i,j}$  como función objetivo) en el  $k$ -ésimo conjunto de datos de la categoría; y  $m$  es el número de conjuntos de datos experimentales que pertenecen a la categoría (de condiciones ambientales/de crecimiento) analizada.

Como puede verse, este problema de optimización bi-nivel tiene como objetivo externo la minimización de Error biomasa, definido en la sección 4.4. Así que es este criterio fue el que se utilizó para determinar qué funciones objetivo compartimentalizada del FBA permiten un modelado más fiel del comportamiento celular. Si bien en la ecuación [32] se presenta el valor absoluto de Error biomasa como objetivo externo, para el análisis del comportamiento celular sí se tuvo en cuenta el signo de este error, lo cual permite saber directamente si había subestimación o sobreestimación de la tasa de crecimiento celular.

Además de Error biomasa, también se calculó para cada función objetivo compartimentalizada explorada el Error *fluxes*, el cual se definió en la sección 4.4. Si bien este error no se usó como criterio de ajuste de las funciones, sí resulta útil tenerlo en cuenta en el análisis. De esta manera, se considera tanto el ajuste respecto al crecimiento celular, como el ajuste respecto a la producción de metabolitos secundarios para cada función objetivo probada.

Las gráficas de análisis de resultados se mejoraron incluyendo, además de las mejores funciones objetivo para cada categoría de datos experimentales, los errores cometidos en todos los casos con las funciones objetivos ‘maximización de producción de biomasa’ (por ser la función objetivos más popular hasta el momento) y ‘maximización de utilización de ATP’ (representada por  $Z = \max(\text{producción ATP})_{\text{citosol}} + \max(\text{producción ATP})_{\text{mitocondria}}$ ); los resultados con esta función se incluyeron a modo de control, pues generalmente no arroja buenas predicciones, hecho que ya se había visto en estudios anteriores (por ejemplo Knorr *et al.*, 2007), y que se hizo evidente al calcular los errores obtenidos con esa función en este trabajo. En todo caso, si son fijados los valores tanto de la ingesta de oxígeno como de la excreción de dióxido de carbono, la maximización de ATP como función objetivo puede dar buenos resultados (ver por ejemplo Schuetz *et al.*, 2007).

Los pesos relativos para las combinaciones se definieron como proporciones, evaluándose una gran cantidad de éstas (por ejemplo, 1:1:1:1, 1:2:3:4, 1:1:4:4, 1:1:1:8, 1:1:1:0, 1:1:0:0, etc., junto con sus respectivas permutaciones). Entre las proporciones se incluyeron valores cero, de modo que la evaluación considerara de nuevo subconjuntos de objetivos de los compartimentos, incluyendo cada uno de ellos de forma independiente. Esto permite que sean representadas muchas funciones objetivo evaluadas previamente, y que su desempeño sea comparado en este trabajo. Una descripción más detallada de los pesos relativos explorados en el presente trabajo puede encontrarse en los anexos (ver **Anexo B**).

En la **Tabla 7** se resumen los compartimentos celulares considerados y los objetivos propuestos para cada uno de ellos, cuyas combinaciones lineales (usando los pesos relativos mencionados en el párrafo anterior) generaron todas las funciones objetivo probadas.

**Tabla 7.** Posibles objetivos de los compartimentos celulares evaluados en la versión final del trabajo.

Global	Citosol	Mitocondria	Peroxisoma
max (producción biomasa)	max (producción NADH+NADPH)	max (producción ATP)	max (producción ácidos grasos)
min (producción biomasa)	max (consumo NADH+NADPH)	max (producción NADH+NADPH)	max (consumo ATP)
	min (producción NADH)	min (consumo NADH+NADPH)	min (consumo ATP)
	min (consumo NADH+NADPH)	max (transporte de ATP hacia citosol)	max (producción NADH+NADPH)
	min (producción NADH+NADPH)		min (consumo NADH+NADPH)
	max (producción ATP)		
	min (consumo ATP)		
	min (producción acetato)		
	min (producción CO <sub>2</sub> )		
	min (producción etanol)		
	min (producción succinato)		
	min (producción glicerol)		
	max (producción acetato)		
	max (producción CO <sub>2</sub> )		
	max (producción etanol)		
	max (producción succinato)		
	max (producción glicerol)		

Muchas funciones objetivo evaluadas previamente (Kauffman *et al.*, 2003; Knorr *et al.*, 2007; Schuetz *et al.*, 2007) quedan incluidas en el conjunto de funciones objetivo exploradas, entre ellas, maximización de producción de biomasa [ $Z = \max(\text{producción biomasa})_{\text{global}}$ ], minimización de utilización de poder reductor [ $Z = \min(\text{producción NAD(P)H})_{\text{citosol}} + \min(\text{consumo NAD(P)H})_{\text{mitocondria}} + \min(\text{consumo NAD(P)H})_{\text{peroxisoma}}$ ], maximización de tasa de producción de ATP [ $Z = \max(\text{producción ATP})_{\text{citosol}} + \max(\text{producción ATP})_{\text{mitocondria}}$ ], y minimización de producción de NADH [ $Z = \min(\text{producción NADH})_{\text{citosol}}$ ].

## 5. RESULTADOS Y DISCUSIÓN

Tal como se expresó en el capítulo 4, se calcularon los errores cometidos por las diferentes funciones objetivo probadas, tanto de la predicción de la tasa de crecimiento como de la predicción de los valores de los *fluxes* de intercambio. Entre estos errores, se usó el error en la estimación de biomasa para definir las mejores funciones objetivo, y el error en los *fluxes* de intercambio cometido por las mejores funciones fue tenido en cuenta en los análisis.

Vale la pena mencionar que, en cualquier caso, las funciones objetivo que condujeron a mejores estimaciones de la tasa de crecimiento, en general, presentaron también errores en la estimación de los *fluxes* cuyo orden de magnitud estaba entre los más bajos encontrados.

Los resultados (errores cometidos en la estimación de biomasa y *fluxes*) se presentan por medio de diagramas de caja y bigotes, dada la pluralidad de experimentos en cada una de las categorías. Un diagrama de caja y bigotes está conformado por una caja (rectángulo) que se extiende desde el primer hasta el tercer cuartil del conjunto de datos y está atravesada por una línea que indica la posición del segundo cuartil, y unos bigotes (líneas) que representan los demás datos experimentales. Por medio de puntos por fuera de los bigotes se representan los datos que se consideran atípicos, teniendo en cuenta el rango intercuartílico, esto es, la distancia entre el primer y el tercer cuartil (Montgomery y Runger, 2003). Para las categorías con una cantidad de datos baja no se presentan bigotes.

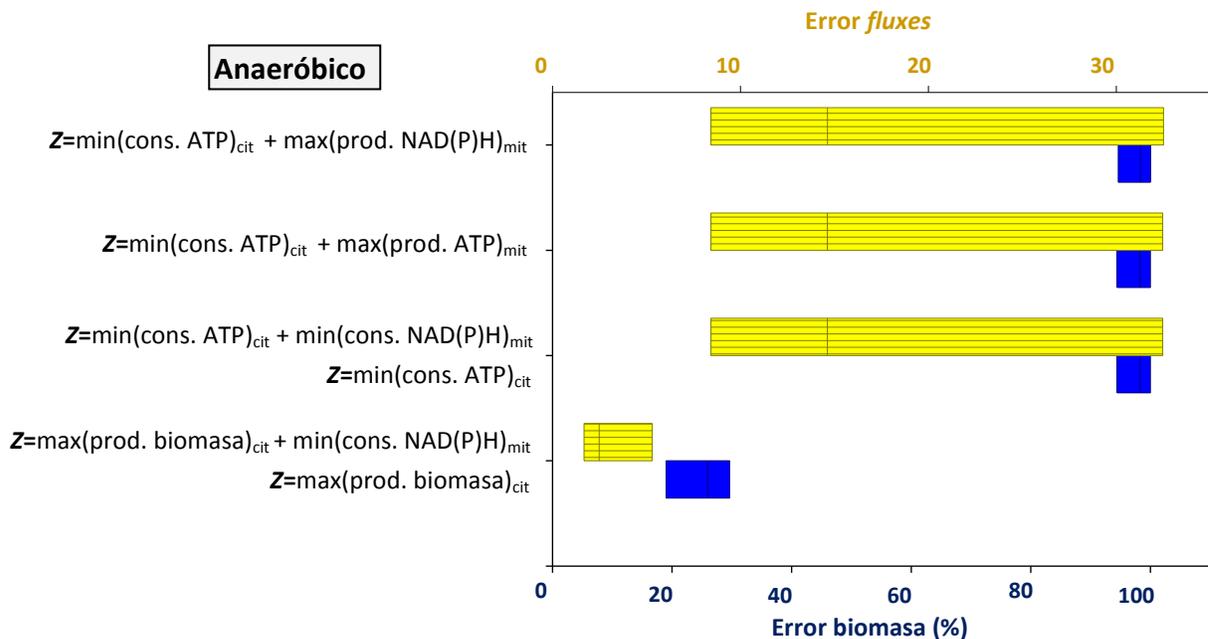
### 5.1. RESULTADOS PRIMERA VERSIÓN

Los datos experimentales fueron clasificados en tres categorías: crecimiento anaerobio (seis conjuntos de datos pertenecían a esta categoría), crecimiento aerobio con tasa de ingesta de oxígeno conocida (25 conjuntos de mediciones experimentales), y crecimiento aerobio sin conocimiento del *flux* de intercambio de oxígeno (14 conjuntos de datos).

Los resultados indicaron que, para todas las categorías, los objetivos del peroxisoma prácticamente no contribuyeron a las estimaciones del FBA; los valores estimados de cualquier combinación de objetivos de los otros dos compartimentos se mantenían casi constantes (las diferencias eran del orden de  $1,0 \times 10^{-4}$  %) al cambiar los objetivos del peroxisoma. Por lo anterior, en las mejores funciones para cada uno de los casos no se mencionan objetivos del compartimento peroxisoma.

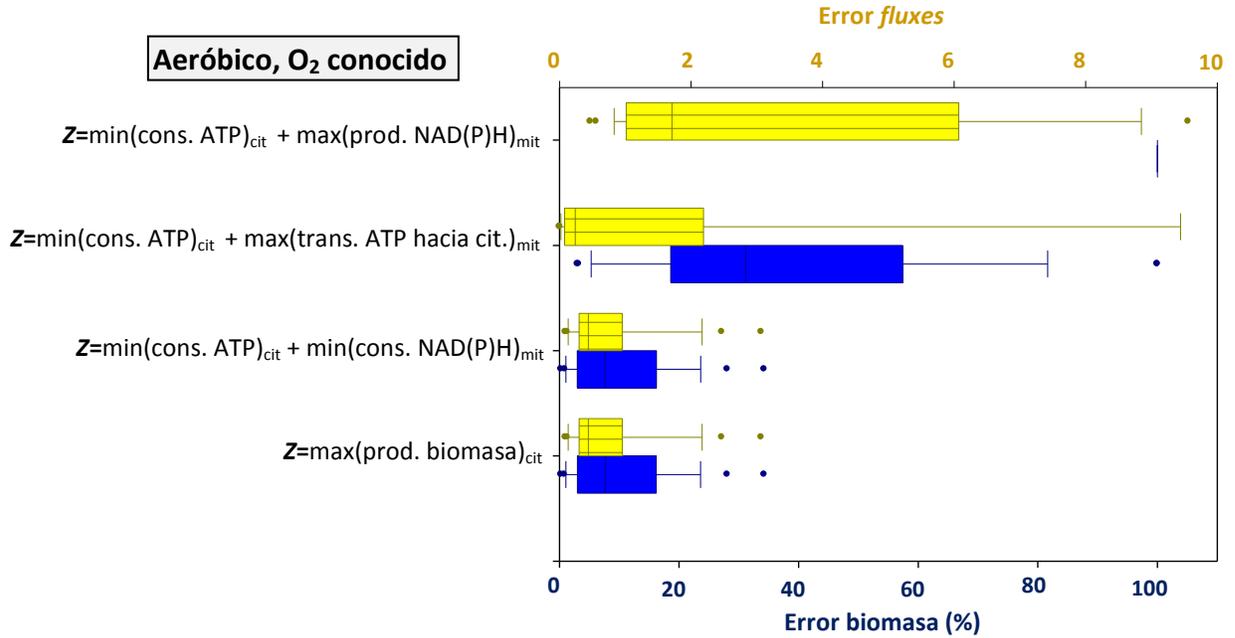
En la **Figura 7** se presentan diagramas de caja y bigotes de los errores de predicción cometidos por el FBA con las mejores funciones objetivo, para el caso de crecimiento anaeróbico de *S. cerevisiae* en cultivo continuo. Debido al bajo número de datos en la categoría (seis), los diagramas no presentan bigotes. Con los objetivos propuestos y las combinaciones permitidas (nótese que en esta primera versión, como ‘maximización de producción de biomasa’ estaba representada como un objetivo del citosol, no podía combinarse con ningún otro de los objetivos definidos para el citosol), es bastante evidente que ‘maximización de producción de biomasa’ da los mejores resultados como función objetivo del FBA, tanto en la predicción de crecimiento celular como en la de excreción de metabolitos secundarios.

**Figura 7.** Mejores funciones objetivo en FBA y sus errores de predicción, para el caso de crecimiento anaeróbico de *S. cerevisiae* en cultivo continuo, con la versión inicial de combinaciones de objetivos. En amarillo se muestran los errores en la estimación de *fluxes* de intercambio, mientras en azul se presentan los porcentajes de error en la predicción de la tasa específica de crecimiento.

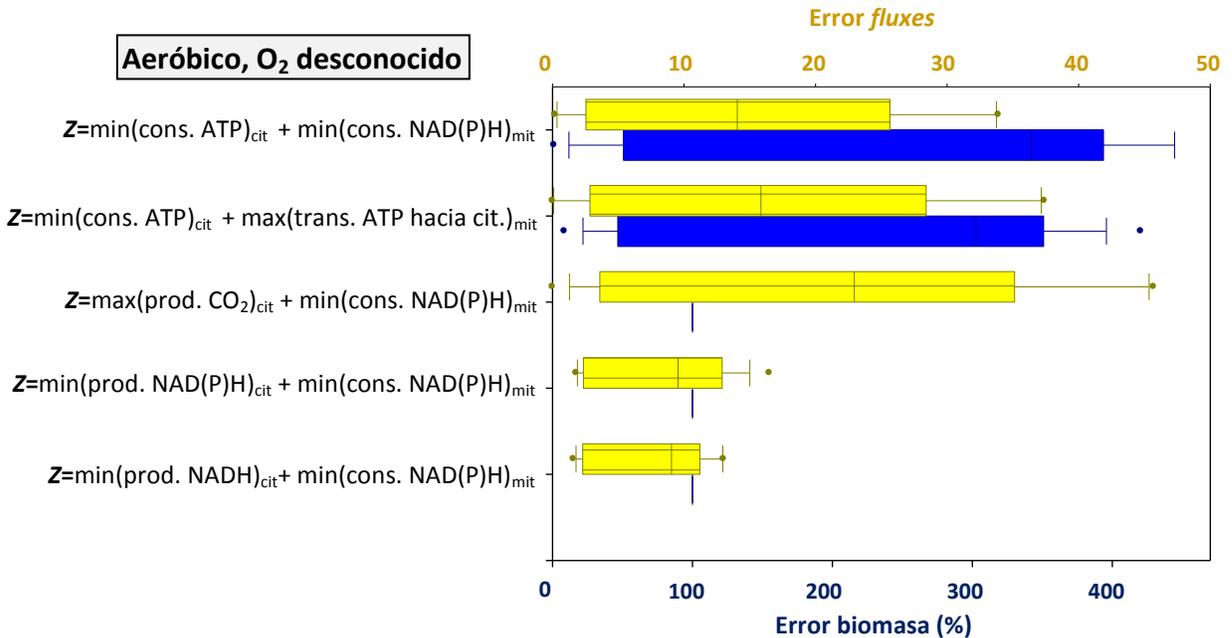


En la figura 7, también se aprecia que la combinación de ‘maximización de producción de biomasa’ con ‘minimización de consumo de NAD(P)H en mitocondria’ permitió obtener las mismas predicciones que ‘maximización de producción de biomasa’, pero hay que tener en cuenta que como eran casos de crecimiento anaerobio, los *fluxes* en la mitocondria no deben tener magnitudes relevantes para la célula. Todas las demás combinaciones de objetivos probadas conducen a errores de predicción muy superiores, excediendo el 90% en la predicción de la tasa específica de crecimiento y siendo más de tres veces superiores respecto a los *fluxes* de intercambio.

**Figura 8.** Errores cometidos por las mejores funciones objetivo en FBA usando las combinaciones de objetivos de la versión inicial, para el crecimiento aeróbico de *S. cerevisiae* en cultivo continuo con *flux* de intercambio de oxígeno conocido. En amarillo se muestran los errores en la estimación de *fluxes* de intercambio, mientras en azul se presentan los porcentajes de error en la predicción de la tasa específica de crecimiento.



**Figura 9.** Mejores funciones objetivo en FBA y sus errores de predicción, para el caso de crecimiento aeróbico de *S. cerevisiae*, sin datos acerca del valor del *flux* de intercambio de oxígeno, usando las combinaciones de objetivos de la versión inicial. En amarillo se muestran los errores en la estimación de producción de metabolitos, mientras en azul se presentan los porcentajes de error en la predicción de la tasa específica de crecimiento.



La **Figura 8** muestra los diagramas de caja y bigotes de los errores cometidos por las mejores cuatro funciones objetivo para los experimentos aerobios en los cuales la tasa de ingesta de oxígeno era conocida. Se puede observar que las mejores predicciones son obtenidas con el objetivo ‘maximización de crecimiento’, pero curiosamente, la función ‘minimización de consumo de ATP en citosol más minimización de consumo de NAD(P)H en mitocondria’ permite obtener unas predicciones iguales. Esto es debido a que, dado que el *flux* de oxígeno está fijo, minimizar los consumos de ATP en citosol y NAD(P)H en mitocondria implica para la célula dirigir los *fluxes* hacia el crecimiento. Este mismo efecto no pudo presentarse en el caso anaeróbico, pues en ausencia de oxígeno la relevancia de los *fluxes* mitocondriales es casi despreciable. Con todas las demás combinaciones de objetivos, los errores son significativamente mayores.

En la **Figura 9** se presentan los errores en las predicciones de las mejores cinco funciones objetivos probadas, en la categoría correspondiente al crecimiento aerobio sin *flux* de oxígeno conocido. Ninguna de las combinaciones de objetivos probadas en la versión inicial permite obtener buenas estimaciones; como se aprecia en la figura 9, las tres mejores funciones tienen un error del 100% respecto al crecimiento, lo que se debe a que estas tres funciones predicen que la célula no produce biomasa. Otras combinaciones, relacionadas principalmente con los objetivos ‘maximización de producción de biomasa’ y ‘minimización de consumo de ATP en el citosol’, sí predicen crecimiento, pero lo sobreestiman ostensiblemente, cometiendo errores incluso superiores al 400% en algunos casos.

Debe mencionarse que en esta primera versión, en la categoría cuyos resultados se muestran en la figura 9 se incluyeron tanto los experimentos continuos como los de crecimiento en fase exponencial de cultivo por lotes. Esto contribuyó a la alta dispersión de los errores cometidos por las funciones, dado que en el crecimiento en fase exponencial las células crecen a la mayor tasa posible, lo cual no ocurre en los cultivos continuos, ya que la tasa de crecimiento depende de la tasa de dilución del cultivo, y por consiguiente el modelado de las dos situaciones debiera diferir.

## 5.2. RESULTADOS VERSIÓN FINAL

Como se indicó en la sección 4.6, para la versión final del trabajo se introdujeron cambios en los siguientes aspectos:

1. Se creó un compartimento virtual, llamado ‘global’, para incluir en él únicamente los objetivos relacionados con la producción de biomasa. De esta manera, se podían probar las combinaciones de estos objetivos con posibles objetivos del citosol, que tienen mayor tendencia a ser influyentes que los objetivos de los otros compartimentos.

2. Las combinaciones entre los posibles objetivos ahora podían tener diferentes ponderaciones, las cuales se representaron con una serie de posibles vectores de pesos relativos.
3. Se incrementó el número de posibles objetivos del compartimento 'citosol', incluyendo objetivos relacionados con minimización de producción de metabolitos secundarios.
4. A pesar de la inexistente influencia de los objetivos del compartimento 'peroxisoma' evidenciada en la versión inicial, para la versión final se conservaron este compartimento y sus objetivos, teniendo presente la posibilidad de que se evidenciaran interacciones entre los objetivos del peroxisoma y los de otros compartimentos, o que los pesos relativos usados ocasionaran una mayor influencia del peroxisoma.

Y respecto a la definición de las categorías y al análisis de los resultados, estas son las diferencias principales:

1. En la clasificación de los datos experimentales respecto a la disponibilidad de oxígeno, para el caso de desconocimiento del *flux* de oxígeno se separaron los datos provenientes de cultivos continuos de aquellos provenientes de experimentos por lotes.
2. Se clasificaron los conjuntos de datos experimentales respecto a otros dos aspectos: la tasa específica de crecimiento y el tipo de sustrato. En la clasificación por tasas de crecimiento, los datos provenientes de cultivos continuos con *flux* de oxígeno conocido fueron divididos en tres categorías: tasas menores o iguales a  $0,15 \text{ h}^{-1}$  ( $\mu \leq 0,15 \text{ h}^{-1}$ ), tasas superiores a  $0,15 \text{ h}^{-1}$  pero no mayores a  $0,28 \text{ h}^{-1}$  ( $0,15 \text{ h}^{-1} < \mu \leq 0,28 \text{ h}^{-1}$ ), y tasas superiores a  $0,28 \text{ h}^{-1}$  ( $0,28 \text{ h}^{-1} < \mu$ ). Respecto al tipo de sustrato, los experimentos de crecimiento en reactores continuos en los cuales se midió la ingesta de oxígeno, se dividieron en experimentos en los que el sustrato era exclusivamente glucosa, y experimentos con sustratos diferentes a glucosa.
3. Para el análisis de resultados, además de los errores cometidos por las cinco mejores funciones objetivo de cada categoría, se presentaron también los errores de las predicciones hechas con las funciones objetivo 'maximización de producción de biomasa' [ $Z = \max(\text{producción biomasa})_{\text{global}}$ ] y 'maximización de producción de ATP' [ $Z = \max(\text{producción ATP})_{\text{citosol}} + \max(\text{producción ATP})_{\text{mitocondria}}$ ] en cada una de las categorías en que se dividieron los datos; la primera función se incluyó por ser la función con mayor soporte previo, y la segunda a modo de control, debido a que usar la maximización de producción de ATP como función objetivo conduce

consistentemente a malos resultados, producto de la subestimación del crecimiento celular. Esto se debe a que esta función objetivo desvía los *fluxes* celulares (y, por consiguiente, los nutrientes) hacia las reacciones productoras de ATP, y abandona las reacciones productoras de biomasa.

4. Si bien las mejores funciones objetivo fueron escogidas de acuerdo con el valor absoluto del promedio del porcentaje de error en la estimación de la producción de biomasa (ecuación [32]), para efectos de análisis en las figuras se presentan los diagramas de caja y bigotes del porcentaje de error en la predicción del crecimiento, pero conservando el signo del error, de manera que es posible identificar si las funciones objetivo conducen a sobreestimaciones o subestimaciones del verdadero valor de la producción de biomasa.

De nuevo, los posibles objetivos del peroxisoma no afectaron significativamente las predicciones obtenidas al aplicar el FBA, y por consiguiente no se muestran objetivos de ese compartimento entre las combinaciones que permitieron mejores estimaciones.

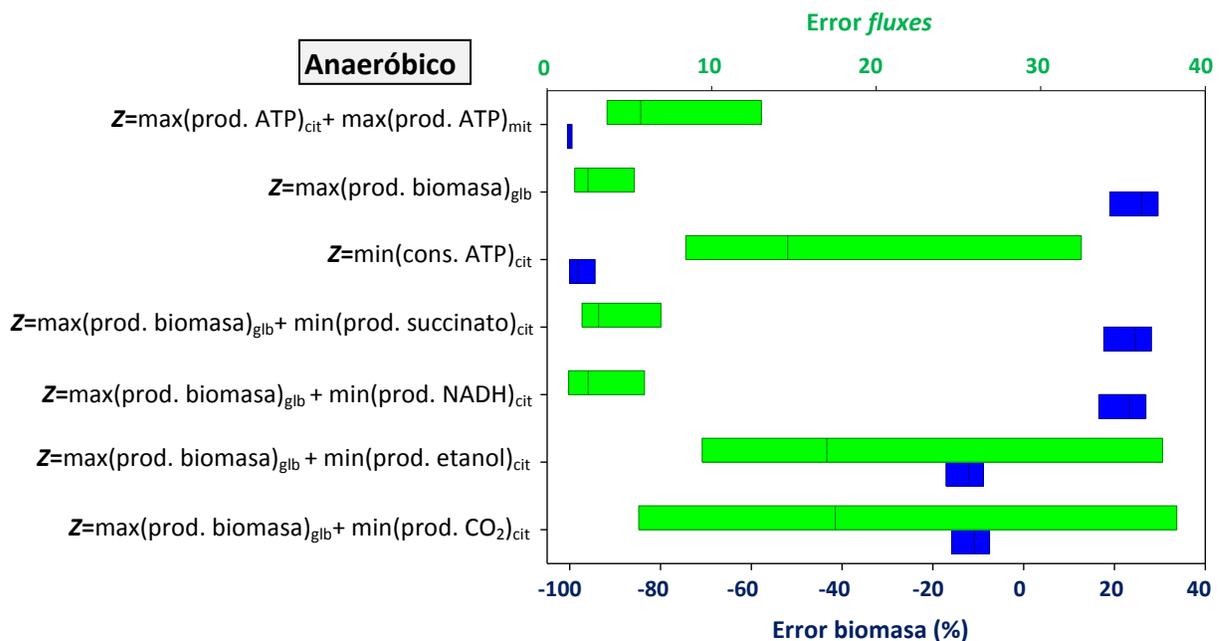
Adicionalmente, se presentó un efecto inesperado respecto a los pesos relativos: todas las predicciones obtenidas con combinaciones de funciones usando ponderaciones diferentes a vectores compuestos exclusivamente de unos y ceros, dieron los mismos resultados que alguna combinación (de los mismos objetivos) con ponderaciones formadas únicamente de unos y ceros. Por lo anterior, en los resultados no aparecen pesos relativos, porque se eligieron como representantes de estos conjuntos de ponderaciones con los mismos resultados a aquellas que estaban compuestas sólo por unos y ceros (es decir, cada objetivo que aparece tiene como peso relativo uno; los compartimentos que no aportan objetivos a una solución particular tienen un peso relativo igual a cero, indicando que esos compartimentos no están activos en esa combinación). Esto es fundamentalmente debido a la diferencia de orden de magnitud de los valores de los *fluxes* que integran los diferentes objetivos de los compartimentos propuestos, que ocasionó que las proporciones de pesos relativos exploradas no fueran suficientemente diferentes para ocasionar un cambio de tendencia respecto a los objetivos dominantes en las distintas situaciones. Sin embargo, es posible que con un mayor espectro de pesos relativos, éstos tengan efecto sobre las predicciones obtenidas.

## **1. Potencial predictivo respecto a la disponibilidad de oxígeno**

Respecto a la disponibilidad de oxígeno, los datos experimentales fueron clasificados en cuatro categorías: crecimiento anaerobio (seis conjuntos de datos en esta categoría), crecimiento aerobio en cultivo continuo con *flux* de oxígeno conocido (25 experimentos), crecimiento aerobio con ingesta de oxígeno desconocida en cultivo continuo (seis conjuntos de datos) y crecimiento aerobio con *flux* de entrada de oxígeno desconocido en

crecimiento en fase exponencial en cultivo por lotes (ocho datos experimentales). La subdivisión respecto al conocimiento del *flux* de oxígeno fue necesaria debido a la gran diferencia en la precisión de las predicciones del FBA entre los casos en que dicho *flux* es conocido y en los que no; cuando la ingesta de oxígeno no es conocida con precisión, el error en las estimaciones tiende a ser mucho más alto. Las figuras 10 a 13 muestran (por medio de gráficos de caja y bigotes) los errores en las predicciones cometidos con las mejores cinco funciones objetivo para cada una de las categorías respecto a la disponibilidad de oxígeno, así como los errores cometidos usando ‘maximización de producción de biomasa’ y ‘maximización de producción de ATP’ en las mismas clasificaciones de datos.

**Figura 10.** Mejores funciones objetivo en FBA y sus errores de predicción, para el caso de crecimiento anaeróbico de *S. cerevisiae* en cultivo continuo, en la versión final del trabajo.



La **Figura 10** presenta el caso de crecimiento anaeróbico, y se puede observar que las mejores predicciones fueron obtenidas con ‘maximización de crecimiento’, ‘maximización de crecimiento más minimización de producción de NADH en el citosol’, y ‘maximización de crecimiento más minimización de producción de succinato en el citosol’. Con los datos usados, estos tres objetivos condujeron a una sobreestimación de la tasa específica de crecimiento de alrededor del 25%, y un bajo error en la predicción de los *fluxes* de intercambio.

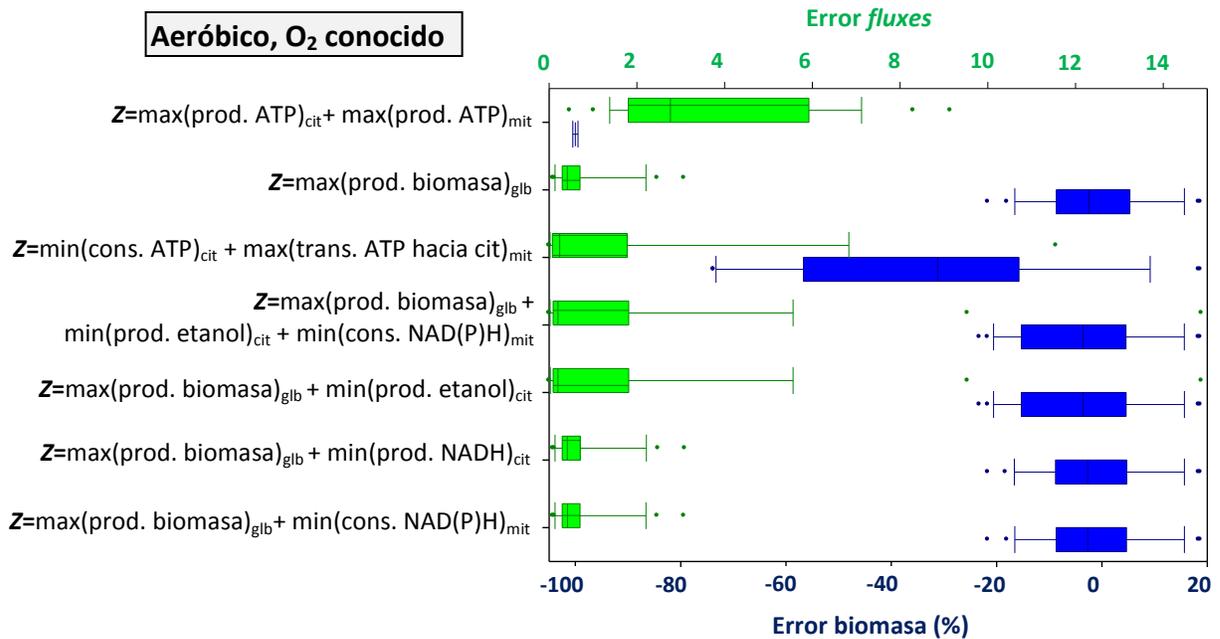
Los otros dos objetivos interesantes fueron ‘maximización de crecimiento más minimización de producción de etanol en el citosol’ y ‘maximización de crecimiento más

minimización de producción de CO<sub>2</sub> en el citosol', que condujeron a una predicción de la tasa de crecimiento ligeramente superior (el error fue de menos del 20%), pero subestimando dicha tasa. Por otra parte, el error aumentó mucho más en la estimación respecto a los *fluxes* de intercambio, y por esto es que no resultan mejores que los tres objetivos mencionados previamente.

En los casos de crecimiento anaerobio, todas las demás funciones evaluadas ocasionaron errores significativamente más grandes que los de las cinco funciones mencionadas. Por ejemplo, como también puede verse en la figura 10, con la quinta mejor función objetivo ('minimización de consumo de ATP en el citosol') se obtuvo un error en la predicción de la tasa de crecimiento de casi un 100%.

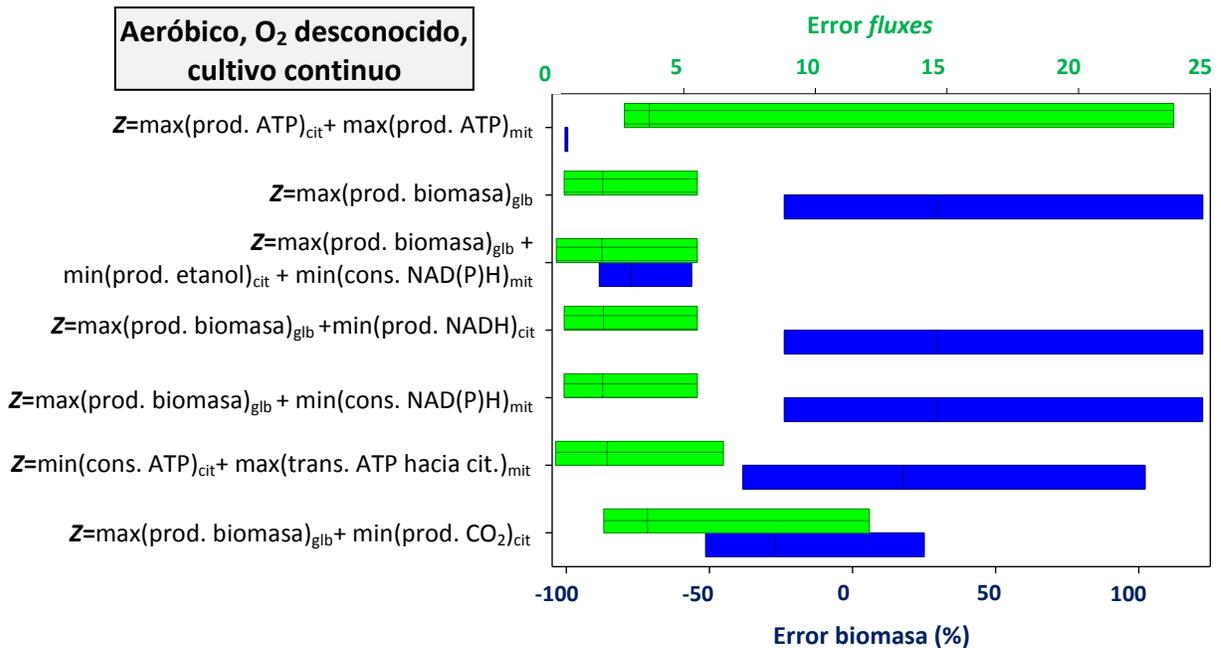
Para el caso del crecimiento aerobio en cultivo continuo con ingesta de oxígeno conocida, cuyos resultados se presentan en la **Figura 11**, las combinaciones 'maximización de crecimiento' y 'maximización de crecimiento más minimización de producción de NADH en el citosol' estuvieron de nuevo entre las mejores funciones objetivo para FBA. Con el objetivo 'maximización de crecimiento más minimización de consumo de NAD(P)H en la mitocondria' se obtuvo el mismo error de predicción. Adicionalmente, otras dos funciones objetivo permitieron obtener un nivel de predicción comparable, y el resto de combinaciones condujeron a estimaciones de una calidad muy inferior.

**Figura 11.** Mejores funciones objetivo en FBA y sus errores de predicción, para el caso de crecimiento aeróbico de *S. cerevisiae* en cultivo continuo conociendo el valor del *flux* de intercambio de oxígeno, en la versión final del trabajo.



En la tercera categoría, crecimiento aerobio en cultivo continuo con ingesta de oxígeno conocida (**Figura 12**), se presentaron nuevamente algunas funciones con una precisión comparable a la obtenida con ‘maximización de crecimiento’, y de nuevo entre estas funciones se encuentra ‘maximización de crecimiento más minimización de producción de NADH en el citosol’.

**Figura 12.** Mejores funciones objetivo (entre las combinaciones de objetivos exploradas en la versión final) en FBA y sus errores de predicción, para el caso de crecimiento aeróbico de *S. cerevisiae* en cultivo continuo, sin conocer el valor del *flux* de intercambio de oxígeno.

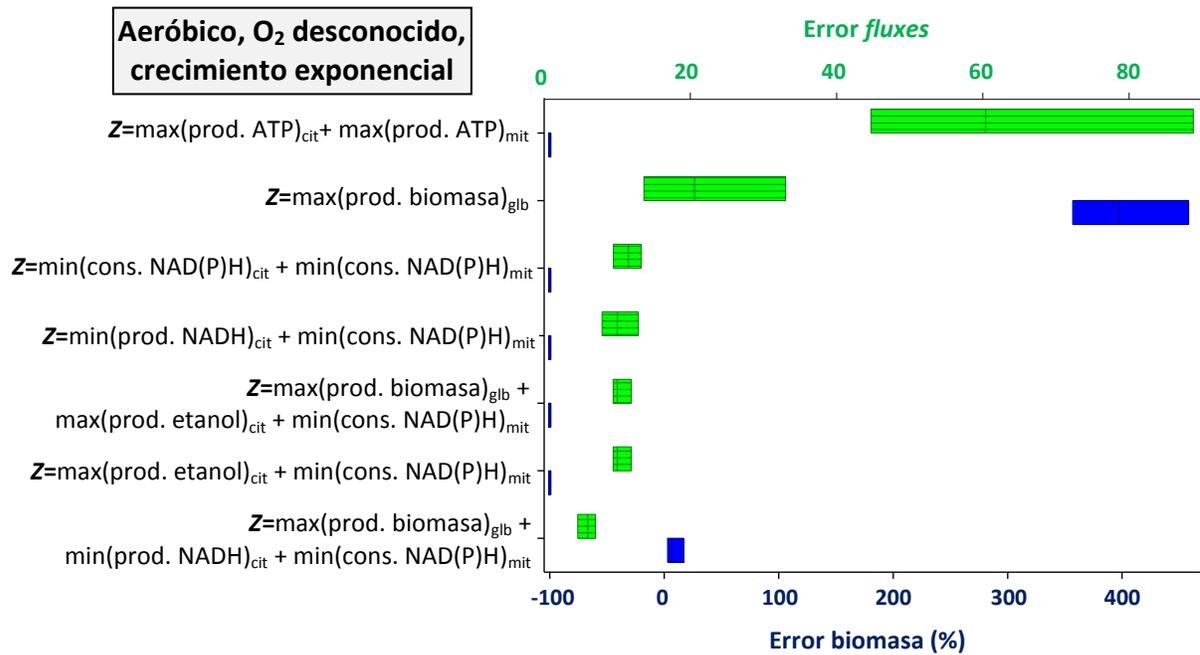


Se hace evidente en la figura 12 que la falta de conocimiento acerca del *flux* de entrada de oxígeno ocasiona una tendencia de sobreestimar considerablemente la tasa específica de crecimiento, excediendo incluso el 100% de error, para las combinaciones que incluyen ‘maximización de crecimiento’ entre sus elementos. Sólo una de las mejores funciones, ‘maximización de crecimiento más minimización de producción de CO<sub>2</sub> en el citosol’, tiende a subestimar el crecimiento en vez de sobreestimarlos, presentando unas predicciones de los *fluxes* de intercambio ligeramente peores. Pero en general, las predicciones son peores que las obtenidas cuando es conocida la tasa de ingesta de oxígeno (está claro que en el caso anaerobio esta tasa es conocida, siendo igual a cero).

La tendencia fue totalmente diferente en la cuarta categoría de la clasificación, es decir, en el caso de crecimiento aerobio en fase exponencial en un cultivo por lotes con desconocimiento de la tasa de ingesta de oxígeno, cuyos resultados se presentan en la **Figura 13**. La función ‘maximización de crecimiento’ ahora causa un error en la predicción

de la tasa específica de crecimiento superior al 300% en todos los conjuntos de datos experimentales usados.

**Figura 13.** Mejores funciones objetivo en FBA y sus errores de predicción, para el caso de crecimiento aeróbico de *S. cerevisiae* en fase exponencial de crecimiento en cultivo por lotes, sin conocer el valor del *flux* de intercambio de oxígeno, en la versión final del trabajo.



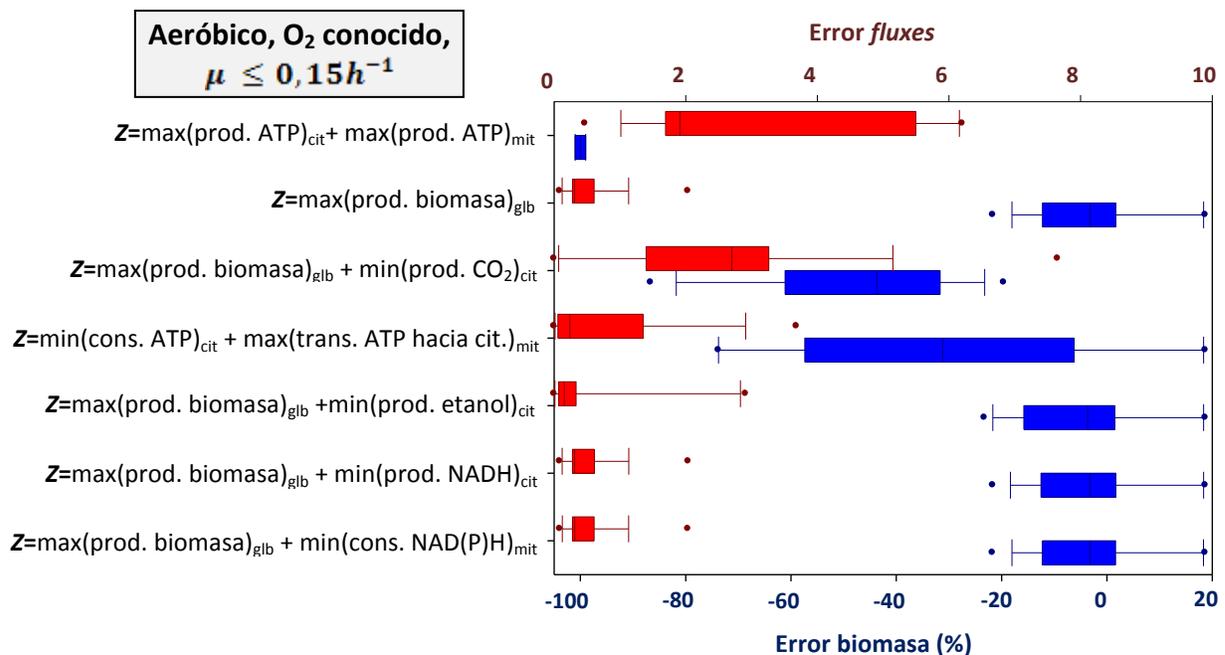
Cuantitativamente, las mejores predicciones fueron obtenidas con funciones que predicen que no hay crecimiento (es decir, subestiman el crecimiento en un 100%), excepto por una función que realmente se destaca entre todas: la combinación ‘maximización de crecimiento más minimización de producción de NADH en el citosol más minimización de consumo de NAD(P)H en la mitocondria’, que permite obtener predicciones de la tasa específica de crecimiento con un error inferior al 20%, y estimaciones del intercambio de metabolitos con el ambiente con un *Error fluxes* (ecuación [30]) menor a 10. La figura 13 muestra claramente la enorme diferencia entre las predicciones obtenidas con ‘maximización de crecimiento más minimización de producción de NADH en el citosol más minimización de consumo de NAD(P)H en la mitocondria’ y las obtenidas con cualquier otra combinación de objetivos para los conjuntos de datos evaluados, siendo esta mejor combinación una opción muy atractiva para modelar el crecimiento aerobio en fase exponencial de *S. cerevisiae*.

## 2. Potencial predictivo respecto a la tasa de crecimiento, en crecimiento aerobio en cultivo continuo y con tasa de ingesta de oxígeno conocida

Los conjuntos de datos correspondientes a cultivos continuos en los que se determinó la ingesta de oxígeno provenían de experimentos con diferentes tasas de dilución, por lo cual se decidió explorar el desempeño de las funciones respecto a las tasas específicas de crecimiento. Así, los datos experimentales de crecimiento en cultivo continuo con *flux* de oxígeno conocido fueron clasificados en tres grupos: tasa específica de crecimiento menor o igual a  $0,15 \text{ h}^{-1}$  (19 conjuntos de datos en esta categoría), tasa específica de crecimiento superior a  $0,15 \text{ h}^{-1}$  y no mayor a  $0,28 \text{ h}^{-1}$  (tres conjuntos de datos), y tasa específica de crecimiento superior a  $0,28 \text{ h}^{-1}$  (tres conjuntos de datos).

Las figuras 14, 15 y 16 muestran las mejores cinco funciones objetivo (entre las combinaciones probadas) del FBA en cada una de estas categorías, así como los dos objetivos usados para comparación. El porcentaje de error en la estimación de la producción de biomasa se muestra por medio de cajas y bigotes azules, mientras que el error en la estimación de los demás *fluxes* (ecuación [30]) se presenta con diagramas de caja y bigotes de color rojo.

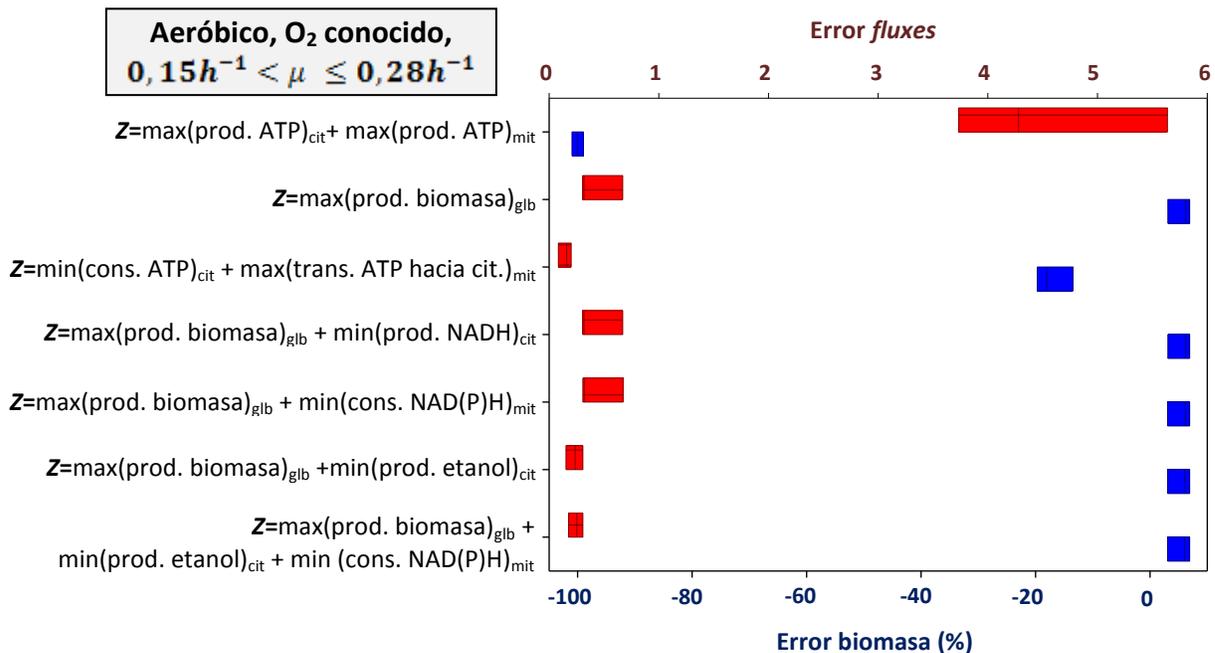
**Figura 14.** Mejores funciones objetivo en FBA y sus errores de predicción, para el caso de crecimiento aeróbico de *S. cerevisiae* en cultivo continuo, con tasas de dilución iguales o menores a  $0,15 \text{ h}^{-1}$ , en la versión final del trabajo. El *flux* de oxígeno es conocido.



En la **Figura 14** se aprecian los resultados obtenidos con las tasas de dilución más bajas. El objetivo ‘maximización de producción de biomasa’ conduce a los mejores resultados, con los mismos errores cometidos por otras dos funciones: ‘maximización de producción de biomasa más minimización de consumo de NAD(P)H en la mitocondria’ y ‘maximización de producción de biomasa más minimización de producción de NADH en el citosol’. Con una tercera función se alcanza un error comparable en la predicción de la tasa específica de crecimiento, pero arroja peores predicciones sobre los *fluxes* de intercambio. Las demás combinaciones probadas conducen a estimaciones mucho más malas al ser usados como objetivos del FBA.

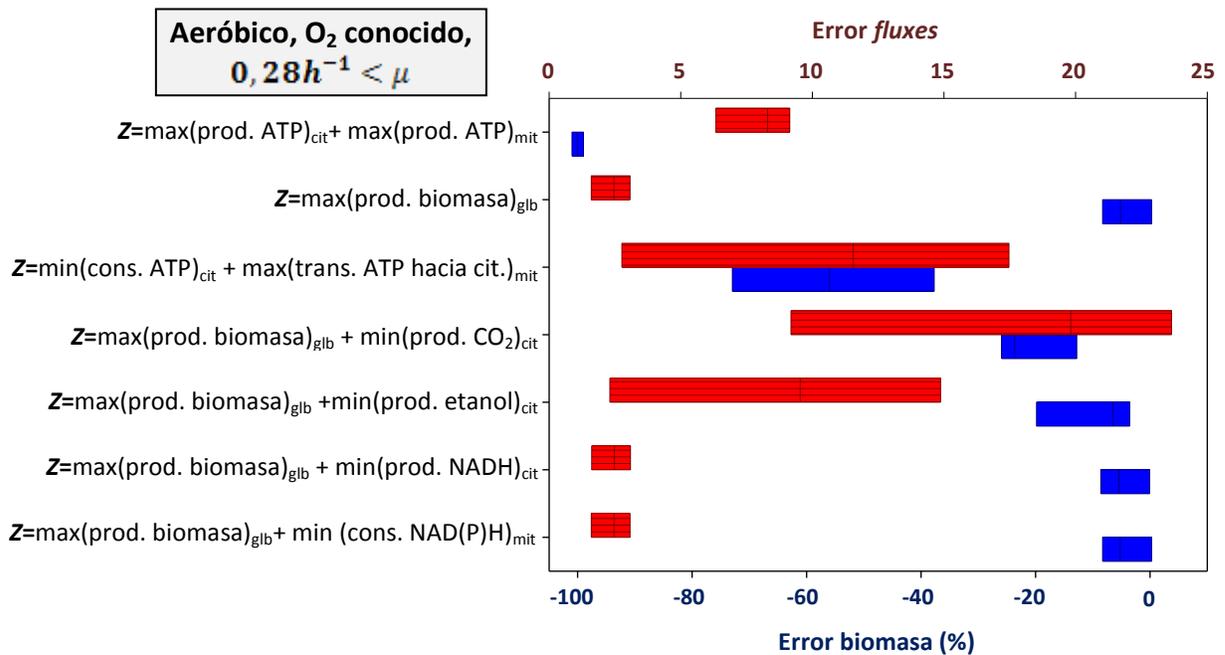
En la categoría correspondiente a las tasas de dilución medias, cuyos resultados se presentan en la **Figura 15**, se encuentra que el objetivo ‘maximización de producción de biomasa’ está entre los que permiten obtener mejores estimaciones de la tasa específica de crecimiento, pero hay un par de funciones que lo superan respecto a la predicción de la producción de metabolitos (teniendo el mismo error en la estimación del crecimiento): las funciones ‘maximización de producción de biomasa más minimización de producción de etanol más minimización de consumo de NAD(P)H en la mitocondria’ y ‘maximización de producción de biomasa más minimización de producción de etanol’. En cualquier caso, la diferencia en los errores entre estos objetivos no es muy grande.

**Figura 15.** Mejores funciones objetivo en FBA y sus errores de predicción, para el caso de crecimiento aeróbico de *S. cerevisiae* en cultivo continuo, con tasas de dilución mayores a  $0,15 \text{ h}^{-1}$  y no superiores a  $0,28 \text{ h}^{-1}$ , en la versión final del trabajo. El *flux* de oxígeno es conocido.



Todas las funciones que se encuentran entre las mejores, y que pueden ser observadas en la figura 15, involucran la maximización del crecimiento entre sus elementos, excepto una, que es ‘minimización de consumo de ATP en el citosol más maximización de transporte de ATP hacia citosol en la mitocondria’. Esta combinación puede parecer una contradicción de objetivos, pero esa aparente contradicción es lo que explicaría su buen desempeño en este caso: se busca que se minimice el consumo de ATP en el citosol, y a la vez se busca que la mitocondria se encargue de su principal labor (producir ATP) pero sujeta a la anterior restricción, por lo cual gran parte de los nutrientes pueden redirigirse hacia el crecimiento celular.

**Figura 16.** Mejores funciones objetivo en FBA y sus errores de predicción, para el caso de crecimiento aeróbico de *S. cerevisiae* en cultivo continuo, con tasas de dilución mayores a  $0,28 \text{ h}^{-1}$ , en la versión final del trabajo. El *flux* de oxígeno es conocido.



La **Figura 16** presenta los resultados de la categoría de mayores tasas de dilución. Curiosamente, con los datos experimentales usados, los resultados son cualitativamente muy similares a los observados en el caso de bajas tasas de dilución, aunque podía esperarse que el objetivo celular fuera diferente en este caso debido a la redirección de *fluxes* hacia producción de etanol que se presenta en altas tasas de dilución. Es posible que esto sea debido al conocimiento de la tasa de consumo de oxígeno, con lo cual se limitaba el crecimiento; sin embargo, una mejor explicación estaría relacionada con los conjuntos de datos usados, pues se aprecia claramente que con todas las funciones objetivo probadas la tasa específica de crecimiento se subestimó. Esto se evidencia fácilmente porque el signo de Error biomasa es negativo con todas las funciones objetivo

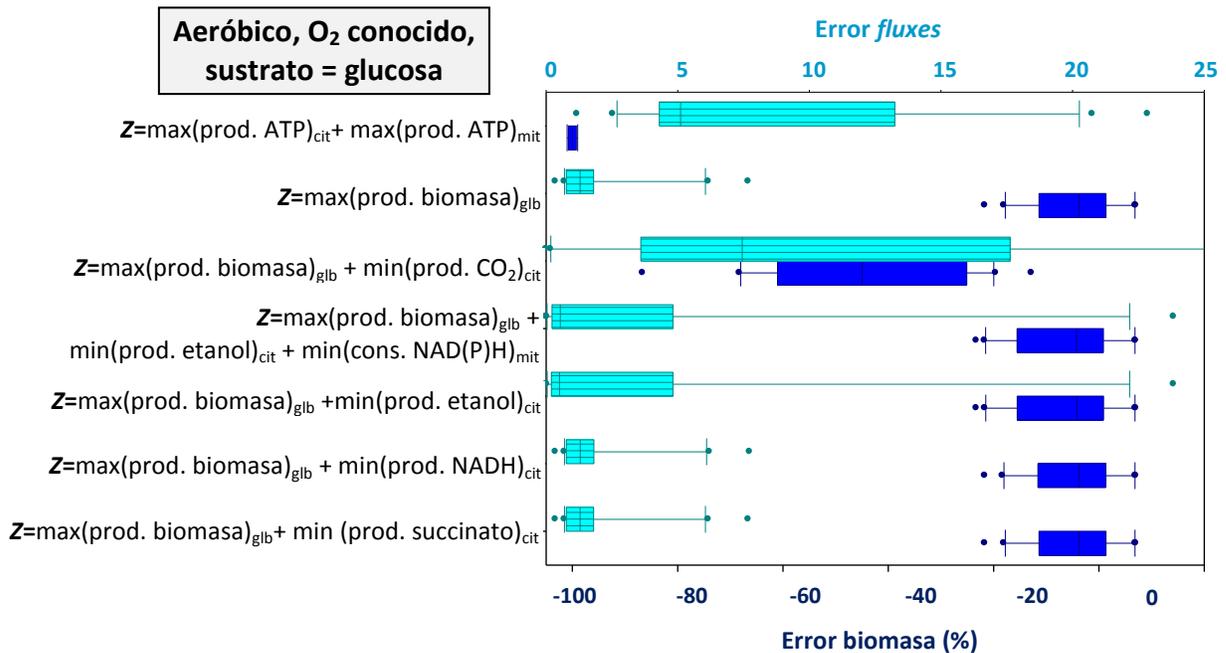
evaluadas, y como se explicó en la ecuación [29], un signo negativo implica que la tasa específica de crecimiento estimada  $\mu_{estimado}$  es inferior a la experimental  $\mu_{medido}$ .

Lo mencionado en el anterior párrafo manifiesta un problema en el balance del carbono de algunos datos experimentales, y explica por qué la ‘maximización de producción de biomasa’ no sobreestimó la tasa de crecimiento, lo cual permitió que fuera uno de los objetivos que arrojó mejores predicciones.

### 3. Potencial predictivo respecto al tipo de sustrato, en crecimiento aerobio en cultivo continuo con tasa de ingesta de oxígeno conocida

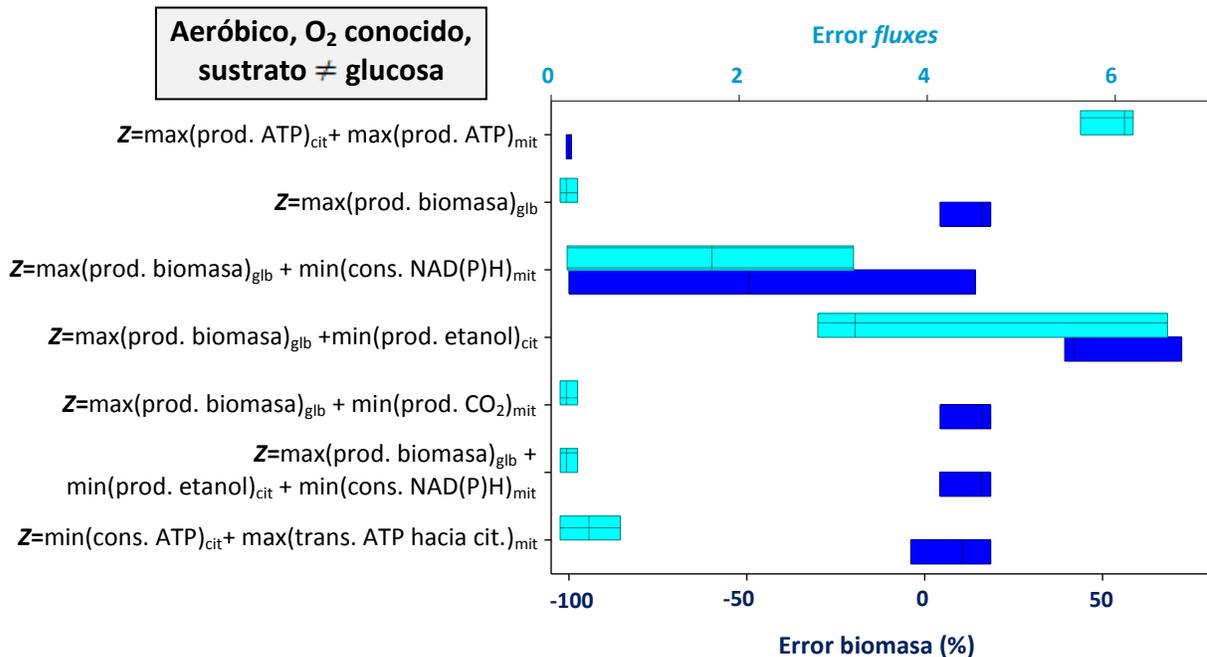
Las predicciones del FBA con las diferentes combinaciones probadas fueron evaluadas respecto al tipo de sustrato, clasificando los datos experimentales provenientes de cultivos continuos con ingesta de oxígeno medida en dos grupos: experimentos con glucosa como sustrato (20 datos experimentales), y aquellos en los que el sustrato no era exclusivamente glucosa (cuatro conjuntos de datos). Las figuras 17 y 18 presentan los resultados de estas categorías, usando el color azul oscuro para presentar el porcentaje de error en la estimación de la biomasa, y un color celeste para mostrar las cajas relacionadas con el error en la predicción de los *fluxes* de intercambio.

**Figura 17.** Mejores funciones objetivo en FBA y sus errores de predicción, para el caso de crecimiento aeróbico de *S. cerevisiae* en cultivo continuo, con glucosa como sustrato, y *flux* de oxígeno conocido.



Las **Figuras 17 y 18** muestran que en ambos casos, las mejores funciones son ‘maximización de producción de biomasa’ y un par de combinaciones que la incluyen. La única excepción se presenta en el caso de sustratos diferentes a glucosa, con la función ‘minimización de consumo de ATP en el citosol más maximización de transporte de ATP hacia citosol en la mitocondria’, con la cual se obtienen predicciones de crecimiento ligeramente mejores, y estimaciones de producción de metabolitos casi iguales. Vale la pena notar que había una menor cantidad de datos experimentales en esta segunda categoría, y por consiguiente los análisis realizados eran más sensibles a la presencia de ruido en los datos y a sesgos causados por datos particulares; adicionalmente, cinco de los seis conjuntos de datos de “sustrato diferente a glucosa” sólo contenían mediciones experimentales de oxígeno y dióxido de carbono, por lo cual para comparar el error en los fluxes sólo se contaba con un dato. De todas formas, este objetivo podría ser un objetivo del FBA interesante para analizar tomando mayor cantidad de datos de crecimiento de *S. cerevisiae* en sustratos diferentes a glucosa, en caso de que la mejor función objetivo sí dependa del sustrato usado, como lo han sugerido estudios previos (ver por ejemplo Ibarra *et al.*, 2002).

**Figura 18.** Mejores funciones objetivo en FBA y sus errores de predicción, para el caso de crecimiento aeróbico de *S. cerevisiae* en cultivo continuo, con sustratos diferentes a glucosa, y *flux* de oxígeno conocido.



En la **Tabla 8** se presenta un resumen de las mejores funciones objetivo obtenidas para cada condición. Se presenta también una comparación de los rangos de valores de los errores arrojados al usar estas funciones, y al usar la función “maximización de producción de biomasa”. En general, si el *flux* de oxígeno es conocido y se usa como dato de entrada para el FBA, las estimaciones obtenidas con “maximización de producción de biomasa” están entre las mejores. Por otra parte, cuando el *flux* de oxígeno es desconocido y no se puede usar como dato de entrada, las predicciones obtenidas con “maximización de producción de biomasa” tienen muy baja calidad, pues se sobreestima en gran medida el crecimiento celular. En estas situaciones, unas cuantas funciones objetivo específicas tienen un desempeño muy superior.

También vale la pena recalcar que, en general, las magnitudes de los errores de predicción tanto de crecimiento como de *fluxes* de intercambio son actualmente muy grandes. Esto se debe parcialmente a la alta incertidumbre experimental que se presenta, tanto por la misma naturaleza de los microorganismos como por las variadas técnicas de medición usadas, y parcialmente a la calidad de los modelos metabólicos existentes. A pesar de que se usó el mejor modelo metabólico compartimentalizado de *S. cerevisiae* existente, se vio que este también presenta múltiples falencias y vacíos en la red, lo cual se comprobó cuando se eligieron los compartimentos que se propondrían como aportantes de objetivos compartimentales (ver secciones 4.5 y 4.6).

**Tabla 8.** Resumen de los resultados obtenidos en el presente estudio.

Condición experimental		Núm. de exp.	Rango de los errores usando “max(prod. biomasa) <sub>glb</sub> ”	Mejores funciones objetivo	Rango de los errores usando las mejores funciones objetivo	Comentarios
Cultivo continuo, anaeróbico		6	11% a 34% en biomasa  1 a 7 en <i>fluxes</i> de intercambio	<ul style="list-style-type: none"> <li>max(prod. biomasa)<sub>glb</sub> + min(prod. etanol)<sub>cit</sub></li> <li>max(prod. biomasa)<sub>glb</sub> + min(prod. CO<sub>2</sub>)<sub>cit</sub></li> </ul>	-8% a -22% en biomasa  5 a 47 en <i>fluxes</i> de intercambio	Las funciones “max(prod. biomasa) <sub>glb</sub> + min(prod. NADH) <sub>cit</sub> ” y “max(prod. biomasa) <sub>glb</sub> + min(prod. NADH) <sub>cit</sub> ” arrojan predicciones similares a las obtenidas con “max(prod. biomasa) <sub>glb</sub> ”.
Cultivo continuo, aeróbico, ingesta de oxígeno conocida	Sin condiciones adicionales	25	-21% a 19% en biomasa  0,15 a 2 en <i>fluxes</i> de intercambio	<ul style="list-style-type: none"> <li>max(prod. biomasa)<sub>glb</sub> + min(cons. NAD(P)H)<sub>mit</sub></li> <li>max(prod. biomasa)<sub>glb</sub></li> </ul>	-21% a 19% en biomasa  0,15 a 2 en <i>fluxes</i> de intercambio	Las funciones “max(prod. biomasa) <sub>glb</sub> + min(prod. etanol) <sub>cit</sub> ” y “max(prod. biomasa) <sub>glb</sub> + min(prod. etanol) <sub>cit</sub> +

				<ul style="list-style-type: none"> <li>+ min(prod. NADH)<sub>cit</sub></li> <li>• max(prod. biomasa)<sub>glb</sub></li> </ul>		min(cons. NAD(P)H) <sub>mit</sub> ” arrojan predicciones ligeramente inferiores de los <i>fluxes</i> de intercambio, respecto a las mejores funciones.
$\mu$ menor o igual a 0,15 h <sup>-1</sup>	19	<p>-18% a 19% en biomasa</p> <p>0,15 a 2 en <i>fluxes</i> de intercambio</p>	<ul style="list-style-type: none"> <li>• max(prod. biomasa)<sub>glb</sub>+ min(cons. NAD(P)H)<sub>mit</sub></li> <li>• max(prod. biomasa)<sub>glb</sub> + min(prod. NADH)<sub>cit</sub></li> <li>• max(prod. biomasa)<sub>glb</sub></li> </ul>	<p>-18% a 19% en biomasa</p> <p>0,15 a 2 en <i>fluxes</i> de intercambio</p>	La función “max(prod. biomasa) <sub>glb</sub> + min(prod. etanol) <sub>cit</sub> ” es sólo ligeramente inferior a las mejores.	
$\mu$ mayor a 0,15 y menor o igual a 0,28 h <sup>-1</sup>	3	<p>3% a 7% en biomasa</p> <p>0,3 a 0,7 en <i>fluxes</i> de intercambio</p>	<ul style="list-style-type: none"> <li>• max(prod. biomasa)<sub>glb</sub>+ min(prod. etanol)<sub>cit</sub></li> <li>• max(prod. biomasa)<sub>glb</sub> + min(prod. etanol)<sub>cit</sub> + min (cons. NAD(P)H)<sub>mit</sub></li> </ul>	<p>3% a 7% en biomasa</p> <p>0,2 a 0,3 en <i>fluxes</i> de intercambio</p>	Otros dos objetivos arrojaron los mismos errores que “max(prod. biomasa) <sub>glb</sub> ”. Estos fueron ligeramente peores que los de las mejores funciones.	
$\mu$ mayor a 0,28 h <sup>-1</sup>	3	<p>-8% a 0,3% en biomasa</p> <p>1,6 a 3,1 en <i>fluxes</i> de intercambio</p>	<ul style="list-style-type: none"> <li>• max(prod. biomasa)<sub>glb</sub>+ min(cons. NAD(P)H)<sub>mit</sub></li> <li>• max(prod. biomasa)<sub>glb</sub> + min(prod. NADH)<sub>cit</sub></li> <li>• max(prod. biomasa)<sub>glb</sub></li> </ul>	<p>-8% a 0,3% en biomasa</p> <p>1,6 a 3,1 en <i>fluxes</i> de intercambio</p>	Las predicciones de las demás funciones objetivo fueron significativamente peores.	
Glucosa como sustrato	20	<p>-22% a 7% en biomasa</p> <p>0,1 a 3 en <i>fluxes</i> de intercambio</p>	<ul style="list-style-type: none"> <li>• max(prod. biomasa)<sub>glb</sub>+ min(prod. succinato)<sub>cit</sub></li> <li>• max(prod. biomasa)<sub>glb</sub> + min(prod. NADH)<sub>cit</sub></li> <li>• max(prod. biomasa)<sub>glb</sub></li> </ul>	<p>-22% a 7% en biomasa</p> <p>0,1a 3 en <i>fluxes</i> de intercambio</p>	Otras dos funciones tuvieron predicciones ligeramente más malas respecto a crecimiento, pero significativamente peores respecto a los demás <i>fluxes</i> de intercambio.	
Sustrato diferente a	5	1% a 19% en biomasa	<ul style="list-style-type: none"> <li>• min(cons. ATP)<sub>cit</sub> +</li> </ul>	-6% a 19% en biomasa	Las funciones “max(prod.	

	glucosa		0,1 a 0,3 en <i>fluxes</i> de intercambio	max(trans. ATP hacia cit.) <sub>mit</sub>	0,1 a 0,7 en <i>fluxes</i> de intercambio	biomasa) <sub>glb</sub> + min(prod. CO <sub>2</sub> ) <sub>mit</sub> " y "max(prod. biomasa) <sub>glb</sub> + min(prod. etanol) <sub>cit</sub> + min(cons. NAD(P)H) <sub>mit</sub> " dan predicciones similares a las obtenidas con "max(prod. biomasa) <sub>glb</sub> ".
Cultivo continuo, aeróbico, ingesta de oxígeno desconocida	6	-28% a 190% en biomasa 0,5 a 11 en <i>fluxes</i> de intercambio	• max(prod. biomasa) <sub>glb</sub> + min(prod. CO <sub>2</sub> ) <sub>cit</sub>	-62% a 60% en biomasa 0 a 16 en <i>fluxes</i> de inter-cambio	La función "min(cons. ATP) <sub>cit</sub> + max(trans. ATP hacia cit.) <sub>mit</sub> " también es superior a "max(prod. biomasa) <sub>glb</sub> ", pero la diferencia en las predicciones es pequeña.	
Cultivo <i>batch</i> en fase exponencial, aeróbico, ingesta de oxígeno desconocida	8	330% a 485% en biomasa 9 a 34 en <i>fluxes</i> de intercambio	• max(prod. biomasa) <sub>glb</sub> + min(prod. NADH) <sub>cit</sub> + min(cons. NAD(P)H) <sub>mit</sub>	-6% a 20% en biomasa 0,5 a 34 en <i>fluxes</i> de inter-cambio	Después de la mejor función, las siguientes en exactitud numérica son las que no predicen crecimiento (error en biomasa de -100%). Las otras funciones objetivo que sí predicen crecimiento tienen errores mucho más altos, comparables a los obtenidos con "max(prod. biomasa) <sub>glb</sub> ".	

### 5.3. DISCUSIÓN

A pesar de los grandes avances en las técnicas de medición de componentes celulares en las últimas décadas, aún falta mucho para lograr un completo conjunto de medidas de todos los diferentes metabolitos y enzimas en tiempo real. El resultado de esta limitación se refleja en la sub-determinación que se obtiene cuando se intenta modelar un sistema celular cualquiera; si el objetivo es modelar de una manera exhaustiva una red celular, por ejemplo, se obtienen invariablemente más incógnitas que ecuaciones.

La manera más versátil de resolver esa sub-determinación del sistema es recurriendo a la teoría de optimización, pues de esta forma se puede incluso pensar en aplicar el modelo

resultante como modelo predictivo del comportamiento celular. Esta es la estrategia usada por el FBA: resolver la sub-determinación de un modelo metabólico transformando el problema en uno de optimización. Podría objetarse de este enfoque el hecho de que debe suponerse una función objetivo celular que será maximizada o minimizada por el problema de optimización, pues puede ponerse en duda la existencia de tal función objetivo celular, o al menos, que este objetivo pueda representarse de manera única e invariable. Independientemente de tales consideraciones teóricas, en este caso, la optimización resulta ser una poderosa herramienta que permite concretar un modelado matemático bastante útil del problema, obteniéndose un modelo predictivo cuya exactitud dependerá de la calidad del modelo metabólico, de las restricciones impuestas y de la función objetivo usada.

El presente trabajo fue dedicado al análisis del error cometido por la técnica de optimización (FBA), al usar múltiples diferentes funciones objetivo. Representa un avance en el conocimiento sobre el modelado matemático del comportamiento celular en estado estacionario, al evaluar el efecto de diferentes combinaciones de objetivos intracelulares como función objetivo del problema de optimización usado para determinar y resolver el sistema.

Inicialmente, se probaron combinaciones de objetivos de tres compartimentos celulares (citósol, mitocondria, peroxisoma), sin usar pesos relativos. Esta primera aproximación, denominada 'versión inicial' en este trabajo, permitió evidenciar algunos efectos importantes:

- El objetivo 'maximización de producción de biomasa' tiene gran importancia en el modelado por medio de FBA, desde el punto de vista de que se presentó recurrentemente en las combinaciones que produjeron las mejores funciones objetivo.
- No todas las combinaciones que incluían a 'maximización de producción de biomasa' condujeron a funciones objetivo que arrojan buenas predicciones. Esto permitió descartar la posibilidad de que ese objetivo fuera el único que tuviera importancia en las combinaciones que lo incluían (es decir, permitió concluir que los otros objetivos supuestos sí alteran las predicciones del FBA al aparecer en una combinación particular).
- Los objetivos supuestos para el peroxisoma presentaron un efecto prácticamente despreciable en la alteración de los *fluxes* supuestos por el FBA. Lo más probable es que esto se deba al valor relativamente bajo (en comparación con *fluxes* del metabolismo central, por ejemplo) que toman los *fluxes* dentro del peroxisoma.

- La ausencia de conocimiento de la tasa de ingesta de oxígeno ocasiona en el modelo una falta de restricción sobre su valor, lo que a su vez causa que el objetivo ‘maximización de crecimiento celular’ tienda a sobreestimar el crecimiento celular forzando al *flux* de ingesta de oxígeno a tomar valores mayores a los reales.
- La variabilidad de los datos experimentales de diferentes fuentes es aún muy alta; esto se evidenció con las significativas diferencias en los balances de carbono de los diferentes conjuntos de datos, siendo algunos incluso inferiores a un 90% de balance (los cuales no fueron tenidos en cuenta en el trabajo).

Esas observaciones obtenidas a partir de la versión inicial de las combinaciones, condujeron a cambios valiosos para la versión final del trabajo. Como se mencionó en la sección 4.6, se aislaron los objetivos relacionados con producción de biomasa en un compartimento independiente para poderlos combinar con otros objetivos del citosol, se agregaron como posibles objetivos citosolares las minimizaciones de las producciones de metabolitos secundarios, y se consideraron ponderaciones relativas en las combinaciones de objetivos de los compartimentos. No se eliminó el compartimento peroxisoma, a pesar de lo indicado por la versión inicial, porque no podía descartarse la presencia de interacción, o una mayor influencia con algunos de los pesos relativos.

La calidad de las predicciones obtenidas con el FBA depende dramáticamente del conocimiento de la tasa de consumo de oxígeno, lo que se aprecia fácilmente comparando los errores que se cometen con las mejores funciones objetivo para los casos en los cuales el oxígeno es conocido (figuras 10 y 11) con los errores obtenidos cuando la ingesta de oxígeno no se conoce (figuras 12 y 13). Estudios previos han usado el valor del *flux* de oxígeno para evaluar la calidad de las predicciones del FBA, pero utilizando todos los demás datos de los *fluxes* de intercambio como datos de entrada (Knorr *et al.*, 2007). Al utilizar solamente el consumo de oxígeno como criterio de comparación entre la estimación con el FBA y la medición experimental, la validez de la que sea elegida como mejor función objetivo queda limitada a la estimación de la ingesta de oxígeno en experimentos ya realizados. Por otra parte, el presente trabajo se enfocó en la comparación de las estimaciones (tanto de tasa de crecimiento, como de *fluxes* de intercambio) obtenidas con diferentes funciones objetivo en el FBA, pensando en una aplicación predictiva de la técnica y no solamente en la valoración de la cantidad de oxígeno ingerido por el microorganismo en un experimento particular.

Como las predicciones tienen menor calidad cuando no se conoce la tasa de consumo de oxígeno, la ausencia de esta información es una gran barrera para aplicar el FBA como técnica predictiva para modelar el crecimiento aerobio de *S. cerevisiae*, pues puede ser complicado diseñar un proceso de manera que la tasa de ingesta de oxígeno quede fija en un valor determinado, y esto es un requisito para poder implementar experimentalmente una estimación realizada usando un FBA en el cual se fija la tasa de ingesta de oxígeno. En

cualquier caso, para ciertas condiciones experimentales, es posible determinar límites inferior y superior de la tasa de ingesta de oxígeno. En consecuencia, la restricción sobre la ingesta de oxígeno puede hacerse más ajustada, y las predicciones con el FBA pueden mejorarse significativamente.

En la mayoría de clasificaciones estudiadas, las mejores predicciones fueron obtenidas con el objetivo 'maximización de producción de biomasa', y combinaciones de objetivos que lo incluían. Este resultado es acorde con estudios previos que han respaldado el uso de la maximización de crecimiento como función objetivo para el FBA (por ejemplo Edwards *et al.*, 2001; Gianchandani *et al.*, 2008; Ibarra *et al.*, 2002).

Se presentaron dos objetivos que recurrentemente tendieron a aparecer combinados con 'maximización de producción de biomasa' entre las mejores funciones objetivo; estos objetivos son 'minimización de producción de NADH en el citosol' y 'minimización de consumo de NAD(P)H en la mitocondria'. Estos dos objetivos no causaron grandes cambios en las predicciones respecto a las obtenidas con 'maximización de producción de biomasa' sola, sino que las alteraban ligeramente. Esta condición bioquímica ha sido observada en *S. cerevisiae*, en el cual se presenta una tendencia a minimizar el exceso de formación de NADH bajo condiciones anaeróbicas (Albers *et al.*, 1996). El rendimiento de biomasa ha presentado una correlación positiva con respecto a la tasa neta de NADH reoxidado en la respiración y la formación de glicerol, indicando que la rotación (*turnover*) del exceso de NADH de la biosíntesis es un factor importante que influye en el rendimiento de biomasa en condiciones de limitación de oxígeno (Franzén, 2003). La minimización del potencial redox también ha mostrado buenos resultados como objetivo de FBA para *E. coli* (Knorr *et al.*, 2007), de modo que la relevancia de los objetivos relacionados con este propósito no es novedosa.

El caso de crecimiento en fase exponencial en medio aerobio sin conocimiento de la tasa de ingesta de oxígeno es la excepción más importante respecto a la tendencia mencionada previamente. Un estudio previo ya había indicado que el mejor objetivo para el FBA puede ser diferente de acuerdo con la estrategia de cultivo, es decir, si las células crecen en un cultivo por lotes o en un cultivo continuo (Schuetz *et al.*, 2007). En el presente trabajo, la función objetivo 'maximización de producción de biomasa más minimización de producción de NADH en el citosol más minimización de consumo de NAD(P)H en la mitocondria' arrojó estimaciones del comportamiento metabólico celular mucho más precisas que las obtenidas con cualquier otra función objetivo explorada, para los casos en que *S. cerevisiae* crece en fase exponencial en un cultivo por lotes, y no se fija el valor de la ingesta de oxígeno. Con la combinación recién mencionada, los errores en la predicción de la tasa de crecimiento no superaron el 20%, mientras que cualquier otra función probada arrojó errores o del -100% (esto es, las funciones con que no se predecía crecimiento), o superiores al 300%, en la estimación de la tasa específica de crecimiento. Este es un resultado realmente interesante, y parece valioso explorar el desempeño de

esta función objetivo en el FBA usando mayor cantidad de datos, e incluso probándola con otros tipos de células en la misma condición (cultivo por lotes, fase exponencial, ingesta de oxígeno desconocida).

Se ha observado una redistribución amplia de los *fluxes* en condiciones anaerobias respecto a todas las condiciones aerobias (Jouhten *et al.*, 2008). Adicionalmente, bajo condiciones aeróbicas, *S. cerevisiae* regenera el NAD<sup>+</sup> principalmente a través de la respiración celular. Cuando la célula posee una disponibilidad de oxígeno limitada, se restringe la respiración celular, y las células usan otros medios para regenerar el NAD<sup>+</sup>, y el NADH mitocondrial debe ser transportado al citosol para ser re-oxidado. De esta manera se justifica la interrelación de estos diferentes objetivos de los compartimentos para explicar el comportamiento celular (Bakker *et al.*, 2000; Daran-Lapujade *et al.*, 2004).

Finalmente, los resultados no indicaron que diferentes funciones objetivo permitieran un mejor desempeño del FBA dependiendo del tipo de sustrato (si es glucosa o no lo es), lo cual había sido sugerido por un estudio previo (Förster *et al.*, 2003). El único resultado extraño que se evidenció fue el desempeño inesperadamente bueno que se obtuvo con la función 'minimización de consumo de ATP en el citosol más maximización de transporte de ATP hacia citosol en la mitocondria' para los casos en que el sustrato no era glucosa. En cualquier caso, la cantidad de datos experimentales entre ambas categorías fue muy diferente, teniéndose muy pocos datos de crecimiento con sustratos diferentes a (exclusivamente) glucosa. Para obtener mejores predicciones respecto al tipo de sustrato, sería necesario realizar los análisis usando mayor cantidad de datos experimentales de crecimiento con diferentes sustratos, para evaluar la posible existencia de diferencias en la mejor manera de modelar el objetivo celular de acuerdo con, por ejemplo, las rutas en las que se integran los diferentes sustratos a la red metabólica. De la misma manera, la evaluación del efecto de otros factores, como pH o presión osmótica del medio, requiere una mayor cantidad y calidad de datos experimentales.

En la **Tabla 9** se presenta un resumen comparativo de los diferentes estudios que se han dedicado a la comparación del desempeño de diferentes funciones objetivo en el FBA. Se incluyen las propuestas consistentes en generar la función objetivo que más se ajusta a un conjunto de datos. Los resultados de la presente tesis se muestran en la referencia García Sánchez *et al.*, 2012. Hay que tener en cuenta que entre los diferentes estudios hay diferencias en los microorganismos usados, en los modelos metabólicos, en los números de datos experimentales, e incluso en los *fluxes* usado como datos de entrada (valores de los *fluxes* que son fijados en el FBA como restricciones de igualdad) y como datos de salida (valores de los *fluxes* que fueron usados para evaluar qué tan buenas eran las predicciones de cada posible función objetivo).

**Tabla 9.** Resumen de estudios de comparación del desempeño de diferentes funciones objetivo en el FBA.

Referencia	Condiciones y microorganismo		Núm. de datos exp.	Datos de entrada	Datos de salida	Funciones comparadas	Conclusiones
Burgard y Maranas, 2003	Modelo de metabolismo central (62 reacc. y 48 met.) de <i>E. coli</i> , aeróbico, cultivo en quimiostato		1	Ingesta de glucosa	28 <i>fluxes</i> calculados experimentalmente por marcación isotópica	Funciones resultantes de la suma de las reacciones del modelo	La reacción más importante es la producción de biomasa
	Modelo de metabolismo central (62 reacc. y 48 met.) de <i>E. coli</i> , anaeróbico, cultivo en quimiostato		1	Ingesta de glucosa, ingesta de oxígeno	28 <i>fluxes</i> calculados experimentalmente por marcación isotópica	Funciones resultantes de la suma de las reacciones del modelo	La reacción más importante es la producción de biomasa
Knorr et al., 2007	Modelo a escala genómica iJR904 de <i>E. coli</i> , succinato como sustrato, cultivo <i>batch</i> , diferentes temperaturas y concentraciones de succinato		24	Ingesta de succinato	Ingesta de oxígeno y producción de acetato	<ul style="list-style-type: none"> <li>• Maximización de tasa de crecimiento</li> <li>• Minimización de la tasa de producción de potencial redox</li> <li>• Minimización de tasa de producción de ATP</li> <li>• Maximización de tasa de producción de ATP</li> <li>• Minimización de tasa de ingesta de nutrientes</li> </ul>	La mejor función es la minimización de la tasa de producción de potencial redox
Schuetz et al., 2007	Modelo de metabolismo central (98 reacc. y 60 met.) de <i>E. coli</i> ; además de las funciones se probaron simultáneamente 8 restricciones sobre el modelo	Cultivo <i>batch</i> , aerobio, $\mu = 0,6 \text{ h}^{-1}$	1	Ingesta de glucosa, ingesta de oxígeno, excreción de CO <sub>2</sub>	10 <i>fluxes</i> calculados experimentalmente por marcación isotópica	<ul style="list-style-type: none"> <li>• Maximización de tasa de crecimiento</li> <li>• Maximización de rendimiento de ATP</li> </ul>	La mayor fidelidad predictiva se obtuvo con maximización de rendimiento de ATP por <i>flux</i> unitario
		Cultivo <i>batch</i> , anaerobio, $\mu = 0,3 \text{ h}^{-1}$	1	Ingesta de glucosa, ingesta de oxígeno, excreción de CO <sub>2</sub>	10 <i>fluxes</i> calculados experimentalmente por marcación isotópica	<ul style="list-style-type: none"> <li>• Minimización del <i>flux</i> global intracelular</li> <li>• Maximización de rendimiento de ATP por <i>flux</i> unitario</li> </ul>	La mayor fidelidad predictiva se obtuvo con maximización de rendimiento de ATP por <i>flux</i> unitario
		Cultivo <i>batch</i> , anaerobio, $\mu = 0,2 \text{ h}^{-1}$	1	Ingesta de glucosa, ingesta de oxígeno, excreción de CO <sub>2</sub>	10 <i>fluxes</i> calculados experimentalmente por marcación isotópica	<ul style="list-style-type: none"> <li>• Minimización de consumo de glucosa</li> <li>• Minimización de pasos de reacción</li> </ul>	Todas las funciones arrojaron predicciones de calidad comparable.
		Cultivo en quimiostato, limitación en carbono, $\mu = 0,1 \text{ h}^{-1}$	1	Ingesta de glucosa, ingesta de oxígeno, excreción de CO <sub>2</sub>	10 <i>fluxes</i> calculados experimentalmente por marcación isotópica	<ul style="list-style-type: none"> <li>• Maximización de rendimiento de ATP por paso de reacción</li> <li>• Minimización de potencial redox</li> </ul>	Las mejores funciones son maximización de rendimiento de ATP y maximización de producción de biomasa
		Cultivo en quimiostato, li	1	Ingesta de glucosa,	10 <i>fluxes</i> calculados	<ul style="list-style-type: none"> <li>• Minimización de</li> </ul>	Las mejores funciones son

		mitación en carbono, $\mu = 0,4 \text{ h}^{-1}$		ingesta de oxígeno, excreción de $\text{CO}_2$	experimentalmente por marcación isotópica	<i>fluxes</i> productores de ATP • Maximización de <i>fluxes</i> productores de ATP	maximización de rendimiento de ATP y maximización de producción de biomasa
		Cultivo en quimiostato, limitación en nitrógeno, $\mu = 0,4 \text{ h}^{-1}$	1	Ingesta de glucosa, ingesta de oxígeno, excreción de $\text{CO}_2$	10 <i>fluxes</i> calculados experimentalmente por marcación isotópica		Las mejores funciones son maximización de rendimiento de ATP y maximización de producción de biomasa
Gianchandani <i>et al.</i> , 2008	Modelo de metabolismo central (62 reacc. y 60 met.) de <i>S. cerevisiae</i> , cultivo <i>batch</i> , limitación de carbono, $\mu = 0,37 \text{ h}^{-1}$		1	Ingesta de glucosa	Alrededor de 25 <i>fluxes</i> calculados experimentalmente por marcación isotópica	Funciones resultantes de la combinación lineal de las reacciones del modelo	La mejor función objetivo es la maximización de producción de biomasa
García Sánchez <i>et al.</i> , 2012	Modelo de <i>S. cerevisiae</i> compartimentalizado a escala genómica iMM904 (1412 reacc. y 1228 met.)	Cultivo en quimiostato, anaeróbico	6	Ingesta de sustrato, ingesta de oxígeno	Tasa específica de crecimiento, <i>fluxes</i> de intercambio medidos experimentalmente	Todas las combinaciones de objetivos supuestos (en total, 1620 funciones objetivo diferentes)	Las mejores funciones son maximización de biomasa más maximización de etanol, y maximización de biomasa más minimización de $\text{CO}_2$ . Maximización de biomasa y otras dos funciones dan resultados ligeramente peores.
		Cultivo en quimiostato, aeróbico, ingesta de oxígeno conocida	25	Ingesta de sustrato, ingesta de oxígeno			Las mejores funciones son maximización de biomasa, y otras dos combinaciones que la incluyen.
		Cultivo en quimiostato, aeróbico, ingesta de oxígeno conocida, $\mu \leq 0,15 \text{ h}^{-1}$	19	Ingesta de sustrato, ingesta de oxígeno			Las mejores funciones son maximización de biomasa, y otras dos combinaciones que la incluyen.
		Cultivo en quimiostato, aeróbico, ingesta de oxígeno conocida, $0,15 \text{ h}^{-1} \leq \mu < 0,28 \text{ h}^{-1}$	3	Ingesta de sustrato, ingesta de oxígeno			Las mejores funciones consisten en la combinación de maximización de biomasa y minimización de etanol. La predicción es ligeramente superior a la obtenida con maximización de biomasa.
		Cultivo en quimiostato, aeróbico, ingesta de oxígeno	3	Ingesta de sustrato, ingesta de oxígeno			Las mejores funciones son maximización de biomasa, y otras dos combinaciones que

		conocida, $0,28 \text{ h}^{-1} < \mu$					la incluyen.
		Cultivo en quimiostato, aeróbico, ingesta de oxígeno conocida, glucosa como sustrato	20	Ingesta de sustrato, ingesta de oxígeno			Las mejores funciones son maximización de biomasa, y otras dos combinaciones que la incluyen.
		Cultivo en quimiostato, aeróbico, ingesta de oxígeno conocida, sustrato diferente a glucosa	5	Ingesta de sustrato, ingesta de oxígeno			La mejor función es minimización de consumo de ATP en el citosol más maximización de transporte de ATP hacia citosol. Esta función es ligeramente superior a maximización de biomasa y otras dos funciones.
		Cultivo en quimiostato, aeróbico, ingesta de oxígeno desconocida	6	Ingesta de sustrato			La mejor función objetivo es maximización de biomasa más minimización de $\text{CO}_2$ . Las predicciones con esta función son mucho mejores que las obtenidas con maximización de biomasa.
		Cultivo <i>batch</i> en fase exponencial, aeróbico, ingesta de oxígeno desconocida	8	Ingesta de sustrato			La mejor función objetivo es maximización de biomasa más minimización de producción de NADH en citosol más minimización de consumo de NAD(P)H en mitocondria. Las predicciones con esta función son mucho mejores que las obtenidas con maximización de biomasa.

## 6. CONCLUSIONES

1. Las funciones objetivo del FBA pudieron ser representadas por medio de combinaciones de objetivos menores de la célula, con lo cual fue posible interpretar desde un punto de vista bioquímico estas funciones, lo cual permitió mejorar la comprensión del comportamiento celular por medio del análisis de las funciones objetivo que mejor representan las distribuciones de *fluxes* metabólicos.
2. Entre los diferentes objetivos probados, el de mayor importancia en las funciones objetivo que mejor modelaban el comportamiento celular fue la 'maximización de producción de biomasa'. Lo anterior no quiere decir que cualquier combinación que incluya dicho objetivo sea una buena función objetivo para el FBA, sino que en general las funciones que mejor se desempeñaron en cada una de las categorías exploradas eran algunas combinaciones que la incluían.
3. Pensando en aplicar el FBA para predecir el comportamiento celular en diferentes condiciones, lo ideal, desde el punto de vista de modelado de los procesos celulares, es usar como datos de entrada únicamente la ingesta de sustratos y oxígeno. Por consiguiente, en el presente trabajo se usó sólo esta información para crear restricciones igualdad sobre los valores de los *fluxes*, lo cual lo diferenció de los demás estudios sobre adecuación de funciones objetivo para el FBA que se han realizado. Sin embargo, se comprobó a partir de varios conjuntos de datos que el desconocimiento de la tasa de ingesta de oxígeno afecta significativamente el potencial predictivo del FBA. De esta manera, fue posible comprobar que un obstáculo para un buen modelado predictivo del FBA es la correcta determinación de la tasa de ingesta de oxígeno para un conjunto de condiciones ambientales supuestas, dado que se requiere que los *fluxes* que se usen como datos de entrada para el FBA puedan fijarse de antemano a través del montaje experimental, y en algunas situaciones esto puede ser complicado.
4. Para el modelado de crecimiento de *S. cerevisiae* en fase exponencial en un cultivo por lotes sin determinar la tasa de consumo de oxígeno, la función objetivo 'maximización de producción de biomasa más minimización de producción de NADH en el citosol más minimización de consumo de NAD(P)H en la mitocondria' permitió obtener predicciones de una calidad muy superior a las obtenidas con cualquier otra función probada en el presente estudio, siendo el error en la estimación de la biomasa menor al 20%, en comparación con las demás funciones con las que se obtuvieron errores del 100% para aquellas que no predecían crecimiento, o superiores al 300% para las que sobreestimaron el crecimiento (incluyendo maximización de biomasa).

5. Aquellos posibles objetivos relacionados con *fluxes* que no tienen un valor relativamente alto dentro de la célula, no tuvieron impacto significativo en las estimaciones obtenidas por el FBA cuando hacen parte de la función objetivo. Esto se reflejó claramente con los objetivos supuestos del peroxisoma, pero también fue evidente con otros objetivos supuestos en otros compartimentos (como por ejemplo la maximización de producción de acetato, objetivo del citosol), los cuales no alteraban las predicciones obtenidas con el FBA.
  
6. Las ponderaciones relativas probadas entre las diferentes combinaciones de objetivos no tuvieron efecto alguno; esto es, las combinaciones de objetivos usando cualquier conjunto de pesos relativos dieron resultados iguales a los obtenidos con la combinación de objetivos usando una ponderación uniforme para todos estos objetivos. Esto puede atribuirse a una mezcla de varios factores, principalmente a la diferencia en la magnitud de los distintos *fluxes* intracelulares involucrados en los diferentes objetivos supuestos, a la desigualdad en la cantidad de *fluxes* incluidos en los distintos objetivos probados, y al funcionamiento propio de la célula, que causa que los objetivos compartimentales supuestos en realidad tengan niveles de influencia muy diferentes.

## 7. FUTURAS DIRECCIONES

1. La aplicación del enfoque usado en este trabajo a otros tipos de microorganismos o células permitiría evaluar si es posible generalizar las mejores funciones objetivo encontradas para las categorías evaluadas en el presente estudio, a otros tipos de sistemas vivos.
2. Una mayor cantidad y calidad de datos experimentales se requieren para analizar si existen diferencias en el modelado más adecuado del objetivo celular en condiciones ambientales diferentes. De esta manera se podría evaluar la respuesta celular a distintos estímulos externos, como el pH extracelular, la presión osmótica externa o la temperatura.
3. Las combinaciones de objetivos celulares como funciones objetivo para el FBA pueden ser probadas también para analizar el comportamiento celular tras sufrir diferentes alteraciones a nivel genético, ya sean borrados o inserciones. De esta forma se podría analizar si la adaptación del microorganismo a una nueva situación genética puede representarse adecuadamente por medio de relaciones entre objetivos menores de la célula.

## BIBLIOGRAFÍA

- Åkesson, M., Förster, J. y Nielsen, J. 2004. Integration of gene expression data into genome-scale metabolic models. *Metabolic Eng.* 6: 285-293.
- Albe, K. y Wright, B. 1992. Systems analysis of the tricarboxylic acid cycle in *Dictyostelium discoideum*: II. Control analysis. *J. Biol. Chem.* 267: 3106-3114.
- Albers, E., Larsson, C., Lidén, G., Niklasson, C. y Gustafsson, L. 1996. Influence of the nitrogen source on *Saccharomyces cerevisiae* anaerobic growth and product formation. *Appl. Environ. Microbiol.* 62: 3187-3195.
- Albert, R., Jeong, H. y Barabasi, A. L. 2000. Error and attack tolerance of complex networks. *Nature* 406: 378-382.
- Albert, R. y Othmer, H. G. 2003. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in *Drosophila melanogaster*. *J. Theor. Biol.* 223: 1-18.
- Allen, J., Davey, H. M., Broadhurst, D., Rowland, J. J., Oliver, S. G. y Kell, D. B. 2004. Discrimination of the modes of action of antifungal substances by use of metabolic footprinting. *Appl. Env. Micr.* 70: 6157-6165.
- Allen, T. E. y Palsson, B. Ø. 2003. Sequence-based analysis of metabolic demands for protein synthesis in prokaryotes. *J. Theor. Biol.* 220(1): 1-18.
- Almaas, E., Kovács, B., Vicsek, T., Oltvai, Z. N. y Barabási, A.-L. 2004. Global organization of metabolic fluxes in the bacterium *Escherichia Coli*. *Nature* 427: 839-843.
- Almaas, E., Oltvai, Z. N. y Barabási, A.-L. 2005. Metabolic networks, en Vaidyanathan, S., Harrigan, G. G. y Goodacre, R. (eds.), *Metabolome Analyses: Strategies for Systems Biology*. 243-264.
- Aon, J. C. y Cortassa, S. 2001. Involvement of nitrogen metabolism in the triggering of ethanol fermentation in aerobic chemostat cultures of *Saccharomyces cerevisiae*. *Metab. Eng.* 3(3): 250-264.
- Avis, D. y Fukuda, K. 1992. A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra. *Discrete Comput. Geom.* 8(3): 295-313.

- Bai, F. W., Anderson, W. A., y Moo-Young, M. 2008. Ethanol fermentation technologies from sugar and starch feedstocks. *Biotechnol. Adv.* 26: 89-105.
- Bakker, B. M., Bro, C., Kötter, P., Luttik, M. A., van Dijken, J. P. y Pronk J. T. 2000. The mitochondrial alcohol dehydrogenase Adh3p is involved in a redox shuttle in *Saccharomyces cerevisiae*. *J. Bacteriol.* 182(17): 4730-4737.
- Balabanian, N. y Bickart T. A. 1981. Linear network theory: analysis, properties, design, and synthesis. Matrix, Beaverton.
- Baxter, C.J., Redestig, H., Schauer, N., Repsilber, D., Patil, K.R., Nielsen, J., Selbig, J., Liu, J., Fernie, A.R. y Sweetlove, L.J., 2007. The metabolic response of heterotrophic *Arabidopsis* cells to oxidative stress. *Plant Physiol.* 143: 312–325.
- Beard, D. A., Liang, S.-D. y Qian, H. 2002. Energy balance for analysis of complex metabolic networks. *Biosphys. J.* 83(1): 79-86.
- Becker, S. A. y Palsson, B. Ø. 2005. Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: an initial draft to the two-dimensional annotation. *BMC Microbiol.* 5: 8.
- Becker, S. A., Price, N. D. y Palsson, B. Ø. 2006. Metabolite coupling in genome-scale metabolic networks. *BMC Bioinformatics* 7: 111.
- Beste, D. J. V., Hooper, T., Stewart, G., Bonde, B., Avignone-Rossa, C., Bushell, M. E., Wheeler, P., Klamt, S., Kierzek, A. M. y McFadden, J. 2007. GSMN-TB: a web-based genome-scale network model of *Mycobacterium tuberculosis* metabolism. *Genome Biol.* 8: R89.
- Beyer, A., Hollunder, J., Nasheuer, H.-P. y Wilhelm, T. 2004. Post-transcriptional expression regulation in the yeast *Saccharomyces cerevisiae* on a genomic scale. *Mol. Cel. Proteomics* 3(11): 1083-1092.
- Blanch, H. W. 1981. Microbial growth kinetics. *Chem. Eng. Commun.* 8: 181–211.
- Bohnert, H.J. y Sheveleva, E. 1998. Plant stress adaptations – making metabolism move. *Curr. Opin. Plant Biol.* 1: 267–274.
- Borodina, I y Nielsen, J. 2005. From genomes to *in silico* cells via metabolic networks. *Curr. Opin. Biotechnol.* 16:350-355.

Braunstein, A., Mulet, R. y Pagnani, A. 2008. Estimating the size of the solution space of metabolic networks. *BMC Bioinformatics* 9: 240.

Brock, T. D., Madigan, M. T., Martinko, J. y Parker, J. Brock's biology of microorganisms. Benjamin Cummings, 2002.

Burgard, A. P. y Maranas, C. D. 2003. Optimization-based framework for inferring and testing hipotesized metabolic objective functions. *Biotechnol. Bioeng.* 82(6): 670-677.

Burgard, A. P., Pharkya, P. y Maranas, C. D. 2003. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* 84: 647–657.

Çakır, T., Efe, Ç., Dikicioglu, D., Hortacsu, A., Kırdar, B. y Oliver, S. 2007. Flux balance analysis of a genome-scale yeast model constrained by exometabolomic data allows metabolic system identification of genetically different strains. *Biotechnol.Prog.* 23: 320-326.

Cerulo, L., Elkan, C. y Ceccarelli, M. 2010. Learning gene regulatory networks from only positive and unlabeled data. *BMC Bioinformatics* 11: 228.

Cho, D.-Y., Cho, K.-H., Zhang, B.-T. 2006. Identification of biochemical networks by S-tree based genetic programming. *Bioinformatics* 22(13): 1631-1640.

Choi, J. H., Lee, S. J., Lee, S. J. y Lee, S. Y. 2003. Enhanced production of insulin-like growth factor I fusion protein in *Escherichia coli* by coexpression of the down-regulated genes identified by transcriptome profiling. *Appl. Environ.Microbiol.* 69:4737–4742.

Covert, M. W., Knight, E. M., Reed, J. L., Herrgård, M. J. y Palsson, B. Ø. 2004. Integrating high-throughput and computational data elucidates bacterial networks. *Nature* 429:92-96.

Covert, M. W. y Palsson, B. Ø. 2002. Transcriptional regulation in constraint-based metabolic models of *Escherichia coli*. *J. Biol. Chem.* 277:28058-28064.

Covert, M. W., Schilling, C. H. y Palsson, B. Ø. 2001. Regulation of gene expression in flux balance models of metabolism. *J. Theor. Biol.* 213:73-88.

Covert, M. W., Famili, I. y Palsson, B. Ø. 2003. Identifying constraints that govern cellbehavior: a key to converting conceptual to computational models in biology? *Biotechnol. Bioeng.* 84(7): 763–772.

Cox, M. y Nelson, D. L. 2005. *Lehninger Principles of Biochemistry*. Palgrave Macmillan.

- Daran-Lapujade, P., Jansen, M. L. A., Daran, J.-M., van Gulik, W., de Winde, J. H. y Prock, J. T. 2004. Role of transcriptional regulation in controlling fluxes in central carbon metabolism of *Saccharomyces cerevisiae*. *J. Biol. Chem.* 279(10): 9125-9138.
- Desai, M., Antonopoulos, D., Gilbert, J. A., Glass, E. M. y Meyer, F. 2012. From genomics to metagenomics. *Curr. Opin. Biotechnol.* 23(1): 72-76.
- Doble, M., Kruthiventi, A. K. y Gaikar, V. G. 2004. *Biotransformations and bioprocesses*, Marcel Decker.
- Dobson, P. D., Smallbone, K., Jameson, D., Simeonidis, E., Lanthaler, K., Pir, P., Lu, C., Swainston, N., Dunn, W. B., Fisher, P., Hull, D., Brown, M., Oshota, O., Stanford, N. J., Kell, D. B., King, R. D., Oliver, S. G., Stevens, R. D. y Mendes, P. 2010. Further developments towards a genome-scale metabolic model of yeast. *BMC Syst. Biol.* 4: 145.
- Dolinski, K. y Botstein, D. 2005. Changing perspectives in yeast research nearly a decade after the genome sequence. *Genome Res.* 15: 1611-1619.
- Duarte, N. C., Becker, S. A., Jamshidi, N., Thiele, I., Mo, M. L., Vo, T. D., Srivas, R. y Palsson, B. Ø. 2007. Global reconstruction of the humanmetabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. USA* 104: 1777-1782.
- Duarte, N. C., Palsson, B. Ø. y Fu, P. 2004. Integrated analysis of metabolic phenotypes in *Saccaromyces cerevisiae*. *BMC Genomics* 5: 63.
- Edwards, J. S. y Palsson, B. Ø. 1999. Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *J. Biol. Chem.* 274: 17410–17416.
- Edwards, J. S., Ibarra, R. U. y Palsson, B. Ø. 2001. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat. Biotechnol.*, 19:125–130.
- Edwards, J. S. y Palsson, B. 2000. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics and capabilities. *Proc. Natl. Acad. Sci. USA* 97: 5528-5533.
- Edwards, J. S. y Palsson, B. 1998. How will bioinformatics influence metabolic engineering? *Biotechnol. Bioeng.* 58: 162-169.

- Famili, I., Förster, J., Nielsen, J. y Palsson, B. Ø. 2003. *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. PNAS 100(23):13134-13139.
- Feist, A. M. y Palsson, B. Ø. 2010. The biomass objective function. Curr. Opin. Microbiol. 13(3): 344-349.
- Fernie, A.R., Trethewey, R.N., Krotzky, A.J. y Willmitzer, L., 2004. Metabolite profiling: from diagnostics to systems biology. Nat. Rev.Mol. Cell Biol. 5: 763–769.
- Fischer, E. y Sauer, U. 2003. A novel metabolic cycle catalyzes glucose oxidation and anaplerosis in hungry *Escherichia coli*. J. Biol. Chem. 278: 46446–46451.
- Förster, J., Famili, I., Fu, P., Palsson, B. y Nielsen, J. 2003. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. Genome Res. 13:244-253.
- Franzén, C. J. 2003. Metabolic flux analysis of RQ –controlled microaerobic ethanol production by *Saccharomyces cerevisiae*. Yeast 20: 117-132.
- Frick, O. y Wittmann, O. 2005. Characterization of the metabolic shift between oxidative and fermentative growth in *Saccharomyces cerevisiae* by comparative <sup>13</sup>C flux analysis. Microbial Cell Factories 4: 30.
- Friedman, N. 2004. Inferring cellular networks using probabilistic graphical models. Science 303: 799-805.
- García Sánchez, C. E., Vargas García, C. A. y Torres Sáez, R. G. 2012. Predictive potential of Flux Balance Analysis of *Saccharomyces cerevisiae* using as optimization function combinations of cell compartmental objectives. PLoS ONE 7(8): e43006.
- Gevorgyan, A., Poolman, M. G. y Fell, D. A. 2008. Detection of stoichiometric inconsistencies in biomolecular models. Bioinformatics 24(19): 2245-2251.
- Gheshlaghi, R., Scharer, J. M., Moo-Young, M. y Douglas, P. L. 2007. Metabolic flux analysis for optimizing the specific growth rate of recombinant *Aspergillus niger*. Bioprocess Biosyst. Eng. 30: 397-418.
- Gianchandani, E. P., Oberhardt, M. A., Burgard, A. P., Maranas, C. D. y Papin, J. A. 2008. Predicting biological system objectives *de novo* from internal state measurements. BMC Bioinformatics 9: 43.

- Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., et al. 2002. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418: 387–391.
- Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H. y Oliver, S. G. 1996. Life with 6000 genes. *Science* 274: 563–567.
- Goldbeter, A. 1996. Biochemical oscillations and cellular rhythms: the molecular bases of periodic and chaotic behavior. Cambridge Univ. Press, Cambridge.
- Gombert, A. K., Moreira dos Santos, M., Christensen, B. y Nielsen, J. 2001. Network identification and flux quantification in the central metabolism of *Saccharomyces cerevisiae* under different conditions of glucose repression. *J. Bacteriol.* 183(4): 1441-1451.
- Gombert, A. K. y Nielsen, J. 2000. Mathematical modelling of metabolism. *Curr. Opin. Biotechnol.* 11: 180-186.
- Greenbaum, D., Colangelo, C., Williams, K. y Gerstein, M. 2003. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol.* 4: 117.1–117.8.
- Grotkjaer, T., Akesson, M., Christensen, B., Gombert, A. K. y Nielsen, J. 2004. Impact of transamination reactions and protein turnover on labeling dynamics in <sup>13</sup>C-labeling experiments. *Biotechnol. Bioeng.* 86:209–216.
- Gudmundsson, S. y Thiele, I. 2010. Computationally efficient flux variability analysis. *BMC Bioinformatics* 11: 489.
- Guebel, D. V. 2004. Canonical sensitivities: a useful tool to deal with large perturbations in metabolic network modeling. *In Silico Biol.* 4: 163-182.
- Guimera, R. y Amaral, L. A. N. 2005. Functional cartography of complex metabolic networks. *Nature* 433: 895-900.
- Guimera, R., Sales-Pardo, M. y Amaral, L. A. 2007. A network-based method for target selection in metabolic networks. *Bioinformatics* 23: 1616–1622.
- Haag, J. E., Wouwer, A. V. y Bogaerts, P. 2005. Dynamic modeling of complex biological systems: a link between metabolic and macroscopic description. *Math. Biosci.* 193: 25-49.

Hatzimanikatis, V., Floudas, C. y Bailey, J. 1996. Optimization of regulatory architectures in metabolic reaction networks. *Biotechnol. Bioeng.* 52: 485-500.

Heijnen, J. J. 2005. Approximative kinetics formats used in metabolic network modeling. *Biotechnol. Bioeng.* 91(5): 534-545.

Heinrich, R. y Rapoport, T. A. 1974. A linear steady-state treatment of enzymatic chains: general properties, control and effector strength. *Eur. J. Biochem.* 42:89-95.

Heinrich, R. y Schuster, S. 1996. *The regulation of cellular systems.* Chapman & Hall, New York.

Henry, C. S., Broadbelt, L. J. y Hatzimanikatis, V. 2007. Thermodynamics-based metabolic flux analysis. *Biophys. J.* 92: 1792-1805.

Henson, M. A. 2003. Dynamic modeling of microbial cell populations. *Curr. Opin. Biotechnol.* 14: 460-467.

Herrgård, M. J., Lee, B. S., Portnoy, V. y Palsson, B. Ø. 2006. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome Res.* 16: 627-635.

Herrgård, M. J., Swainston, N., Dobson, P., Dunn, W. B., Arga, Y., Arvas, M., Blüthgen, N., Borger, S., Costenoble, R., Heinemann, M., Hucka, M., Le Novère, N., Li, P., Liebermeister, W., Mo, M. L., Oliveira, A. P., Petranovic, D., Pettifer, S., Simeonidis, E., Smallbone, K., Spasić, I., Weichart, D., Brent, R., Broomhead, D. S., Westerhoff, H. V., Kirdar, B., Penttilä, M., Klipp, E., Palsson, B. Ø., Sauer, U., Oliver, S. G., Mendes, P., Nielsen, J. y Kell, D. B. 2008. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nature Biotechnol.* 26(10):1155-1160.

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

Hjersted, J. L., Henson, H. A. y Mahadevan, R. 2006. Genome-scale analysis of *Saccharomyces cerevisiae* metabolism and ethanol production in fed-batch culture. *Biotechnol. Bioeng.* 97(5): 1190-1204.

Ibarra, R. U., Edwards, J. S. y Palsson, B. Ø. 2002. *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature* 420(6912): 186-189.

- Ibarra, R. U., Fu, P., Palsson, B. Ø., DiTonno, J. R. y Edwards, J. S. 2003. Quantitative analysis of *Escherichia coli* metabolic phenotypes within the context of phenotypic phase planes. *J. Mol. Microbiol. Biotechnol.* 6: 101-108.
- Ishii, N., Robert, M., Nakayama, Y., Kanai, A. y Tomita, M. 2004. Toward large-scale modeling of the microbial cell for computer simulation. *J. Biotechnol.* 113: 281-294.
- Ishii, N., Soga, T., Nishioka, T. y Tomita, M. 2005. Metabolome analysis and metabolic simulation. *Metabolomics* 1(1): 29-37.
- Iwatani, S., Yamada, Y. y Usuda, Y. 2008. Metabolic flux analysis in biotechnology processes. *Biotechnol. Lett.* 30: 791-799.
- Iyer, V. V., Sriram, G. y Shanks, J. V. 2007. Metabolic flux maps of central carbon metabolism in plant systems, *en Nikolau, B. J. y Syrkin Wurtele, E. (eds.), Concepts in Plant Metabolomics.* 125 - 144.
- Jamshidi, N. y Palsson, B. Ø. 2007. Investigating the metabolic capabilities of *Mycobacterium tuberculosis* H37Rv using the *in silico* strain iNJ661 and proposing alternative drug targets. *BMC Syst. Biol.* 1:26.
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. y Barabási, A.-L. 2000. The large-scale organization of metabolic networks. *Nature* 407: 651-654.
- Jin, Y.-S. y Jeffries, T. W. 2004. Stoichiometric network constraints on xylose metabolism by recombinant *Saccharomyces cerevisiae*. *Metabolic Eng.* 6: 229-238.
- Jouhten, P., Rintala, E., Huuskonen, A., Tamminen, A., Toivari, M., Wiebe, M., Rouhonen, L., Penttilä, M. y Maaheimo, H. 2008. Oxygen dependence of metabolic fluxes and energy generation of *Saccharomyces cerevisiae* CEN.PK113-1A. *BMC Syst. Biol.* 2: 60.
- Joyce, A. R. y Palsson, B. Ø. 2007. Toward whole cell modeling and simulation: comprehensive functional genomics through the constraint-based approach. *Prog. Drug Res.* 64: 267-309.
- Kauffman, K. J., Prakash, P. y Edwards, J. S. 2003. Advances in flux balance analysis. *Curr. Opin. Biotechnol.* 14: 491-496.
- Kell, D. B. 2006. Systems biology, metabolic modeling and metabolomics in drug discovery and development. *Drug Discovery Today*, 11(23-24): 1085-1092.

- Kikuchi, S., Tominaga, D., Arita, M., Takahashi, K. y Tomita, M. 2003. Dynamic modeling of genetic networks using genetic algorithm and S-system. *Bioinformatics* 19: 643–650.
- Kitano, H. 2004. Biological robustness. *Nature Rev. Genetic* 5: 826-837.
- Klamt, S. y Gilles, E. D. 2004. Minimal cut sets in biochemical reaction networks. *Bioinformatics* 20: 226–234.
- Klamt, S., Schuster, S. y Gilles, E. D. 2002. Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism in purple nonsulfur bacteria. *Biotechnol. Bioeng.* 77(7): 734-751.
- Klamt, S. y Stelling, J. 2003. Two approaches for metabolic pathway analysis? *Trends Biotechnol.* 2003, 21: 64–69
- Kleessen, S. y Nikoloski, Z. 2012. Dynamic regulatory on/off minimization for biological systems under internal temporal perturbations. *BMC Syst. Biol.* 6: 16.
- Klipp, E. 2007. Modelling dynamic processes in yeast. *Yeast* 24: 943-959.
- Klipp, E., Nordlander, B., Kruger, R., Gennemark, P. y Hohmann, S. 2005. Integrative model of the response of yeast to osmotic shock. *Nat. Biotechnol.* 23: 975–982.
- Knorr, A. L., Jain, R. y Srivastava, R. 2007. Bayesian-based selection of metabolic objective functions. *Bioinformatics* 23(3): 351-357.
- Kovářova-Kovar, K. y Egli, T. 1998. Growth kinetics of suspended microbial cells: from single-substrate-controlled growth to mixed-substrate kinetics. *Microbiol. Mol. Biol. Reviews* 62(3): 646-666.
- Koffas, M., Roberge, C., Lee, K., y Stephanopoulos, G. 1999. Metabolic engineering. *Annu. Rev. Biomed. Eng.* 01: 535–557.
- Kresnowati, M. T. A. P., van Winden, W. A., Almering, M. J. H., ten Pierick, A., Ras, C., Knijnenburg, T. A., Daran-Lapujade, P., Pronk, J. T., Heijnen, J. J. y Daran, J. M. 2006. When transcriptome meets metabolome: fast cellular responses of yeast to sudden relief of glucose limitation. *Mol. Syst. Biol.* 2: 49.
- Kuepfer, L., Sauer, U. y Blank, L.M. 2005. Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Res.* 15: 1421–1430.

Kümmel, A., Panke, S. y Heinemann, M. 2006. Systematic assignment of thermodynamics constraints in metabolic network models. *BMC Bioinformatics* 7:512.

Kun, A., Papp, B. y Szathmary, E. 2008. Computational identification of obligatorily autocatalytic replicators embedded in metabolic networks. *Genome Biol.* 9: R51.

Leduc, M., Tikhomiroff, C., Cloutier, M., Perrier, M. y Jolicoeur, M. 2006. Development of a kinetic metabolic model: application to *Catharanthus roseus* hairy root. *Bioprocess Biosyst. Eng.* 28: 295-313.

Lee, D.-Y., Fan, L. T., Park, S., Lee, S. Y., Shafie, S., Bertók, B. y Fiedler, F. 2005. Complementary identification of multiple flux distributions and multiple metabolic pathways. *Metabolic Eng.* 7(3):182-200.

Lee, J. H., Lee, D. E., Lee, B. U. y Kim, H. S. 2003. Global analyses of transcriptomes and proteomes of a parent strain and an L-threonine overproducing mutant strain. *J. Bacteriol.* 185:5442–5451.

Lee, J. M., Gianchandani, E. P. y 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. y Grossmann, I.E., 2000. Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. *Comput. Chem. Eng.* 24: 711–716.

Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I., Zeitlinger, J., Jennings, E. G., Murray, H. L., Gordon, D. B., Ren, B., Wyrick, J. J., Tagne, J.-B., Volkert, T. L., Fraenkel, E., Gifford, D. K. y Young, R. A. 2002. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298: 799-804.

Leonard, E., Fowler, Z. y Koffas, M. 2007. Metabolic Engineering, en *M. Al-Rubeai y M. Fussenegger (eds.), Systems Biology*, 301–359.

Li, F., Long, T., Lu, Y., Ouyang, Q. y Tang, C. 2004. The yeast cell-cycle network is robustly designed. *Proc. Natl. Acad. Sci. USA* 101: 4781-4786.

Lidén, G., Persson, A., Gustafsson, L. y Niklasson, C. 1995. Energetics and product formation by *Saccharomyces cerevisiae* grown in anaerobic chemostats under nitrogen limitation. *Appl. Microbiol. Biotechnol.* 43(6): 1034-1038.

Lindon, J. C., Keun, H. C., Ebbels, T. M., Pearce, J. M., Holmes, E. y Nicholson, J. K. 2005. The Consortium for Metabonomic Toxicology (COMET): aims, activities and achievements. *Pharmacogenomics* 6: 691–699.

Llaneras, F. y Picó, J. 2007. A procedure for the estimation over time of metabolic fluxes in scenarios where measurements are uncertain and/or insufficient. *BMC Bioinformatics*, 8: 421.

Llaneras, F. y Picó, J. 2008. Stoichiometric modelling of cell metabolism. *J. of Bioscience and Bioeng.* 105:1-11.

Mahadevan, R., Edwards, J. S. y Doyle, F. J. 2002. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys. J.* 83:1331-1340.

Mahadevan, R. y Lovley, D. R. 2008. The degree of redundancy in metabolic genes is linked to mode of metabolism. *Biophys. J.* 94: 1216–1220.

Mahadevan, R. y Palsson, B. Ø. 2005. Properties of metabolic networks: structure versus function. *Biophys. J.* 88: L7-L9.

Mahadevan, R. y Schilling, C. H. 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Eng.* 5: 264-276.

Mason, G. F. y Rothman, D. L. 2004. Basic principles of metabolic modeling of NMR <sup>13</sup>C isotopic turnover to determine rates of brain metabolism in vivo. *Metabolic Eng.* 6: 75-84.

Michaelis, L. y Menten, M. L. 1913. Die kinetik der invertinwirkung. *Biochemische Zeitschrift.* 49: 333–369.

Min Lee, J., Gianchandani, E. P., Eddy, J. A. y Papin, J. A. 2008. Dynamic analysis of integrated signaling, metabolic, and regulatory networks. *PLoS Comput. Biol.* 4(5): e1000086.

Mo, M.L., Palsson, B.Ø. y Herrgård, M.J. 2009. Connecting extracellular measurements to intracellular flux states in yeast. *BMC Syst. Biol.* 3: 37.

Montgomery, D. C. y Runger, G. C. *Applied Statistics and Probability for Engineers.* 3 ed., John Wiley and Sons, 2003.

Morgan, J. A. y Rhodes, D. 2002. Mathematical modeling of plant metabolic pathways. *Metabolic Eng.* 4(1):80-89.

- Mori, H. 2004. From the sequence to cell modeling: comprehensive functional genomics in *Escherichia coli*. *J. Biochem. Mol. Biol.* 37(1): 83-92.
- Müller, D., Aguilera-Vázquez, L., Reuss, M., y Mauch, K. 2005. Integration of metabolic and signaling networks, en L. Alberghina y H. V. Westerhoff (Eds.), *Systems Biology* 13: 227-241, Springer Berlin, Heidelberg.
- Nagrath, D., Avila-Elchiver, M., Berthiaume, F., Tilles, A. W., Messac, A. y Yarmush, M. L. 2007. Integrated energy and flux balance based multiobjective framework for large-scale metabolic networks. *Annals Biomed. Eng.* 35(6): 863-885.
- Nash, R., Weng, S., Hitz, B., Balakrishnan, R., Christie, K. R., Costanzo, M. C., Dwight, S. S., Engel, S. R., Fisk, D. G., Hirschman, J. E., Hong, E. L., Livstone, M. S., Oughtred, R., Park, J., Skrzypek, M., Theesfeld, C. L., Binkley, G., Dong, Q., Lane, C., Miyasato, S., Sethuraman, A., Schroeder, M., Dolinski, K., Botstein, D. y Cherry, J. M. 2007. Expanded protein information at SGD: new pages and proteome browser. *Nucleic Acids Res.* 35: D468–D471.
- Newman, M. E. J. y Leicht, E. A. 2007. Mixture models and exploratory analysis in networks. *PNAS* 104(23): 9564-9569.
- Nielsen, J. y Jewett, M. C. 2007. Impact of systems biology on metabolic engineering of *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 8: 122-131.
- Nikerel, I. E., van Winden, W. A., van Gulik, W. M. y Heijnen, J. J. 2006. A method for estimation of elasticities in metabolic networks using steady state and dynamic metabolomics data and linlog kinetics. *BMC Bioinformatics* 7: 540.
- Nissen T., Schulze U., Nielsen J., Villadsen J. 1997. Flux distributions in anaerobic, glucose-limited continuous cultures of *Saccharomyces cerevisiae*. *Microbiology* 143: 203-218.
- Nookaew, I., Jewett, M. C., Meechai, A., Thammamongtham, C., Laoteng, K., Cheevadhanarak, S., Nielsen, J. y Bhumiratana, S. 2008. The genome-scale metabolic model iIN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. *BMC Syst. Biol.* 2: 71.
- Oberhardt, M. A., Puchałka, J., Fryer, K. E., dos Santos, V. A. P. M. y Papin, J. A. 2008. Genome-scale metabolic network analysis of the opportunistic pathogen *Pseudomonas aeruginosa* PAO1. *J. Bacteriol.* 190: 2790–2803.
- Oberhardt, M. A., Palsson, B. Ø. y Papin, J. A. 2009. Applications of genome-scale metabolic reconstructions. *Mol. Syst. Biol.* 5: 320.

- Oh, Y.-K., Palsson, B. Ø., Park, S. M., Schilling, C. H. y Mahadevan, R. 2007. Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *J. Biol. Chem.* 282: 28791–28799.
- Oh, M. K., Rohlin, L., Kao, K. C. y Liao, J. C. 2002. Global expression profiling of acetate-grown *Escherichia coli*. *J. Biol. Chem.* 277:13175–13183.
- Oliver, S. G., van der Aart, Q. J., Agostoni-Carbone, M. L. *et al.* 1992. The complete DNA sequence of yeast chromosome III. *Nature* 357: 38–46.
- Oliver, S. G., Winson, M. K., Kell, D. B. y Baganz, F. 1998. Systematic functional analysis of the yeast genome. *Trends Biotechnol.* 16: 373–378.
- Ostergaard, S., Olsson, L. y Nielsen, J. 2000. Metabolic Engineering of *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. R.* 64: 34-50.
- Österlund, T., Nookaew, I. y Nielsen, J. 2011. Fifteen years of large scale metabolic modeling of yeast: developments and impacts. *Biotechnol. Adv.* 30(5): 979-988.
- Overkamp, K. M., Bakker, B. M., Kotter, P., van Tuijl, A., de Vries, S., van Dijken, J. P., y Pronk, J. T. 2000. In vivo analysis of the mechanisms for oxidation of cytosolic NADH by *Saccharomyces cerevisiae* mitochondria. *J. Bacteriol.* 182: 2823–2830.
- Ow, D. S., Lee, D. Y., Yap, M. G. y Oh, S. K. 2009. Identification of cellular objective for elucidating the physiological state of plasmid-bearing *Escherichia coli* using genome-scale *in silico* analysis. *Biotechnol. Prog.* 25: 61–67.
- Palsson, B. Ø. 2000. The challenges of *in silico* biology. *Nat. Biotechnol.*, 18: 1147–1150.
- Palsson, B. Ø. 2006. *Systems biology: properties of reconstructed networks*. Cambridge University Press New York, New York.
- Papin, J. A., Hunter, T., Palsson, B. Ø. y Subramaniam, S. 2005. Reconstruction of cellular signalling networks and analysis of their properties. *Nat. Rev. Mol. Cell. Biol.* 6(2): 99–111.
- Papin, J. A., Price, N. D. y Palsson, B. Ø. 2002. Extreme pathway lengths and reaction participation in genome-scale metabolic networks. *Genome Res.* 12: 1889–1900.
- Papin, J. A., Price, N. D., Wiback, S. J., Fell, D. A. y Palsson, B. Ø. 2003. Metabolic pathways in the post-genome era. *Trends Biochem. Sci.* 28: 250–258.

- Park, J. H., Lee, S. Y., Kim, T. Y. y Kim, H. U. 2008. Application of systems biology for bioprocess development. *Trends in Biotechnol.* 26(8): 404-412.
- Park, S. J., Lee, S. Y., Cho, J., Kim, T. Y., Lee, J. W., Park, J. H. y Han, M.-J. 2005. Global physiological understanding and metabolic engineering of microorganisms based on omics studies. *Appl. Microbiol. Biotechnol.* 68: 567-579.
- Patil, K. R., Rocha, I., Förster, J. y Nielsen, J. 2005. Evolutionary programming as a platform for *in silico* metabolic engineering. *BMC Bioinformatics* 6: 308.
- Pham, T. K., Chong, P. K., Gan, C. S. y Wright, P. C. 2006. Proteomic analysis of *Saccharomyces cerevisiae* under high gravity fermentation conditions. *J. of Proteome Research* 5: 3411-3419.
- Picard, F.J. y Bergeron, M. G. 2002. Rapid molecular theranostics in infectious diseases. *Drug Discov. Today* 7: 1092–1101.
- Pitt, J. J., Eggington, M. y Kahler, S. G. 2002. Comprehensive screening of urine samples for inborn errors of metabolism by electrospray tandem mass spectrometry. *Clin. Chem.* 48: 1970–1980.
- Polisetty, P. K., Voit, E. O. y Gatzke, E. P. 2006. Identification of metabolic system parameters using global optimization methods. *Theoret. Biol. and Med. Modelling* 3:4.
- Pramanik, J. y Keasling, J. D. 1997. Stoichiometric model of *Escherichia coli* metabolism: Incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. *Biotechnol. Bioeng.* 56: 398–421.
- Price, N. D., Reed, J. L. y Palsson, B. Ø. 2004. Genome-scale models of microbial cells:evaluating the consequences of constraints. *Nat. Rev. Microbiol.* 2(11): 886–897.
- Price, N. D., Reed, J. L., Papin, J. A., Famili, I. y Palsson, B. Ø. 2003. Analysis of metabolic capabilities using singular value decomposition of extreme pathway matrices. *Biophys. J.* 84: 794-804.
- Price, N. D. y Shmulevich, I. 2007. Biochemical and statistical network models for systems biology. *Curr. Opin. Biotechnol.* 18: 365-370.
- Provost, A. y Bastin, G. 2004. Dynamic metabolic modelling under the balanced growth condition. *J. Process Contr.* 14: 717-728.

- Puchałka, J., Oberhardt, M. A., Godinho, M., Bielecka, A., Regenhardt, D., Timmis, K. N., Papin, J. A. y dos Santos, V. A. P. M. 2008. Genome-scale reconstruction and analysis of the *Pseudomonas putida* KT2440 metabolic network facilitates applications in biotechnology. *PLoS Comput. Biol.* 4(10): e1000210.
- Qian, H. 2001a. Mathematical formalism for isothermal linear irreversibility. *Proc. R. Soc. Lond. A.* 457:1645–1655.
- Qian, H. 2001b. Mesoscopic nonequilibrium thermodynamics of single macromolecules and dynamic entropy-energy compensation. *Phys. Rev. E.* 65:016102: 1–5.
- Qian, H. y Beard, D. A. 2005. Thermodynamics of stoichiometric biochemical networks in living systems far from equilibrium. *Biophysical Chem.* 114: 213-220.
- Raab, R. M., Tyo, K. y Stephanopoulos, G. 2005. Metabolic engineering. *Adv. Biochem. Engin./Biotechnol.* 100: 1-17.
- Ramakrishna, R., Edwards, J. S., McCulloch, A., Palsson, B.O. 2001. Flux balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 280:R695-R704.
- Raman, K. y Chandra, N. 2009. Flux balance analysis of biological systems: applications and challenges. *Briefings in Bioinformatics* 10(4):453-449.
- Reed, J. L. y Palsson, B. Ø. 2003. Thirteen years of building constraintbased *in silico* models of *Escherichia coli*. *J. Bacteriol.* 185: 2692-2699.
- Reed, J. L., Vo, T. D., Schilling, C. H. y Palsson, B. Ø. 2003. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol.* 4: R54.
- Rodríguez, J., Kleerebezem, R., Lema, J. M. y van Loosdrecht, M. C. M. 2005. Modeling product formation in anaerobic mixed culture fermentations. *Biotechnol. Bioeng.* 93(3): 592-606.
- Sainz, J., Pizarro, F., Pérez-Correa, J. R. y Agosin, E. 2002. Modeling of yeast metabolism and process dynamics in batch fermentation. *Biotechnol. Bioeng.* 81(7): 818-828.
- Samal, A., Singh, S., Giri, V., Krishna, S., Raghuram, N. y Jain, S. 2006. Low degree metabolites explain essential reactions and enhance modularity in biological networks. *BMC Bioinformatics* 7: 118.

- Sauer, U. 2006. Metabolic networks in motion:  $^{13}\text{C}$ -based flux analysis. *Mol. Syst. Biol.* 2: 62.
- Savageau, M. A. Biochemical system analysis. 1976. Addison-Wesley Publishing Company, Reading.
- Savinell, J. M. y Palsson, B. Ø. 1992. Network analysis of intermediary metabolism using linear optimization. II. Interpretation of hybridoma cell metabolism. *J. Theor. Biol.* 154: 455–473.
- Schaffter, T., Marbach, D., y Floreano, D. 2011. GeneNetWeaver: *In silico* benchmark generation and performance profiling of network inference methods. *Bioinformatics*, 27(16): 2263-2270.
- Schellenberger J., Que R., Fleming R. M. T., Thiele I., Orth J. D., Feist A. M., Zielinski D. C., Bordbar A., Lewis N. E., Rahmanian S., Kang J., Hyduke D. R., Palsson B. O. 2011. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc.* 6: 1290-1307.
- Schilling, C. H., Edwards, J. S., Letscher, D. y Palsson, B. Ø. 2000a. Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnol. Bioeng.* 71: 286-306.
- Schilling, C. H., Letscher, D. y Palsson, B. Ø. 2000b. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theor. Biol.* 203: 229-248.
- Schilling, C. H. y Palsson, B. Ø. 2000. Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. *J. Theor. Biol.* 203: 249-283.
- Schilling, C. H., Schuster, S., Palsson, B. Ø. y Heinrich, R. 1999. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol. Prog.* 15: 296–303.
- Schilling, C. H., Covert, M. W., Famili, I., Church, G. M., Edwards, J. S. Y Palsson, B. Ø. 2002. Genome-scale metabolic model of *Helicobacter pylori* 26695. *J. Bacteriol.* 184(16): 4582-4593.
- Schoondermark-Stolk, S. A., Jansen, M., Veurink, J. H., Verkleij, A. J., Verrips, C. T., Euverink, G.-J. W., Boonstra, J. y Dijkhuizen, L. 2006. Rapid identification of target genes

for 3-methyl-1-butanol production in *Saccharomyces cerevisiae*. Appl. Microbiol. Biotechnol. 70: 237-246.

Schuetz, R., Kuepfer, L. y Sauer, U. 2007. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. Mol. Syst. Biol. 3: 119.

Schuster, S., Dandekar, T. y Fell, D. A. 1999. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. Trends Biotech. 17: 53-60.

Schuster, S., Pfeiffer, T., Moldenhauer, F., Koch, I. y Dandekar, T. 2002. Exploring the pathway structure of metabolism: decomposition into subnetworks and application to *Mycoplasma pneumoniae*. Bioinformatics 18: 351-361.

Schwartz, J.-M. y Kanehisa, M. 2006. Quantitative elementary mode analysis of metabolic pathways: the example of yeast glycolysis. BMC Bioinformatics 7: 186.

Schweder, T. y Hecker, M. 2004. Monitoring of stress responses. Adv. Biochem. Eng. Biotechnol. 89: 47-71.

Segrè, D., Vitkup, D. y Church, G. M. 2002. Analysis of optimality in natural and perturbed metabolic networks. PNAS 99: 15112-15117.

Sendín, J-O. H., Alonso, A. A. y Banga, J. R. 2009. Multi-objective optimization of biological networks for prediction of intracellular fluxes, en *IWPACBB 2008, J. M. Corchado et al. (Eds.):* 197-205.

Serre, V., Penverne, B., Souciet, J.-L., Potier, S., Guy, H., Evans, D., Vicart, P. y Hervé, G. 2004. Integrated allosteric regulation in the *S. cerevisiae* carbamylphosphate synthetase – aspartate transcarbamylase multifunctional protein. BMC Biochemistry 5: 6.

Shinto, H., Tashiro, Y., Yamashita, M., Kobayashi, G., Sekiguchi, T., Hanai, T., Kuriya, Y., Okamoto, M. y Sonomoto, K. 2007. Kinetic modeling and sensitivity analysis of acetone-butanol-ethanol production. J. Biotechnol. 131: 45-56.

Shiraishi, F. y Savageau, M. 1992. The tricarboxylic acid cycle in *Dictyostelium discoideum*: III. Analysis of steady state and dynamic behaviour. J. Biol. Chem. 267: 22926-22933.

Shlomi, T., Berkman, O. y Ruppin, E. 2005. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. PNAS 102: 7695-7700.

- Shlomi, T., Eisenberg, Y., Sharan, R., y Ruppín, E. 2007. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol. Syst. Biol.* 3:101.
- Smallbone, K., Simeonidis, E., Broomhead, D. S. y Kell, D. B. 2007. Something from nothing -bridging the gap between constraint-based and kinetic modelling. *FEBS J.* 274: 5576-5585.
- Smid, E. J., van Enckevort, F. J. H., Wegkamp, A., Boekhorst, J., Molenaar, D., Hugenholtz, J., Siezen, R. J. y Teusink, B. 2005. Metabolic models for rational improvement of lactic acid bacteria as cell factories. *J. Appl. Microbiol.* 98: 1326-1331.
- Steinmetz, L. M., Scharfe, C., Deutschbauer, A. M., Mokranjac, D., Herman, Z. S., Jones, T., Chu, A. M., Giaever, G., Prokisch, H., Oefner, P. J., et al. 2002. Systematic screen for human disease genes in yeast. *Nat. Genet.* 22: 22.
- Stelling, J. 2004. Mathematical models in microbial systems biology. *Curr. Opin. Microbiol.* 7: 513-518.
- Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S. y Gilles, E. D. 2002. Metabolic network structure determines key aspects of functionality and regulation. *Nature* 420: 190-193.
- Stephanopoulos, G. N., Aristidou, A. A. y Nielsen, J. *Metabolic Engineering: Principles and Methodologies.* Academic Press, Orlando, 1998.
- Steuer, R., Gross, T., Selbig, J. y Blasius, B. 2006. Structural kinetic modeling of metabolic networks. *PNAS* 103(32): 11868-11873.
- Streit, W. R. y Schmitz, R. A. 2004. Metagenomics - the key to the uncultured microbes. *Curr. Opin. Microbiol.* 7: 492-498.
- Tai, S. L., Boer, V. M., Daran-Lapujade, P., Walsh, M. C., de Winde, J. H., Daran, J. M. y Pronk, J. T. 2005. Two-dimensional transcriptome analysis in chemostat cultures. Combinatorial effects of oxygen availability and macronutrient limitation in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 280(1): 437-447.
- Tai, S. L., Daran-Lapujade, P., Luttik, M. A., Walsh, M. C., Diderich, J. A., Krijger, G. C., van Gulik, W. M., Pronk, J. T. y Daran, J. M. 2007. Control of the glycolytic flux in *Saccharomyces cerevisiae* grown at low temperature: a multi-level analysis in anaerobic chemostat cultures. *J. Biol. Chem.* 282(14): 10243-10251.

Takors, R., Wiechert, W. y Weusters-Botz, D. 1997. Experimental design for the identification of macrokinetic models and model discrimination. *Biotechnol. Bioeng.* 56: 564-567.

Terzer, M., Maynard, N. D., Covert, M. y Stelling, J. 2009. Genome-scale metabolic networks. *WIREs Syst. Biol. Med.* 1(3): 285-297.

Teusink, B., Passarge, J., Reijenga, C. A., Esgalhado, E., van der Weijden, C. C., Schepper, M., Walsh, M. C., Bakker, B. M., van Dam, K., Westerhoff, H. V. y Snoep, J. L. 2000. Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. *Eur. J. Biochem.* 267: 5313-5329.

The UniProt Consortium. 2008. The universal protein resource (UniProt). *Nucleic Acids Res.* 36: D190–D195.

Thiele, I. y Palsson, B. Ø. 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* 5(1): 93-121.

Tominaga, D., Koga, N. y Okamoto, M. 2000. Efficient numerical optimization algorithm based on genetic algorithm for inverse problem, in *Proceedings of the 2000 Genetic and Evolutionary Computation Conference*. Morgan Kaufmann: 251–258.

Torres, N. V. 1994. Modelling approach to control of carbohydrate metabolism during citric acid accumulation by *Aspergillus niger*: I. Model definition and stability of the steady state. *Biotechnol. Bioeng.* 44: 04-111.

Usaite, R., Patil, K. R., Grotkjaer, T., Nielsen, J. y Regenberg, B. 2006. Global transcriptional and physiological responses of *Saccharomyces cerevisiae* to ammonium, L-alanine, or L-glutamine limitation. *Appl. Environ. Microbiol.* 72(9): 6194-6203.

Vaidyanathan, S. *et. al.* (eds). 2006. *Metabolome analyses: strategies for systems biology*, Springer.

Vallino, J. J., Stephanopoulos, G. 1993. Metabolic flux distributions in *Corynebacterium glutamicum* during growth and lysine overproduction. *Biotechnol. Bioeng.* 41: 633–645.

van Gulik, W. y Heijnen, J. 1995. A metabolic network stoichiometry analysis of microbial growth and product formation. *Biotechnol. Bioeng.* 48: 681–698.

van Hoek, P., van Dijken, J. P. y Pronk, J. T. 1998. Effect of specific growth rate on fermentative capacity of baker's yeast. *App. Env. Microbiol.* 64(11): 4226-4233.

van Winden, W. A., van Dam, J. C., Ras, C., Kleijn, R. J., Vinke, J. L., van Gulik, W. M. y Heijnen, J. J. 2005. Metabolic-flux analysis of *Saccharomyces cerevisiae* CEN.PK113-7D based on mass isotopomer measurements of <sup>13</sup>C-labeled primary metabolites. *FEMS Yeast Res.* 5: 559-568.

Vargas García, C. E. 2012. Estimación de funciones objetivo de redes metabólicas basada en compartimentalización y análisis de balance de flujos. Tesis de Maestría en Ingeniería de Sistemas, Universidad Industrial de Santander.

Varma, A. y Palsson, B. Ø. 1994. Metabolic flux balancing: basic concepts, scientific and practical use. *Biotechnol.* 12: 994-998.

Varma, A., Boesch, B. W. y Palsson, B. Ø. 1993. Biochemical production capabilities of *Escherichia coli*. *Biotechnol. Bioeng.* 42: 59-73.

Velagapudi, V. R., Wittmann, C., Schneider, K. y Heinzle, E. 2007. Metabolic flux screening of *Saccharomyces cerevisiae* single knockout strains on glucose and galactose supports elucidation of gene function. *J. Biotechnol.* 132: 395-404.

Vemuri, G. N., Eiteman, M. A. McEwen, J. E., Olsson, L. y Nielsen, J. 2007. Increasing NADH oxidation reduces overflow metabolism in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 104(7): 2402-2407.

Visser, D. y Heijnen, J. J. 2002. The mathematics of metabolic control analysis revisited. *Metabolic Eng.* 4: 114-123.

Visser, D. y Heijnen, J. J. 2003. Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics. *Metabolic Eng.* 5: 164-176.

Voit, E. O. 2000. Computational analysis of biochemical systems. Cambridge University Press, Cambridge, UK.

von Lieres, E., Petersen, S. y Wiechert, W. 2004. A multi-scale modeling concept and computational tools for the integrative analysis of stationary metabolic data. *J. Integrative Bioinf.* 1(1): 4.

Wang, Q., Yang, Y., Ma, H. y Zhao, X. 2007. Metabolic network properties help assign weights to elementary modes to understand physiological flux distributions. *Bioinformatics* 23(9): 1049-1052.

- Wiback, S., Famili, I., Greenberg, H. J. y Palsson, B. Ø. 2004. Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space. *J. Theor. Biol.* 228: 437–447.
- Wiback, S. J., Mahadevan, R. y Palsson, B. Ø. 2003. Reconstructing metabolic flux vectors from extreme pathways: defining the alpha-spectrum. *J. Theor. Biol.* 224: 313–324.
- Wiebe, M.G., Rintala, E., Tamminen, A., Simolin, H., Salusjärvi, L., Toivari, M., Kokkonen, J. T., Kiuru, J., Ketola, R. A., Jouhten, P., Huuskonen, A., Maaheimo, H., Ruohonen, L. y Penttilä, M. 2008. Central carbon metabolism of *Saccharomyces cerevisiae* in anaerobic, oxygen-limited and fully aerobic steady-state conditions and following a shift to anaerobic conditions. *FEMS Yeast Res.* 8: 140–154.
- Wiechert, W. 2002. Modeling and simulation: tools for metabolic engineering. *J. Biotech.* 94: 37-63.
- Wiechert, W. y de Graaf, A. A. 1996. *In vivo* stationary flux analysis by  $^{13}\text{C}$  labelling experiments. *Adv. Biochem. Eng. Biotechnol.* 54:109-154.
- Wiechert, W., Mollney, M., Petersen, S. y de Graaf, A.A., 2001. A universal framework for  $^{13}\text{C}$  metabolic flux analysis. *Metabolic Eng.* 3: 265–283.
- Wright, J. y Wagner, A. 2008. Exhaustive identification of steady state cycles in large stoichiometric networks. *BMC Syst. Biol.* 2: 61.
- Xiong, M., Zhao, J. y Xiong, H. 2004. Network-based regulatory pathways analysis. *Bioinformatics* 20(13): 2056-2066.
- Xu, X., Cao, L. y Chen, X. 2008. Elementary flux mode analysis for optimized ethanol yield in anaerobic fermentation of glucose with *Saccharomyces cerevisiae*. *Chin. J. Chem. Eng.* 16(1): 135-142.
- Yin, K. K., Yang, H., Daoutidis, P. y Yin, G. G. 2001. Simulation of population dynamics using continuous-time finite-state Markov chains. *Comput. Chem. Eng.* 27: 235-249.
- Young, J. D., Henne, K. L., Morgan, J. A., Konopka, A. E. y Ramkrishna, D. 2008. Integrating cybernetic modeling with pathway analysis provides a dynamic, systems-level description of metabolic control. *Biotechnol. Bioeng.* 100(3): 542-559.
- Zamboni, N., Fischer, E. y Sauer, U. 2005. FiatFlux—a software for metabolic flux analysis from  $^{13}\text{C}$ -glucose experiments. *Bioinformatics* 6: 209.

Zhang, H. y Yao, S. 2007. Simulation of flux distribution in central metabolism of *Saccharomyces cerevisiae* by hibridized genetic algorithm. Chin. J. Chem. Eng. 15(2): 150-156.

Zhao, J., Yu, H., Luo, J., Cao, Z. W. y Li, Y. 2006. Complex networks theory for analizing metabolic networks. Chin. Sci. Bulletin 51(13): 1529-1537.

Zomorodi, A. R. y Maranas, C. D. 2010. Improving the iMM904 *S. cerevisiae* metabolic model using essentiality and synthetic lethality data. BMC Syst. Biol. 4: 178.

## ANEXOS

Anexo A. Conjuntos de datos experimentales usados en el estudio.

Tipo de cultivo	Sustrato	Tasa de crecimiento específico $\mu$	$v_{\text{sustrato}}$	$v_{\text{O}_2}$	$v_{\text{etanol}}$	$v_{\text{CO}_2}$	$v_{\text{glicerol}}$	$v_{\text{acetato}}$	$v_{\text{succinato}}$	$v_{\text{acetaldehído}}$	$v_{\text{piruvato}}$	Referencia
Continuo	Glucosa	0,1	3,002	ND	3,632	ND	0,2111	0,0078	0,0132	0,0048	0,0005	Çakır <i>et al.</i> , 2007
Continuo	Glucosa	0,1	5,5550	0	8,2825	9,0658	0,0768	0,0333	0,0250	ND	0,0111	Nissen <i>et al.</i> , 1997
Continuo	Glucosa	0,2	11,5022	0	17,1153	18,8406	2,0934	0,1035	0,0518	ND	0,046	Nissen <i>et al.</i> , 1997
Continuo	Glucosa	0,3	17,3667	0	25,7374	27,8214	3,2997	0,3126	0,0521	ND	0,1042	Nissen <i>et al.</i> , 1997
Continuo	Glucosa	0,4	23,6541	0	35,2682	37,0423	5,1566	0,7096	0,071	ND	0,1892	Nissen <i>et al.</i> , 1997
Continuo	Glucosa	0,1	1,17	ND	0	ND	0	0	0	ND	0	Gombert <i>et al.</i> , 2001
Continuo	Glucosa	0,1	1,15	2,74	ND	2,85	ND	ND	ND	ND	ND	Daran-Lapujade <i>et al.</i> , 2004
Continuo	Maltosa	0,1	0,61	3,05	ND	3,05	ND	ND	ND	ND	ND	Daran-Lapujade <i>et al.</i> , 2004
Continuo	Etanol	0,1	3,78	6,87	---	3,26	ND	ND	ND	ND	ND	Daran-Lapujade <i>et al.</i> , 2004
Continuo	Acetato	0,1	5,89	7,40	ND	7,45	ND	---	ND	ND	ND	Daran-Lapujade <i>et al.</i> , 2004
Continuo	Glucosa	0,15	1,56	ND	0	3,6075	0,0325	0	ND	ND	ND	Frick y Wittmann, 2005
Continuo	Glucosa	0,3	4,90	ND	2,4239	10,3213	0,2606	0,7037	ND	ND	ND	Frick y Wittmann, 2005
Continuo	Glucosa	0,4	8,23	ND	8,8025	13,7405	0,5725	0,4294	ND	ND	ND	Frick y Wittmann, 2005
Continuo	Glucosa más etanol	0,1	0,94	2,87	0,40 **	2,64	ND	ND	ND	ND	ND	van Winden <i>et al.</i> , 2005
Continuo	Galactosa	0,0970	1,06	ND	0	2,62	0	0	ND	ND	0,0005	Ostergaard <i>et al.</i> , 2000
Continuo	Glucosa	0,1	1,28	2,70	0	ND	0	0	ND	ND	ND	Jouhten <i>et al.</i> ,

												2008
Continuo	Glucosa	0,1	1,36	2,50	0,07	ND	0	0	ND	ND	ND	Jouhten <i>et al.</i> , 2008
Continuo	Glucosa	0,1	1,28	2,50	0,10	ND	0	0	ND	ND	ND	Jouhten <i>et al.</i> , 2008
Continuo	Glucosa	0,1	1,97	1,70	1,56	ND	0	0	ND	ND	ND	Jouhten <i>et al.</i> , 2008
Continuo	Glucosa	0,1	2,13	1,70	2,00	ND	0	0	ND	ND	ND	Jouhten <i>et al.</i> , 2008
Continuo	Glucosa	0,1	2,78	1,20	2,91	ND	0	0	ND	ND	ND	Jouhten <i>et al.</i> , 2008
Continuo	Glucosa	0,1	6,30	0	9,05	ND	1,05	0	ND	ND	ND	Jouhten <i>et al.</i> , 2008
Continuo	Glucosa	0,1	6,58	0	9,47	ND	1,11	0	ND	ND	ND	Jouhten <i>et al.</i> , 2008
Continuo	Glucosa	0,1	1,10	2,70	0	2,60	ND	ND	ND	ND	ND	Wiebe <i>et al.</i> , 2008
Continuo	Glucosa	0,1	1,33	2,50	0,10	3,00	ND	ND	ND	ND	ND	Wiebe <i>et al.</i> , 2008
Continuo	Glucosa	0,1	1,90	1,70	1,60	3,70	ND	ND	ND	ND	ND	Wiebe <i>et al.</i> , 2008
Continuo	Glucosa	0,1	2,38	1,20	2,75	4,60	ND	ND	ND	ND	ND	Wiebe <i>et al.</i> , 2008
Batch (crecimiento exponencial)	Glucosa	0,4	20,2	ND	30,0	34,4	1,9	1,0	ND	ND	ND	Heyland <i>et al.</i> , 2009
Batch (crecimiento exponencial)	Glucosa	0,16	7,2	ND	9,0	15,4	0,6	0,7	ND	ND	ND	Heyland <i>et al.</i> , 2009
Batch (crecimiento exponencial)	Glucosa	0,4	19,9	ND	29,6	33,8	1,7	1,3	ND	ND	ND	Heyland <i>et al.</i> , 2009
Batch (crecimiento exponencial)	Glucosa	0,36	18,4	ND	28,2	32,0	1,5	1,3	ND	ND	ND	Heyland <i>et al.</i> , 2009
Batch (crecimiento exponencial)	Glucosa	0,21	12,3	ND	15,2	24,0	2,6	2,1	ND	ND	ND	Heyland <i>et al.</i> , 2009
Batch (crecimiento exponencial)	Glucosa	0,17	10,2	ND	11,2	20,6	2,4	1,7	ND	ND	ND	Heyland <i>et al.</i> , 2009
Batch (crecimiento exponencial)	Glucosa	0,33	15,1	ND	20,1	25,5	2,9	1,0	ND	ND	ND	Heyland <i>et al.</i> , 2009
Batch (crecimiento exponencial)	Glucosa	0,23	12,2	ND	15,6	23,1	2,8	0,9	ND	ND	ND	Heyland <i>et al.</i> , 2009

Continuo	Glucosa	0,025	0,3	0,8	0	0,8	0	0	ND	ND	0	van Hoek <i>et al.</i> , 1998
Continuo	Glucosa	0,05	0,6	1,3	0	1,4	0	0	ND	ND	0	van Hoek <i>et al.</i> , 1998
Continuo	Glucosa	0,1	1,1	2,5	0	2,7	0	0	ND	ND	0	van Hoek <i>et al.</i> , 1998
Continuo	Glucosa	0,15	1,7	3,9	0	4,2	0	0	ND	ND	0	van Hoek <i>et al.</i> , 1998
Continuo	Glucosa	0,2	2,3	5,3	0	5,7	0	0	ND	ND	0	van Hoek <i>et al.</i> , 1998
Continuo	Glucosa	0,25	2,8	7,0	0	7,5	0	0	ND	ND	0	van Hoek <i>et al.</i> , 1998
Continuo	Glucosa	0,28	3,4	7,4	0,11	8,0	0	0,08	ND	ND	0,01	van Hoek <i>et al.</i> , 1998
Continuo	Glucosa	0,3	4,5	6,1	2,3	8,8	0	0,41	ND	ND	0,01	van Hoek <i>et al.</i> , 1998
Continuo	Glucosa	0,35	8,6	5,1	9,5	14,9	0,05	0,62	ND	ND	0,03	van Hoek <i>et al.</i> , 1998
Continuo	Glucosa	0,4	11,1	3,7	13,9	18,9	0,15	0,6	ND	ND	0,05	van Hoek <i>et al.</i> , 1998

Cultivo, tipo de sustrato, valores de los *fluxes* de intercambio medidos, y fuente de los conjunto de datos usados en el estudio. La tasa específica de crecimiento se indica en  $h^{-1}$ , mientras que los *fluxes* se muestran en unidades de  $mmol/(gDW \cdot h)$ . Los *fluxes* de las columnas  $v_{sustrato}$  y  $v_{O_2}$  corresponden a ingesta de metabolitos, mientras el resto de las columnas de *fluxes* corresponden a excreción de metabolitos.

ND: No hay información en la referencia acerca del valor de este *flux*.

\* : El valor del *flux* no es mostrado acá, debido a que el metabolito es sustrato y por consiguiente su *flux* aparece en la columna

$v_{sustrato}$ .

\*\* : Este *flux* de intercambio corresponde a ingesta de etanol, no a producción de etanol.

Anexo B. Vectores de pesos relativos  $\vec{w}_i$  usados para generar todas las funciones objetivo probadas.

$\vec{w}_i$			
0	0	0	0
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1
1	1	0	0
1	0	1	0
1	0	0	1
0	1	1	0
0	1	0	1
0	0	1	1
1	1	1	0
1	1	0	1
1	0	1	1
0	1	1	1
1	1	1	1
4	1	0	0
4	0	1	0
4	0	0	1
1	4	0	0
1	0	4	0
1	0	0	4
0	4	1	0
0	4	0	1
0	0	4	1
0	1	4	0
0	1	0	4
0	0	1	4
3	2	0	0
3	0	2	0
3	0	0	2
2	3	0	0
2	0	3	0
2	0	0	3
0	3	2	0
0	3	0	2
0	0	3	2

0	2	3	0
0	2	0	3
0	0	2	3
3	7	0	0
3	0	7	0
3	0	0	7
7	3	0	0
7	0	3	0
7	0	0	3
0	3	7	0
0	3	0	7
0	0	3	7
0	7	3	0
0	7	0	3
0	0	7	3
5	1	1	0
5	1	0	1
5	0	1	1
0	5	1	1
1	5	1	0
1	5	0	1
1	0	5	1
0	1	5	1
1	1	5	0
1	1	0	5
1	0	1	5
0	1	1	5
1	3	3	0
1	3	0	3
1	0	3	3
0	1	3	3
3	1	3	0
3	1	0	3
3	0	1	3
0	3	1	3
3	3	1	0
3	3	0	1
3	0	3	1
0	3	3	1
1	2	3	0
1	3	2	0
2	3	1	0
2	1	3	0

3	1	2	0
3	2	1	0
1	2	0	3
1	3	0	2
2	1	0	3
2	3	0	1
3	1	0	2
3	2	0	1
1	0	2	3
1	0	3	2
2	0	1	3
2	0	3	1
3	0	1	2
3	0	2	1
0	1	2	3
0	1	3	2
0	2	1	3
0	2	3	1
0	3	1	2
0	3	2	1
7	1	1	1
1	7	1	1
1	1	7	1
1	1	1	7
4	4	1	1
4	1	4	1
4	1	1	4
1	4	4	1
1	4	1	4
1	1	4	4
3	3	3	1
3	1	3	3
3	3	1	3
1	3	3	3
1	2	3	4
1	2	4	3
1	3	2	4
1	3	4	2
1	4	2	3
1	4	3	2
2	1	3	4
2	1	4	3
2	3	1	4

2	3	4	1
2	4	1	3
2	4	3	1
3	2	1	4
3	2	4	1
3	1	2	4
3	1	4	2
3	4	2	1
3	4	1	2
4	2	3	1
4	2	1	3
4	3	2	1
4	3	1	2
4	1	2	3
4	1	3	2

Las filas muestran los vectores de ponderación relativa usados; la primera, segunda, tercera y cuarta columna corresponden a la ponderación de los compartimentos “global”, “citosol”, “mitocondria” y “peroxisoma”, respectivamente.

Cualquiera de estos vectores  $\vec{w}_i$  multiplicado por una constante daría un vector diferente que generaría una función objetivo distinta (múltiplo de la original), con un valor funcional óptimo diferente, pero con los mismos valores de variables de decisión (es decir, la misma distribución de *fluxes*) al ser usada como objetivo en el FBA. En el estudio realizado, lo importante son las distribuciones de *fluxes* obtenidas con cada objetivo y no específicamente el valor de la función objetivo en el óptimo; por lo anterior, se usaron vectores de ponderación relativa compuestos de números enteros.

Anexo C. Contenido del trabajo *Computational Flux Balance Analysis (FBA) of new representative objective functions using a multiple compartmental objective approach and its application to Saccharomyces cerevisiae biological behavior*, presentado en el evento *ISCB Latin America Conference on Bioinformatics*, celebrado en Santiago de Chile en marzo de 2012.

## **Computational Flux Balance Analysis (FBA) of new representative objective functions using a multiple compartmental objective approach and its application to *Saccharomyces cerevisiae* biological behavior**

### **Introduction**

The most used technique to obtain quantitative predictions of metabolic flux distributions of microorganisms under different environmental and genetic conditions has been flux balance analysis (FBA)<sup>(1)</sup>. FBA assumes that cells attempt to achieve an optimal metabolic flux distribution owing to evolutionary pressure<sup>(2)</sup>, therefore a good FBA prediction depends essentially on the selected cellular objective. Most analyses have assumed that maximization of cell growth (i.e. biomass production) is the most appropriated objective function<sup>(3-8)</sup>. However, it has been shown that not always optimization based on cell growth is the best option<sup>(e.g. 6, 9, 10)</sup>. Some previous studies have worked on the development and analysis of different types of objective functions from experimental data; examples are the method of coefficients of importance (*CoIs*)<sup>(11)</sup>, and the BOSS method<sup>(12)</sup>. Nevertheless, in these approaches the chosen objective functions are very dependent on the set of particular data, and are not easily interpretable from a physiological point of view. In this work, a new kind of objective functions for the FBA are presented and analyzed. The objective functions explored consist in linear combinations of multiple minor cellular objectives, coming from different cellular compartments. The prokaryotic cells don't have multiple compartments, so the approach presented here is valid for eukaryotic cells. For that reason, the quantitative evaluation of the predictions obtained with the FBA using the different considered objective functions was made with growing experiments data of *Saccharomyces cerevisiae*.

### **Methods**

- Experimental data

Experimental data used in this work were obtained from previous studies with *S. cerevisiae*<sup>(13-23)</sup>. The carbon balance of every data set was evaluated, and sets of data with error higher than 10% in the mass balance were discarded. 46 sets of experimental data

passed the balance. These data represent several experimental conditions, including diverse oxygen availability, substrate uptake and measured fluxes.

- Metabolic model

The iMM904<sup>(24)</sup>, a compartmentalized genomic model, was used as metabolic model of *S. cerevisiae*. This model includes 1577 enzyme reactions and 1228 metabolites, distributed in 8 compartments: extracellular space, cytosol, mitochondrion, peroxisome, nucleus, endoplasmic reticulum, Golgi apparatus and vacuole<sup>(24)</sup>.

- Tested objective functions

Initially, the five compartments with higher number of reactions (cytosol, mitochondrion, peroxisome, nucleus and Golgi apparatus) were considered as contributors to the global cellular objective. After some preliminary tests it was decided to have into account only the first three compartments. The evaluated FBA's objective functions consisted in combinations  $Z = F1 + F2 + F3$ , where **F1**, **F2** and **F3** are possible objectives of cytosol, mitochondrion and peroxisome, respectively. Table 1 shows the possible objectives considered for every compartment. In every case one of the possible objectives is "None", taking into account that the compartment may not contribute to obtaining a good prediction.

Compartment	Cytosol	Mitochondrion	Peroxisome	
Possible objective functions	max (biomass production)	max (NADH+NADPH production)	max (ATP production)	max (fatty acids production)
	min (biomass production)	min (NADH+NADPH production)	max (NADH+NADPH prod.)	max (ATP consumption)
	max (acetate production)	max (NADH+NADPH consumption)	max (NADH+NADPH cons.)	min (ATP consumption)
	max (CO <sub>2</sub> production)	min (NADH+NADPH consumption)	max(ATP transport towards cytosol)	max (NADH+NADPH prod.)
	max (ethanol production)	max (ATP production)	None	min (NADH+NADPH cons.)
	max (glycerol production)	max (ATP consumption)		None
	max (succinate production)	min (ATP consumption)		
		None		

**Table 1.** Possible objectives whose linear combinations as FBA objective were evaluated.

- Evaluation of performance of the objective functions in FBA

For every experimental condition, the values of substrate uptake and oxygen exchange (if available) were fixed. Subsequently, an FBA was carried out, using every linear combination (with weightings equal to one) of the possible compartment objectives. The

predictions obtained about the cell growth and the exchange fluxes were compared with the experimental data.

- Design of the algorithm

The simulations were made with Matlab (©The Mathworks, Inc.) software and the COBRA toolbox<sup>(25)</sup>. To solve the linear programming problem was used the glpk solver, using glpkmex as interface with Matlab.

## Results

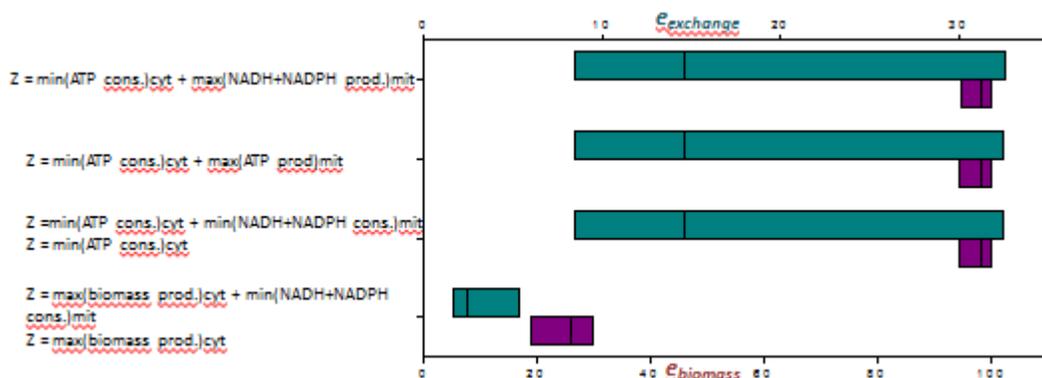
Experimental data were classified in three categories: anaerobic growth, aerobic growth with measured oxygen uptake, and aerobic growth without measured oxygen exchange.

The error in the prediction of growth rate was calculated as the percentage of relative error between the experimental and the predicted specific growth rate (called  $e_{biomass}$ ), and the error in the prediction of the known exchange fluxes was measured with the euclidean distance between the vectors of experimental data and predicted values (called  $e_{exchange}$ ).

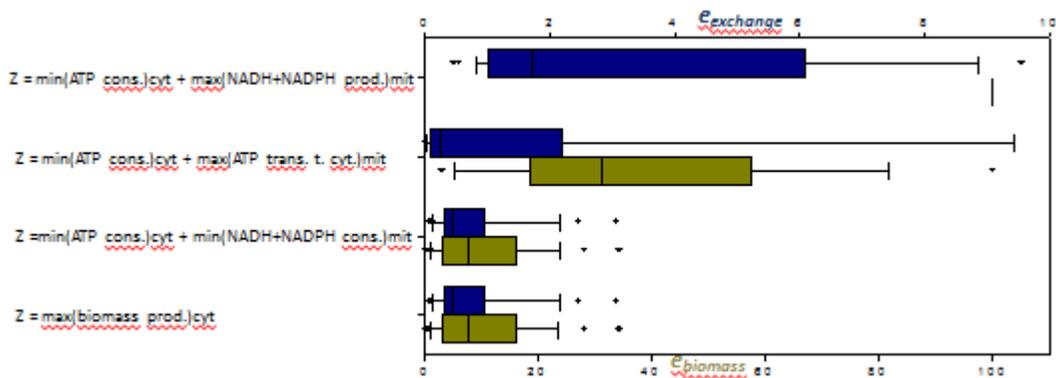
The results showed that, for all cases, the peroxisome objectives didn't contribute to the FBA estimations (i.e., the estimated values were the same using any of the peroxisome possible objectives for all the combinations of cytosol and mitochondrion objectives), so in the graphs the compartment 'peroxisome' is not mentioned.

When the oxygen exchange flux is known (i.e., anaerobic and aerobic with measured oxygen flux – **Figures 1 and 2**) the maximization of growth is the best cytosol objective, and for the mitochondrion it is the minimization of consumption of NADH and NADPH. Other objectives led to significantly worst predictions.

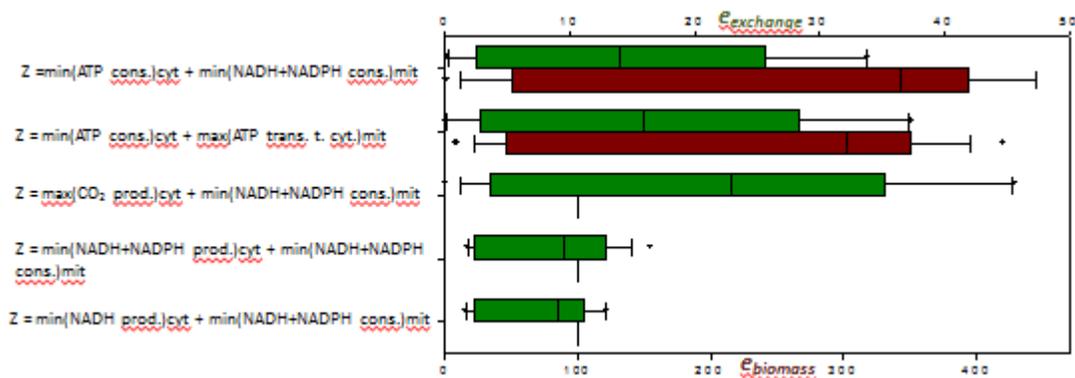
In the other hand, when the oxygen flux is unknown (**Figure 3**), the maximization of growth causes a high error in the growth prediction, overestimating it. In those cases, minimization of objectives relative to the reductive power of the cell led to better estimations.



**Figure 1.** Combinations of compartmental objectives that led to better FBA estimations in **anaerobic growth**. The boxes represent the  $e_{biomass}$  (dark pink boxes) and the  $e_{exchange}$  (dark cyan boxes).



**Figure 2.** Combinations of compartmental objectives that led to better FBA estimations in **aerobic growth, with known oxygen exchange flux**. The boxes represent the  $e_{biomass}$  (dark yellow boxes) and the  $e_{exchange}$  (dark blue boxes).



**Figure 3.** Combinations of compartmental objectives that led to better FBA estimations in **aerobic growth, with unknown oxygen exchange flux**. The boxes represent the  $e_{biomass}$  (dark red boxes) and the  $e_{exchange}$  (green boxes).

## Conclusions

- Maximization of growth is a very important compartmental objective, in the search of obtaining good FBA estimations using the kind of objective functions explored.
- Obtaining good estimations when the oxygen uptake rate is unknown requires the presence of some compartmental objectives that limit the estimation of cell growth that is fueled for the objective 'maximization of growth'.
- Considered peroxisome objectives don't alter the FBA predictions.
- The variability of the errors in the estimations is significantly high (i.e., with different experimental data, quality of the predictions is very different).

All these conclusions are very valuable for the next stages of the study (improving the compartmentalized objective functions and upgrading the algorithm).

## Bibliography

- (1) Papin J A, Price ND, Wiback SJ, Fell DA and Palsson BO. 2003. *TIBS*, 28:250–258.
- (2) Llaneras F and Picó J. 2008. *J. Biosc. Bioeng.* 105:1-11.
- (3) Edwards JS, Ibarra RU and Palsson BO. 2001. *Nat. Biotechnol.*, 19:125–130.
- (4) Edwards JS and Palsson BO. 2000. *Proc. Natl. Acad. Sci. USA* 97:5528-5533.
- (5) Förster J, Famili I, Fu P, Palsson B and Nielsen J. 2003. *Genome Res.* 13:244-253.
- (6) Ibarra RU, Edwards JS and Palsson BO. 2002. *Nature* 420(6912):186-9.
- (7) Schilling CH, Covert MW, Famili I, Church GM, Edwards JS and Palsson BO. 2002. *J. Bacteriol.* 184(16):4582-4593.
- (8) Almaas E, Kovács B, Vicsek T, Oltvai ZN and Barabási A-L. 2004. *Nature* 427:839-843.
- (9) Ramakrishna R, Edwards JS, McCulloch A and Palsson BO. 2001. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 280:R695-R704.
- (10) Schuetz R, Kuepfer L and Sauer U. 2007. *Mol. Syst. Biol* 3:119.
- (11) Burgard AP and Maranas CD. 2003. *Biotechnol. Bioeng.* 82(6):670-677.
- (12) Gianchadani EP, Oberhardt MA, Burgard AP, Maranas CD and Papin JA. 2008. *BMC Bioinformatics*, 9:43.
- (13) Çakır T, Efe Ç, Dikicioglu D, HortaÇsu A, Kırdar B and Oliver SG. 2007. *Biotechnol. Prog.* 23:320-326.
- (14) Nissen T, Schulze U, Nielsen J and Villadsen J. 1997. *Microbiology*, 143:203-218.
- (15) Gombert AK, dos Santos MM, Christensen B and Nielsen J. 2001. *J. Bacteriol.* 183(4):1441–1451.
- (16) Daran-Lapujade P, Jansen MLA, Daran J-M., van Gulik, W, de Winde JH and Pronk JT. 2004. *J Biol Chem.* 279(10):9125-9138.
- (17) Frick O and Wittmann C. 2005. *Microbial Cell Factories* 30(4), doi:10.1186/1475-2859-4-30.
- (18) van Winden WA, van Dam JC, Ras C, Kleijn RJ, Vinke JL, van Gulik, WM and Heijnen JJ. 2005. *FEMS Yeast Research* 5:559-568.
- (19) Ostergaard S, Olsson L and Nielsen, J. 2001. *Biotechnol. Bioeng.* 73(5):412-425.
- (20) Jouhten P, Rintala E, Huuskonen A, Tamminen A, Toivari M, Wiebe M, Ruohonen L, Penttilä M and Maaheimo H. 2008. *BMC Systems Biology* 2:60.
- (21) Wiebe MG, Rintala E, Tamminen A, Simolin H, Salusjärvi L, Toivari M, Kokkonen JT, Kiuru J, Ketola RA, and other 5 authors. 2007. *FEMS Yeast Res.* 8, 140-154.
- (22) Heyland J, Fu J and Blank LM. 2009. *Microbiology* 155:3827-3837. DOI 10.1099/mic.0.030213-0.
- (23) van Hoek P, van Dijken JP and Pronk JT. 1998. *Appl. Environ. Microbiol.* 64(11):4226-4233.
- (24) Mo ML, Palsson BO and Herrgård MJ. 2009. *BMC Systems Biology* 3:37, doi:10.1186/1752-0509-3-37.
- (25) Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, and other 4 authors. 2011. *Nat. Protoc.*, 6, 1290-1307.

Anexo D. Distinción obtenida por el trabajo *Computational Flux Balance Analysis (FBA) of new representative objective functions using a multiple compartmental objective approach and its application to Saccharomyces cerevisiae biological behavior*.



Anexo E. Primera página del artículo *Predictive Potential of Flux Balance Analysis of Saccharomyces cerevisiae Using as Optimization Function Combinations of Cell Compartmental Objectives*, divulgado en la publicación *PLoS One*.

El artículo completo puede accederse en la siguiente dirección electrónica:

<http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0043006>

## Predictive Potential of Flux Balance Analysis of *Saccharomyces cerevisiae* Using as Optimization Function Combinations of Cell Compartmental Objectives

Carlos Eduardo García Sánchez<sup>1\*</sup>, César Augusto Vargas García<sup>2</sup>, Rodrigo Gonzalo Torres Sáez<sup>3</sup>

**1** Escuela de Ingeniería Química, Grupo de Investigación en Bioquímica y Microbiología, Universidad Industrial de Santander, Bucaramanga, Santander, Colombia, **2** Escuela de Ingeniería de Sistemas e Informática, Grupo de Investigación en Ingeniería Biomédica, Universidad Industrial de Santander, Bucaramanga, Santander, Colombia, **3** Escuela de Química, Grupo de Investigación en Bioquímica y Microbiología, Universidad Industrial de Santander, Bucaramanga, Santander, Colombia

### Abstract

**Background:** The main objective of flux balance analysis (FBA) is to obtain quantitative predictions of metabolic fluxes of an organism, and it is necessary to use an appropriate objective function to guarantee a good estimation of those fluxes.

**Methodology:** In this study, the predictive performance of FBA was evaluated, using objective functions arising from the linear combination of different cellular objectives. This approach is most suitable for eukaryotic cells, owing to their multiplicity of cellular compartments. For this reason, *Saccharomyces cerevisiae* was used as model organism, and its metabolic network was represented using the genome-scale metabolic model IMM904. As the objective was to evaluate the predictive performance from the FBA using the kind of objective function previously described, substrate uptake and oxygen consumption were the only input data used for the FBA. Experimental information about microbial growth and exchange of metabolites with the environment was used to assess the quality of the predictions.

**Conclusions:** The quality of the predictions obtained with the FBA depends greatly on the knowledge of the oxygen uptake rate. For the most of studied classifications, the best predictions were obtained with "maximization of growth", and with some combinations that include this objective. However, in the case of exponential growth with unknown oxygen exchange flux, the objective function "maximization of growth, plus minimization of NADH production in cytosol, plus minimization of NAD(P)H consumption in mitochondrion" gave much more accurate estimations of fluxes than the obtained with any other objective function explored in this study.

**Citation:** García Sánchez CE, Vargas García CA, Torres Sáez RG (2012) Predictive Potential of Flux Balance Analysis of *Saccharomyces cerevisiae* Using as Optimization Function Combinations of Cell Compartmental Objectives. *PLoS ONE* 7(8): e43006. doi:10.1371/journal.pone.0043006

**Editor:** Peter Carmaly, Semmelweis University, Hungary

**Received:** April 4, 2012; **Accepted:** July 16, 2012; **Published:** August 9, 2012

**Copyright:** © 2012 García Sánchez et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** CEGS and CAVG received funding from a scholarship of the Universidad Industrial de Santander. Development of the study was supported by the Grupo de Investigación en Bioquímica y Microbiología of the Universidad Industrial de Santander. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: carlos.garcia@correouis.edu.co

### Introduction

Gradual development on genetic manipulation techniques has opened great possibilities for alteration of microorganisms for different purposes. These approaches have ranged from improvements and developments in the production of several metabolites, to multiple biochemical and microbiological investigations [1]. Since early developments in this field, the need for global analysis of cellular systems was evident, because interaction between cellular components does not allow cell functions to be explained simply by characterizing the components comprised in it [2].

This environment led to the emergence of metabolic engineering, which is a combination of systematic analysis from different cellular networks (metabolic, signaling, etc.) with molecular biology techniques to improve cellular properties through rational design and the implementation of genetic modifications [1]. Among the areas studied by metabolic engineering, one of the

most relevant fields is searching for techniques to quantitatively predict the metabolic behavior of microorganisms under different conditions. In this category, the most widely used mathematical modeling approach has been flux balance analysis (FBA) [3].

FBA is based on the assumption that evolutionary pressure has led to the redirection of cellular metabolic fluxes, seeking for an optimal distribution according to a certain cellular goal [4]. This assumption make it possible to solve (i.e. to find a flux distribution based on) the underdetermined system that results from a mass balance in steady state of the intracellular metabolites [3], shown in equation (1), transforming the issue into the optimization problem of the equation (2).

$$S + \bar{v} = 0 \quad (1)$$

Anexo F. Primera página del artículo *Balance de Flujos Metabólicos en Saccharomyces cerevisiae basado en Compartimentalización Intracelular*, aceptado para publicación en la *Revista Colombiana de Biotecnología*.

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

César Augusto Vargas García<sup>1</sup>, Carlos Eduardo García Sanchez<sup>2</sup>, Henry Arguello Fuentes<sup>3</sup>,  
Rodrigo Gonzalo Torres Sáez<sup>4</sup>

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 batch. 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.

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 outperform in 98% growth predictions and 70% ethanol predictions compared with growth maximization.

<sup>1</sup> Ms(c) Ingeniería de Sistemas e Informática, Universidad Industrial de Santander, caugusto.vargas@gmail.com.

<sup>2</sup> Escuela de Ingeniería Química, Grupo de Investigación en Bioquímica y Microbiología, Universidad Industrial de Santander, Bucaramanga, Santander, Colombia

<sup>3</sup> PhD(c) Electrical and Computer Engineering, Docente Asistente Escuela de Ingeniería de Sistemas e Informática, Universidad Industrial de Santander. henarfu@uis.edu.co

<sup>4</sup> PhD Bioquímica, Docente Asistente Facultad de Ciencias Básicas, Universidad Industrial de Santander, rtorres@uis.edu.co

Anexo G. Primera página del artículo *Exploration of the effect of input data on the modeling of cellular objective in Flux Balance Analysis (FBA)*, aceptado para publicación en *Advances in Computational Biology: Proceedings of the Second Colombian Computational Biology Congress*.

## Exploration of the effect of input data on the modeling of cellular objective in Flux Balance Analysis (FBA)

Carlos Eduardo García Sánchez<sup>1</sup>, Rodrigo Gonzalo Torres Sáez<sup>2</sup>

<sup>1</sup>Universidad Industrial de Santander, Escuela de Ingeniería Química, Bucaramanga, Colombia

<sup>2</sup>Universidad Industrial de Santander, Escuela de Química, Bucaramanga, Colombia

carlos.garcia6@correo.uis.edu.co

rtorres@uis.edu.co

**Abstract.** Flux Balance Analysis (FBA) is a technique that allows estimation of metabolic fluxes in established conditions, focusing the flux determination as an optimization problem. For this reason, it is important to use an appropriate objective function in order to adjust estimations of FBA with the real behaviour of the cell. The aim of this work was to examine the effect of a set of input data fluxes (among five different sets) on the accuracy of predictions obtained with FBA with application in seven different objective functions. In this study, *Saccharomyces cerevisiae* was selected as model microorganism, and its metabolism was represented at genomic scale using the model iMM904. Accuracy of obtained predictions was evaluated and compared with eight set of experimental data. Results showed that an objective function representing in a better way the cellular behaviour depends on the set of fluxes used as input data.

**Keywords:** Flux Balance Analysis, input data, objective functions, estimation accuracy.

### 1 Introduction

Flux Balance Analysis (FBA) is a technique of mathematical modeling, which allows determining a distribution of fluxes in a steady state through maximization of a supposed objective function [1]. Obtaining of a suited estimation by using of FBA requires a good metabolic model, a set of adjusted restrictions to real conditions and an objective function that generates realistic results. In other words, a mathematical function should model a cellular objective [1-5]. It also requires a set of input data that allows a feasible region for the model data [6]. Nevertheless, both quality of the ob-

Anexo H. Listado de reacciones involucradas en cada uno de los objetivos de los compartimentos celulares propuestos, en el modelo iMM904. Se presenta, para cada objetivo de cada compartimento, el número de la reacción en el modelo y la reacción completa en la que se incluye el(los) metabolito(s) que conforman el objetivo.

El algoritmo está escrito para ejecutar maximizaciones. Por consiguiente, para representar la maximización de la producción de un metabolito, se suman todas las reacciones en las cuales se produce el metabolito, cada una multiplicada por el coeficiente estequiométrico del metabolito en la reacción. Para representar minimizaciones, se lleva a cabo el mismo procedimiento de identificación de reacciones, pero a todas las reacciones en la combinación se les antepone un signo negativo, de modo que el algoritmo –escrito para maximizar- las minimice.

### Participación de los metabolitos en las funciones objetivo en ‘global’

#### Biomasa

1	(1.1348) 13BDgln + (0.4588) ala-L + (0.046) amp + (0.1607) arg-L + (0.1017) asn-L + (0.2975) asp-L + (59.276) atp + (0.000001) camp + (0.000001) chitin + (0.0447) cmp + (0.000001) coa + (0.0066) cys-L + (0.0036) damp + (0.0024) dcmp + (0.0024) dgmp + (0.0036) dtmp + (0.0007) ergst + (0.000001) fad + (0.1054) gln-L + (0.3018) glu-L + (0.2904) gly + (0.5185) glycogen + (0.046) gmp + (0.000001) gthrd + (59.276) h2o + (0.0663) his-L + (0.1927) ile-L + (0.2964) leu-L + (0.2862) lys-L + (0.8079) mannan + (0.0507) met-L + (0.000001) nad + (0.000006) pa_SC + (0.00006) pc_SC + (0.000045) pe_SC + (0.1339) phe-L + (0.000001) pHEME + (0.1647) pro-L + (0.000017) ps_SC + (0.000053) ptd1ino_SC + (0.000001) q6 + (0.00099) ribflv + (0.1854) ser-L + (0.02) so4 + (0.000001) thf + (0.000001) thmp + (0.1914) thr-L + (0.0234) tre + (0.000066) triglyc_SC + (0.0284) trp-L + (0.102) tyr-L + (0.0599) ump + (0.2646) val-L + (0.0015) zymst --> (59.276) adp + (58.70001) h + (59.305) pi
---	---

### Participación de los metabolitos en las funciones objetivo en citosol

#### Producción NADH+NADPH

#### <nadh>

174	2hb + nad <==> 2obut + h + nadh
349	etoh + nad --> acald + h + nadh
352	etoh + nad <==> acald + h + nadh
353	h2o + nad + pacald --> (2) h + nadh + pac
357	aproa + h2o + nad --> ala-B + (2) h + nadh
435	btd-RR + nad <==> actn-R + h + nadh
443	4mzym_int1 + nad --> 4mzym_int2 + co2 + h + nadh
444	nad + zym_int1 --> co2 + h + nadh + zym_int2
696	fald + gthrd + nad <==> Sfglutth + h + nadh
741	for + nad --> co2 + nadh
801	g3p + nad + pi <==> 13dpg + h + nadh
841	glu-L + h2o + nad --> akg + h + nadh + nh4

981	h2o + imp + nad --> h + nadh + xmp
1003	3c2hmp + nad --> 3c4mop + h + nadh
1022	h2o + lald-L + nad --> (2) h + lac-L + nadh
1060	mal-L + nad <==> h + nadh + oaa
1121	mlthf + nad --> methf + nadh
1125	h2o + n4abutn + nad --> 4aabutn + (2) h + nadh
1254	3pg + nad --> 3php + h + nadh
1311	nad + pphn --> 34hpp + co2 + nadh
1411	h2o + nad + saccrp-L <==> akg + h + lys-L + nadh
1416	nad + sbt-D --> fru + h + nadh
1417	nad + sbt-L --> h + nadh + srb-L
1570	nad + xyln --> h + nadh + xyln-D
929	h2o + histd + (2) nad --> (3) h + his-L + (2) nadh

<nadph>

355	h2o + id3acald + nadp --> (2) h + ind3ac + nadph
359	acald + h2o + nadp --> ac + (2) h + nadph
383	arab-D + nadp --> Ara14lac + h + nadph
419	athr-L + nadp <==> 2aobut + h + nadph
789	g6p + nadp --> 6pgl + h + nadph
858	glyc + nadp --> dha + h + nadph
870	6pgc + nadp --> co2 + nadph + ru5p-D
963	icit + nadp --> akg + co2 + nadph
1037	nadp + ser-L <==> 2amsa + h + nadph
1120	mlthf + nadp <==> methf + nadph
1312	nadp + pphn --> 34hpp + co2 + nadph
1428	dscl + nadp --> h + nadph + scl
1453	h2o + nadp + sucsal --> (2) h + nadph + succ
938	coa + mev-R + (2) nadp <==> (2) h + hmgcoa + (2) nadph
1463	(3) h2o + h2s + (3) nadp <==> (5) h + (3) nadph + so3

Consumo NADH+NADPH

<nadh>

233	L2aadp + atp + h + nadh --> L2aadp6sa + amp + nad + ppi
335	2mbald + h + nadh --> 2mbtoh + nad
338	2mppal + h + nadh --> ibutoh + nad
341	3mbald + h + nadh --> iamoh + nad
344	h + nadh + pacald --> 2phetoh + nad
347	h + id3acald + nadh --> ind3eth + nad
350	acald + h + nadh --> etoh + nad
441	ergtrol + h + nadh + o2 --> ergtetrol + (2) h2o + nad

754	fmn + h + nadh --> fmnh2 + nad
779	dhap + h + nadh --> glyc3p + nad
787	glu5p + h + nadh --> glu5sa + nad + pi
846	akg + gln-L + h + nadh --> (2) glu-L + nad
944	1p3h5c + (2) h + nadh --> 4hpro-LT + nad
948	aspsa + h + nadh --> hom-L + nad
1128	h[c] + nadh[c] + q6[m] --> nad[c] + q6h2[m]
1033	(2) h + lanost + (3) nadh + (3) o2 --> 44mctr + for + (4) h2o + (3) nad
<nadph>	
200	3dsphgn + h + nadph --> nadp + sphgn
232	L2aadp + atp + h + nadph --> L2aadp6sa + amp + nadp + ppi
294	(0.01) 1agly3p_SC + h + nadph --> (0.01) 1ag3p_SC + nadp
334	glyald + h + nadph --> glyc + nadp
337	2mbald + h + nadph --> 2mbtoh + nadp
340	2mppal + h + nadph --> ibutoh + nadp
343	3mbald + h + nadph --> iamoh + nadp
346	h + nadph + pacald --> 2phetoh + nadp
384	arab-L + h + nadph --> abt + nadp
394	4pasp + h + nadph --> aspsa + nadp + pi
439	44mctr + h + nadph --> 44mzym + nadp
440	ergtrol + h + nadph + o2 --> ergtetrol + (2) h2o + nadp
445	4mzym_int2 + h + nadph --> 4mzym + nadp
446	h + nadph + zym_int2 --> nadp + zymst
449	epist + h + nadph + o2 --> ergtrol + (2) h2o + nadp
467	cer1_24 + h + nadph + o2 --> cer2_24 + h2o + nadp
468	cer1_26 + h + nadph + o2 --> cer2_26 + h2o + nadp
473	cer1_24 + h + nadph + o2 --> cer2'_24 + h2o + nadp
474	cer1_26 + h + nadph + o2 --> cer2'_26 + h2o + nadp
479	cer2_24 + h + nadph + o2 --> cer3_24 + h2o + nadp
480	cer2_26 + h + nadph + o2 --> cer3_26 + h2o + nadp
562	h + nadph + o2 + tdcoa --> (2) h2o + nadp + tdecoa
563	h + nadph + o2 + pmtcoa --> (2) h2o + hdcoa + nadp
564	h + nadph + o2 + stcoa --> (2) h2o + nadp + odecoa
565	h + nadph + o2 + odecoa --> (2) h2o + nadp + ocdycocoa
574	dhf + h + nadph --> nadp + thf
603	2dhp + h + nadph --> nadp + pant-R
606	25dhpp + h + nadph --> 25dthpp + nadp
755	fmn + h + nadph --> fmnh2 + nadp
786	glu5p + h + nadph --> glu5sa + nadp + pi
842	akg + h + nadph + nh4 --> glu-L + h2o + nadp
877	gthox + h + nadph --> (2) gthrd + nadp

945	1p3h5c + (2) h + nadph --> 4hpro-LT + nadp
949	aspsa + h + nadph --> hom-L + nadp
1012	Lkynr + h + nadph + o2 --> h2o + hLkynr + nadp
1017	h + mthgxl + nadph --> lald-L + nadp
1123	(2) h + mlthf + nadph --> 5mthf + nadp
1158	(2) Nfortyr + h + nadph --> Nbfortyr + nadp
1220	1pyr5c + (2) h + nadph --> nadp + pro-L
1341	h + nadph + o2 + sphgn --> h2o + nadp + psphings
1410	L2aadp6sa + glu-L + h + nadph <==> h2o + nadp + saccrp-L
1418	glc-D + h + nadph --> nadp + sbt-D
1430	3dhsk + h + nadph --> nadp + skm
1450	(2) frdp + h + nadph --> nadp + (2) ppi + sql
1506	h + nadph + trdox --> nadp + trdrd
1509	h2o + tre6p --> pi + tre
708	(3) h + malcoa + (2) nadph + octa --> co2 + coa + dca + h2o + (2) nadp
710	(3) h + malcoa + (2) nadph + occoa --> co2 + coa + dcacoa + h2o + (2) nadp
711	dca + (3) h + malcoa + (2) nadph --> co2 + coa + ddca + h2o + (2) nadp
713	dcacoa + (3) h + malcoa + (2) nadph --> co2 + coa + ddcacoa + h2o + (2) nadp
714	ddca + (3) h + malcoa + (2) nadph --> co2 + coa + h2o + (2) nadp + ttdca
716	ddcacoa + (3) h + malcoa + (2) nadph --> co2 + coa + h2o + (2) nadp + tdcoa
719	(3) h + malcoa + (2) nadph + ttdca --> co2 + coa + h2o + hdca + (2) nadp
721	(3) h + malcoa + (2) nadph + tdcoa --> co2 + coa + h2o + (2) nadp + pmtcoa
724	(3) h + hdca + malcoa + (2) nadph --> co2 + coa + h2o + (2) nadp + ocdca
726	(3) h + malcoa + (2) nadph + pmtcoa --> co2 + coa + h2o + (2) nadp + stcoa
731	(3) h + malcoa + (2) nadph + ttc --> co2 + coa + h2o + hexc + (2) nadp
447	44mzym + (3) h + (3) nadph + (3) o2 --> 4mzym_int1 + (4) h2o + (3) nadp
448	4mzym + (3) h + (3) nadph + (3) o2 --> (4) h2o + (3) nadp + zym_int1
717	ddca + (4) h + malcoa + (3) nadph + o2 --> co2 + coa + (3) h2o + (3) nadp + ttdcea
722	(4) h + malcoa + (3) nadph + o2 + ttdca --> co2 + coa + (3) h2o + hdcea + (3) nadp
727	(4) h + hdca + malcoa + (3) nadph + o2 --> co2 + coa + (3) h2o + (3) nadp + ocdcea
1032	(2) h + lanost + (3) nadph + (3) o2 --> 44mctr + for + (4) h2o + (3) nadp
730	(9) h + (3) malcoa + (6) nadph + ocdca --> (3) co2 + (3) coa + (3) h2o + (6) nadp + ttc
733	accoa + (9) h + (3) malcoa + (6) nadph --> (3) co2 + (3) coa + (3) h2o + (6) nadp + occoa
734	accoa + (8) h + (3) malcoa + (6) nadph --> (3) co2 + (4) coa + (2) h2o + (6) nadp + octa

#### Producción ATP

1374	adp + h + pep --> atp + pyr
------	-----------------------------

### Consumo ATP

232	L2aadp + atp + h + nadph --> L2aadp6sa + amp + nadp + ppi
233	L2aadp + atp + h + nadh --> L2aadp6sa + amp + nad + ppi
250	accoa + atp + hco3 <==> adp + h + malcoa + pi
281	ac + atp + coa --> accoa + amp + ppi
295	amp + atp <==> (2) adp
302	atp --> camp + ppi
303	adn + atp --> adp + amp + h
307	aps + atp --> adp + h + paps
331	ala-L + atp + trnaala --> alatrna + amp + ppi
387	asp-L + atp + citr-L <==> amp + argsuc + h + ppi
388	arg-L + atp + trnaarg --> amp + argtrna + ppi
397	asp-L + atp + gln-L + h2o --> amp + asn-L + glu-L + h + ppi
398	asn-L + atp + trnaasn --> amp + asntrna + ppi
407	asp-L + atp --> 4pasp + adp
412	asp-L + atp + trnaasp --> amp + asptrna + ppi
420	amp[x] + atp[c] + h[x] --> amp[c] + atp[x] + h[c]
421	adp + atp + h --> ap4a + pi
424	atp + (2) h2o --> amp + (2) h + (2) pi
426	atp + prpp --> ppi + prbatp
427	atp[c] + h2o[c] --> adp[c] + h[e] + pi[c]
432	adp[x] + atp[c] + h[x] --> adp[c] + atp[x] + h[c]
433	atp + btn + h --> btamp + ppi
484	atp + chol --> adp + cholp + h
521	atp + nh4 + utp --> adp + ctp + (2) h + pi
522	atp + gln-L + h2o + utp --> adp + ctp + glu-L + (2) h + pi
530	atp + cys-L + trnacys --> amp + cysrna + ppi
537	atp + cmp <==> adp + cdp
538	atp + dcmp <==> adp + dcdp
545	atp + damp <==> adp + dadp
554	atp + co2 + dann <==> adp + dtbt + (3) h + pi
568	atp + dgmp <==> adp + dgdp
572	atp + dha --> adp + dhap + h
576	atp + dhpt + glu-L --> adp + dhf + h + pi
599	atp + dpcoa --> adp + coa + h
602	5dpmev + atp --> adp + co2 + ipdp + pi
605	atp + drib --> 2dr5p + adp + h
608	atp + dtmp <==> adp + dtdp
611	atp + duri --> adp + dump + h
642	atp + etha --> adp + ethamp + h
681	atp + coa + ttdca <==> amp + ppi + tdcoa

683	atp + coa + ttdcea <==> amp + ppi + tdecoa
685	atp + coa + hdca <==> amp + pmtcoa + ppi
687	atp + coa + hdcea <==> amp + hdcoa + ppi
689	atp + coa + ocdca <==> amp + ppi + stcoa
690	atp + coa + ocdcea <==> amp + odecoa + ppi
691	atp + coa + ocdcya <==> amp + ocycacoa + ppi
752	atp + fnn + h --> fad + ppi
763	atp + f1p --> adp + fdp + h
765	5fthf + atp --> adp + methf + pi
767	5fthf + atp + h2o --> 10fthf + adp + h + pi
768	atp + for + thf --> 10fthf + adp + pi
795	atp + gal --> adp + gal1p + h
822	atp + gmp <==> adp + gdp
831	atp + glu-L + nh4 --> adp + gln-L + h + pi
832	atp + gln-L + trnagln --> amp + glntrna + ppi
838	atp + glu-L --> adp + glu5p
839	atp + cys-L + glu-L --> adp + glucys + h + pi
843	atp + glc-D --> adp + g6p-B + h
847	atp + glu-L + trnaglu --> amp + glutrna + ppi
863	atp + glyc --> adp + glyc3p + h
866	atp + gly + trnagly --> amp + glytrna + ppi
869	atp + gln-L + h2o + xmp --> amp + glu-L + gmp + (2) h + ppi
874	atp + gsn --> adp + gmp + h
884	atp + glucys + gly --> adp + gthrd + h + pi
921	4mhetz + atp --> 4mpetz + adp + h
922	atp + glc-D --> adp + g6p + h
923	atp + man --> adp + h + man6p
924	atp + fru --> adp + f6p + h
931	atp + his-L + trnahis --> amp + histrna + ppi
942	4ahmmp + atp --> 4ampm + adp + h
951	atp + hom-L --> adp + h + phom
974	atp + ile-L + trnaile --> amp + iletrna + ppi
986	atp + ins --> adp + h + imp
1026	atp + leu-L + trnaleu --> amp + leutrna + ppi
1038	atp + lys-L + trnalys --> amp + lystrna + ppi
1068	atp + h2o + met-L --> amet + pi + ppi
1071	atp + met-L + trnamet --> amp + mettrna + ppi
1075	atp + mev-R --> 5pmev + adp + h
1130	atp + nad --> adp + h + nadp
1137	atp + dnad + nh4 --> amp + h + nad + ppi
1148	atp + gdp <==> adp + gtp

1149	atp + udp <==> adp + utp
1150	atp + cdp <==> adp + ctp
1151	atp + dtdp <==> adp + dttp
1152	atp + dgdp <==> adp + dgtp
1153	atp + dudp <==> adp + dutp
1154	atp + dcdp <==> adp + dctp
1155	atp + dadp <==> adp + datp
1156	atp + idp <==> adp + itp
1164	atp + h + nmh --> nad + ppi
1171	atp + h + nicrnt --> dnad + ppi
1223	atp + (0.01) pa_SC --> adp + (0.01) dagpy_SC
1225	ala-B + atp + pant-R --> amp + h + pnto-R + ppi
1231	atp + hco3 + pyr --> adp + h + oaa + pi
1249	atp + f6p --> adp + fdp + h
1250	atp + f6p --> adp + f26bp + h
1251	atp + tag6p-D --> adp + h + tagdp-D
1252	atp + s7p --> adp + h + s17bp
1256	3pg + atp <==> 13dpg + adp
1265	atp + phe-L + trnaphe --> amp + phetrna + ppi
1270	atp + (0.01) ptd3ino_SC --> adp + h + (0.01) ptd134bp_SC
1271	atp + (0.01) ptd3ino_SC --> adp + h + (0.01) ptd135bp_SC
1274	atp + (0.01) ptd4ino_SC --> adp + h + (0.01) ptd145bp_SC
1275	atp + (0.01) ptd1ino_SC --> adp + h + (0.01) ptd3ino_SC
1276	atp + (0.01) ptd1ino_SC --> adp + h + (0.01) ptd4ino_SC
1292	5pmev + atp --> 5dpmev + adp
1294	atp + minohp --> adp + ppmi12346p
1296	atp + mi13456p --> adp + ppmi1346p
1297	4ampm + atp --> 2mahmp + adp
1299	atp + pnto-R --> 4ppan + adp + h
1308	atp + oaa --> adp + co2 + pep
1318	atp + gly + pram <==> adp + gar + h + pi
1319	atp + fpram --> adp + air + (2) h + pi
1322	5aizc + asp-L + atp <==> 25aics + adp + h + pi
1324	atp + fgam + gln-L + h2o --> adp + fpram + glu-L + h + pi
1327	atp + pro-L + trnapro --> amp + ppi + protrna
1330	atp + r5p <==> amp + h + prpp
1355	atp + h + pan4p --> dpcoa + ppi
1369	atp + pydam --> adp + h + pyam5p
1370	atp + pydx --> adp + h + pydx5p
1371	atp + pydxn --> adp + h + pdx5p
1387	atp + ribflv --> adp + fmh + h

1391	atp + rib-D --> adp + h + r5p
1403	atp + rnam --> adp + h + nmn
1404	atp + trdrd --> datp + h2o + trdox
1412	atp + h + so4 --> aps + ppi
1424	atp + ser-L + trnaser --> amp + ppi + sertrna
1431	atp + skm --> adp + h + skm5p
1434	atp + sphgn --> adp + h + sph1p
1435	atp + psphings --> adp + h + psph1p
1471	atp + glu-L + thf <==> adp + h + pi + thfglu
1487	atp + thr-L + trnathr --> amp + ppi + thrtrna
1496	atp + thymd --> adp + dtmp + h
1497	atp + thm --> amp + h + thmpp
1499	atp + thmpp --> adp + thmtp
1502	atp + thmmp <==> adp + thmpp
1519	atp + trnatrp + trp-L --> amp + ppi + trptrna
1529	atp + trnatyr + tyr-L --> amp + ppi + tyrtrna
1541	atp + ump <==> adp + udp
1551	atp + hco3 + urea <==> adp + allphn + h + pi
1552	atp + dump <==> adp + dudp
1554	atp + uri --> adp + h + ump
1561	atp + trnaval + val-L --> amp + ppi + valtrna
1568	atp + xyly-D --> adp + h + xu5p-D
1	[c] : (1.1348) 13BDgln + (0.4588) ala-L + (0.046) amp + (0.1607) arg-L + (0.1017) asn-L + (0.2975) asp-L + (59.276) atp + (0.000001) camp + (0.000001) chitin + (0.0447) cmp + (0.000001) coa + (0.0066) cys-L + (0.0036) damp + (0.0024) dcmp + (0.0024) dgmp + (0.0036) dtmp + (0.0007) ergst + (0.000001) fad + (0.1054) gln-L + (0.3018) glu-L + (0.2904) gly + (0.5185) glycogen + (0.046) gmp + (0.000001) gthrd + (59.276) h2o + (0.0663) his-L + (0.1927) ile-L + (0.2964) leu-L + (0.2862) lys-L + (0.8079) mannan + (0.0507) met-L + (0.000001) nad + (0.000006) pa_SC + (0.00006) pc_SC + (0.000045) pe_SC + (0.1339) phe-L + (0.000001) pheme + (0.1647) pro-L + (0.000017) ps_SC + (0.000053) ptd1ino_SC + (0.000001) q6 + (0.00099) ribflv + (0.1854) ser-L + (0.02) so4 + (0.000001) thf + (0.000001) thmtp + (0.1914) thr-L + (0.0234) tre + (0.000066) triglyc_SC + (0.0284) trp-L + (0.102) tyr-L + (0.0599) ump + (0.2646) val-L + (0.0015) zymst --> (59.276) adp + (58.70001) h + (59.305) pi
454	(2) atp + gln-L + h2o + hco3 --> (2) adp + cbp + glu-L + (2) h + pi

#### Producción acetato

18	ace ->
----	--------

#### Producción dióxido de carbono

37	co2 ->
----	--------

Producción etanol

56	eth ->
----	--------

Producción glicerol

75	gly ->
----	--------

Producción succinato

143	scc ->
-----	--------

**Participación de los metabolitos en las funciones objetivo en mitocondria**

Producción ATP

429	adp[m] + (3) h[c] + pi[m] --> atp[m] + (2) h[m] + h2o[m]
-----	--

Producción NADH+NADPH

<nadh>

219	coa + coucoa + h2o + nad --> 4hbzcoa + accoa + h + nadh
222	4hglusa + h2o + nad <==> e4hglu + (2) h + nadh
237	4abutn + h2o + nad --> 4abut + (2) h + nadh
354	h2o + id3acald + nad --> (2) h + ind3ac + nadh
358	acald + h2o + nad --> ac + (2) h + nadh
810	dhlam + nad <==> h + lpam + nadh
813	dhlpro + nad <==> h + lpro + nadh
859	gly + nad + thf --> co2 + mlthf + nadh + nh4
928	hicit + nad <==> h + nadh + oxag
946	4hpro-LT + nad --> 1p3h5c + (2) h + nadh
962	icit + nad --> akg + co2 + nadh
1061	mal-L + nad <==> h + nadh + oaa
1064	mal-L + nad --> co2 + nadh + pyr
1197	2obut + coa + nad --> co2 + nadh + ppcoa
1219	1pyr5c + (2) h2o + nad --> glu-L + h + nadh
1238	dhlam + nad --> h + lpam + nadh
1239	coa + nad + pyr --> accoa + co2 + nadh
1261	1p3h5c + (2) h2o + nad --> e4hglu + h + nadh
1326	nad + pro-L --> 1pyr5c + (2) h + nadh

<nadph>

356	h2o + id3acald + nadp --> (2) h + ind3ac + nadph
360	acald + h2o + nadp --> ac + (2) h + nadph

947	4hpro-LT + nadp --> 1p3h5c + (2) h + nadph
964	icit + nadp --> akg + co2 + nadph
1065	mal-L + nadp --> co2 + nadph + pyr
1122	mlthf + nadp <==> methf + nadph
1376	glu5sa + h2o + nadp --> glu-L + (2) h + nadph

### Consumo NADH+NADPH

#### <nadh>

191	34hpp + h + nadh --> 34hpl + nad
336	2mbald + h + nadh --> 2mbtoh + nad
339	2mppal + h + nadh --> ibutoh + nad
342	3mbald + h + nadh --> iamoh + nad
345	h + nadh + pacald --> 2phetoh + nad
348	h + id3acald + nadh --> ind3eth + nad
351	acald + h + nadh --> etoh + nad
780	dhap + h + nadh --> glyc3p + nad
919	hemeO + nadh + o2 --> h2o + hemeA + nad
1129	h + nadh + q6 --> nad + q6h2

#### <nadph>

312	acg5p + h + nadph --> acg5sa + nadp + pi
575	dhf + h + nadph --> nadp + thf
604	2dhp + h + nadph --> nadp + pant-R
879	gthox + h + nadph --> (2) gthrd + nadp
1009	alac-S + h + nadph --> 23dhmb + nadp
1010	2ahbut + h + nadph --> 23dhmp + nadp
1508	h + nadph + trdox --> nadp + trdrd
709	(3) h + malACP + (2) nadph + ocACP --> ACP + co2 + dcaACP + h2o + (2) nadp
712	dcaACP + (3) h + malACP + (2) nadph --> ACP + co2 + ddcaACP + h2o + (2) nadp
715	ddcaACP + (3) h + malACP + (2) nadph --> ACP + co2 + h2o + myrsACP + (2) nadp
720	(3) h + malACP + myrsACP + (2) nadph --> ACP + co2 + h2o + (2) nadp + palmACP
725	(3) h + malACP + (2) nadph + palmACP --> ACP + co2 + h2o + (2) nadp + ocdcaACP
718	ddcaACP + (4) h + malACP + (3) nadph + o2 --> ACP + co2 + (3) h2o + (3) nadp + tdeACP
723	(4) h + malACP + myrsACP + (3) nadph + o2 --> ACP + co2 + (3) h2o + hdeACP + (3) nadp
728	(4) h + malACP + (3) nadph + o2 + palmACP --> ACP + co2 + (3) h2o + (3) nadp + octeACP
729	(5) h + malACP + (4) nadph + (2) o2 + palmACP --> ACP + co2 + (5) h2o + (4) nadp + ocdcyaACP
732	acACP + (9) h + (3) malACP + (6) nadph --> (3) ACP + (3) co2 + (3) h2o + (6) nadp + ocACP

## Participación de los metabolitos en las funciones objetivo en peroxisoma

### Síntesis ácidos grasos

679	atp + coa + dca <==> amp + dcacoa + ppi
680	atp + coa + ddca <==> amp + ddcacoa + ppi
682	atp + coa + ttdca <==> amp + ppi + tdcoa
684	atp + coa + ttdcea <==> amp + ppi + tdecoa
686	atp + coa + hdca <==> amp + pmtcoa + ppi
688	atp + coa + hdcea <==> amp + hdcoa + ppi
692	atp + coa + ttc <==> amp + ppi + ttcco
693	atp + coa + hexc <==> amp + hexcco + ppi
694	atp + coa + octa <==> amp + octo + ppi

### Consumo ATP

283	ac + atp + coa --> acco + amp + ppi
653	atp[x] + h2o[x] + tdco[c] --> adp[x] + h[x] + pi[x] + tdco[x]
657	atp[x] + h2o[x] + tdeco[c] --> adp[x] + h[x] + pi[x] + tdeco[x]
661	atp[x] + h2o[x] + pmtco[c] --> adp[x] + h[x] + pi[x] + pmtco[x]
665	atp[x] + h2o[x] + hdco[c] --> adp[x] + h[x] + hdco[x] + pi[x]
669	atp[x] + h2o[x] + stco[c] --> adp[x] + h[x] + pi[x] + stco[x]
672	atp[x] + h2o[x] + odeco[c] --> adp[x] + h[x] + odeco[x] + pi[x]
675	atp[x] + h2o[x] + ocycaco[c] --> adp[x] + h[x] + ocycaco[x] + pi[x]
679	atp + coa + dca <==> amp + dcacoa + ppi
680	atp + coa + ddca <==> amp + ddcacoa + ppi
682	atp + coa + ttdca <==> amp + ppi + tdcoa
684	atp + coa + ttdcea <==> amp + ppi + tdecoa
686	atp + coa + hdca <==> amp + pmtcoa + ppi
688	atp + coa + hdcea <==> amp + hdcoa + ppi
692	atp + coa + ttc <==> amp + ppi + ttcco
693	atp + coa + hexc <==> amp + hexcco + ppi
694	atp + coa + octa <==> amp + octo + ppi

### Producción NADH+NADPH

#### <nadh>

901	3hxcco + nad <==> 3ohxcco + h + nadh
1062	mal-L + nad <==> h + nadh + oaa
706	(3) coa + (3) h2o + (3) nad + (3) o2 + ttcco --> (3) acco + (3) h + (3) h2o2 + (3) nadh + stco
707	(3) coa + (3) h2o + (3) nad + (3) o2 + octo --> (4) acco + (3) h + (3) h2o2 + (3) nadh
697	(6) coa + (6) h2o + (6) nad + nadph + (6) o2 + tdeco --> (7) acco + (5) h + (6) h2o2 + (6) nadh + nadp

698	(6) coa + (6) h2o + (6) nad + (5) o2 + tdecoa --> (7) accoa + (6) h + (5) h2o2 + (6) nadh
699	(7) coa + (7) h2o + hdcoa + (7) nad + nadph + (7) o2 --> (8) accoa + (6) h + (7) h2o2 + (7) nadh + nadp
700	(7) coa + (7) h2o + hdcoa + (7) nad + (6) o2 --> (8) accoa + (7) h + (6) h2o2 + (7) nadh
701	(8) coa + (8) h2o + (8) nad + nadph + (8) o2 + odecoa --> (9) accoa + (7) h + (8) h2o2 + (8) nadh + nadp
702	(8) coa + (8) h2o + (8) nad + (7) o2 + odecoa --> (9) accoa + (8) h + (7) h2o2 + (8) nadh
703	(8) coa + (8) h2o + (8) nad + (2) nadph + (8) o2 + ocdycacoa --> (9) accoa + (6) h + (8) h2o2 + (8) nadh + (2) nadp
704	(8) coa + (8) h2o + (8) nad + nadph + (7) o2 + ocdycacoa --> (9) accoa + (7) h + (7) h2o2 + (8) nadh + nadp
705	(8) coa + (8) h2o + (8) nad + (6) o2 + ocdycacoa --> (9) accoa + (8) h + (6) h2o2 + (8) nadh
<nadph>	
965	icit + nadp --> akg + co2 + nadph

### Consumo NADH+NADPH

<nadh>

902	3odcoa + h + nadh <==> 3hdcoa + nad
903	3oddcoa + h + nadh <==> 3hddcoa + nad
904	3otdcoa + h + nadh <==> 3htdcoa + nad
905	3ohdcoa + h + nadh <==> 3hhdcoa + nad
906	3ohodcoa + h + nadh <==> 3hodcoa + nad

<nadph>

697	(6) coa + (6) h2o + (6) nad + nadph + (6) o2 + tdecoa --> (7) accoa + (5) h + (6) h2o2 + (6) nadh + nadp
699	(7) coa + (7) h2o + hdcoa + (7) nad + nadph + (7) o2 --> (8) accoa + (6) h + (7) h2o2 + (7) nadh + nadp
701	(8) coa + (8) h2o + (8) nad + nadph + (8) o2 + odecoa --> (9) accoa + (7) h + (8) h2o2 + (8) nadh + nadp
704	(8) coa + (8) h2o + (8) nad + nadph + (7) o2 + ocdycacoa --> (9) accoa + (7) h + (7) h2o2 + (8) nadh + nadp
703	(8) coa + (8) h2o + (8) nad + (2) nadph + (8) o2 + ocdycacoa --> (9) accoa + (6) h + (8) h2o2 + (8) nadh + (2) nadp