

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS QUÍMICAS
Departamento de Bioquímica y Biología Molecular I



TESIS DOCTORAL

**Reconstrucción y análisis de modelos
metabólicos a escala genómica
Aplicación al estudio de bacterias endosimbiontes**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Miguel Ponce de León Capurro

Directores

**Francisco Montero Carnerero
Juli Peretó Magraner**

Madrid, 2018

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS QUÍMICAS
Departamento de Bioquímica y Biología Molecular I



**Reconstrucción y análisis de modelos
metabólicos a escala genómica**
Aplicación al estudio de bacterias endosimbiontes

Miguel Ponce de León Capurro

TESIS DOCTORAL

Dirigida por

Francisco Montero Carnerero y Juli Peretó Magraner

Madrid, 2016

"... la naturaleza misma de la investigación (...) nos obliga a atravesar en zigzag un amplio dominio de pensamiento en todas las direcciones. Las anotaciones (...) son como un conjunto de bosquejos de paisajes que han resultado de estos largos y enmarañados viajes. Así pues, este libro es en realidad sólo un álbum..."

Ludwig Wittgenstein, *Investigaciones Filosóficas*

Agradecimientos

Y finalmente ha llegado la hora de dar las gracias. Este parece ser el lugar de la tesis donde uno se relaja y puede aflojar, aunque sea un poco, el tono monocorde del discurso científico. Así que acá vamos! A Paco Montero, por ser como es, por haberme aceptado en su grupo, haber dirigido esta tesis, por haber sido siempre un apoyo incondicional, y obviamente, por sus clases magistrales (teóricas y prácticas) sobre el noble arte del flamenco. A Juli Peretó, por codirigir esta tesis, por lo que me ha enseñado de metabolismo y por todo su apoyo. A Fede Morán, por echar una mano siempre que hizo falta y por brindarme la posibilidad de trabajar en el Máster en Bioinformática. A Laura, mi noble compañera del día a día de trabajo. A los miembros pasados y presentes del grupo de Biofísicas de la UCM. A los profesores del máster en Bioinformática, y al personal de la ENS. Al grupo de Biocomputing del CNIO-INB, en especial a Alfonso y a Belén. A todos los estudiantes del máster, que sin saberlo, hicieron posible esta tesis. A mis colaboradores: Dani, Matteo, don Enrique y el Yuyo. A la gente del Departamento de Bioquímica y Biología Molecular I de la UCM. Al grupo de Genética Evolutiva del Cavanilles, en especial a Amparo. A Madrid por ser una ciudad tan acogedora. A España por ser una tierra tan agradecida. A los amigos y amigas que este país me ha regalado. A la gente de la secciones de Biofísica y Biomatemáticas de la Facultad de Ciencias. A toda mi familia y amigos que los quiero y los extraño. Y en especial a Ceci, por haberse embarcado conmigo en esta aventura de venirse a la península Ibérica, en aquel entonces *Terra ignota*, y que hoy es con creces nuestro segundo hogar, y por sobre todo, por aguantar el tirón siempre que el viento sopló tormenta, ¡te quiero guapa!

Índice General

Agradecimientos.....	v
Índice General.....	vii
Abstract.....	ix
Resumen.....	xiii
1 Introducción.....	17
2 Objetivos.....	23
3 Detección de inconsistencias y refinado de modelos metabólicos.....	25
3.1 Introducción.....	25
3.2 Detección de reacciones bloqueadas en redes metabólicas.....	26
3.3 Inconsistencias termodinámicas en redes estequiométricas.....	33
3.4 Definición y detección de metabolitos <i>gap</i>	42
3.5 <i>Pools</i> conservados y el problema de los metabolitos <i>pseudo-gap</i>	44
3.6 El grafo de inconsistencias y los módulos inconexos.....	47
3.7 Metacoupling y su aplicación al <i>gap-filling</i>	51
4 Revisión y análisis de un modelo metabólico de <i>Blattabacterium cuenoti</i>	55
4.1 Introducción.....	55
4.2 Materiales y Métodos.....	57
4.2.1 Modelo metabólico y anotación funcional.....	57
4.3 Resultados.....	57
4.3.1 Análisis de consistencia del modelo <i>Blattabacterium</i> Bge iCG238.....	57
4.3.2 Actualizando el modelo metabólico de <i>Blatabacterium</i> Bge iCG238.....	71
4.3.3 Análisis de la reorganización de los subconjuntos de reacciones.....	72
4.3.4 Análisis comparativo en la predicción de medios mínimos.....	73
4.3.5 Fragilidad, robustez y reducción genómica.....	75
4.4 Discusión.....	78
5 Complementación metabólica en consorcios endosimbiontes.....	81
5.1 Introducción.....	81
5.2 Materiales y métodos.....	85
5.2.1 Genomas anotados.....	85
5.2.2 Reconstrucción y refinado de modelos metabólicos.....	86
5.2.3 Construcción de una ecuación de biomasa de <i>C. cedri</i>	87
5.2.4 Definición de una restricción de contorno.....	87
5.2.5 Estimación de permeabilidades.....	88
5.2.6 Alineamiento múltiple de secuencias.....	88
5.3 Resultados.....	89
5.3.1 Reconstrucción metabólica de los endosimbiontes del áfido del cedro.....	89
5.3.2 Definición de un modelo metabólico del consorcio endosimbionte.....	101
5.3.3 Definición de una condición de contorno.....	107
5.3.4 Análisis de patrones de crecimiento óptimo.....	108
5.3.5 Experimentos <i>in-silico</i> de reducción: el caso del triptófano.....	110
5.3.6 Complementación metabólica: condiciones necesarias y optimalidad.....	117
5.3.7 El papel de la permeabilidad.....	130
5.4 Discusión.....	133
6 Construcción y análisis de Metamodelos.....	137

6.1	Introducción.....	137
6.2	Materiales y Métodos.....	139
6.2.1	Modelos metabólicos reconstruidos a escala genómica.....	139
6.3	Resultados.....	141
6.3.1	Análisis de consistencia de modelos metabólicos de bacterias.....	141
6.3.2	Construcción y curado del metamodelo.....	143
6.3.3	Análisis de acoplamiento de flujos del metamodelo.....	149
6.3.4	Aplicación del <i>gap-filling</i> al conjunto de 130 GEMMs.....	158
6.3.5	Extendiendo un metamodelo con nueva información bioquímica.....	162
6.3.6	Análisis del impacto de la extensión de un metamodelo.....	163
6.4	Discusión.....	165
6.4.1	Detección de inconsistencias estructurales en modelos metabólicos.....	165
6.4.2	Análisis de consistencia global empleando un metamodelo.....	166
6.4.3	Acoplamiento de flujos y módulos en el metamodelo.....	166
6.4.4	Impacto de la consistencia de un metamodelo en los resultados del <i>gap-filling</i> . .	168
6.4.5	Extensión de metamodelo mediante la adición de nuevos modelos.....	169
6.4.6	Consideraciones finales.....	169
7	Discusión general.....	171
8	Conclusiones.....	175
	Apéndice A.....	177
	A.1 Introducción al CBM.....	177
	A.1.1 Definiciones generales.....	177
	A.1.2 Imposición de restricción.....	178
	A.1.3 Flujos de intercambio con el ambiente.....	180
	A.1.4 La ecuación de Biomasa.....	182
	A.2 Métodos generales.....	183
	A.2.1 Análisis de Balance de Flujos.....	183
	A.2.2 Análisis de Variabilidad de Flujos.....	184
	A.2.3 Análisis de Acoplamiento de Flujos.....	184
	A.2.4 Predicción de medios mínimos de crecimiento.....	186
	A.2.5 Experimentos de <i>knockout in-silico</i>	187
	Apéndice B.....	189
	M.S. - Capítulo 3.....	189
	M.S. - Capítulo 4.....	190
	M.S. - Capítulo 5.....	190
	M.S. - Capítulo 6.....	191
	Apéndice C.....	193
	Listado de publicaciones.....	195
	Bibliografía.....	197

Abstract

Introduction

Metabolic reconstruction is the computational process that aims to elucidate the metabolic network of a given organism from its annotated genome and other sources of information. This integrative process implies the addition of a new level to the genomic annotation. In this process, various enzymatic activities encoded in a genome are put in the context of functional modules, *i.e.* the metabolic pathways. The final goal of a metabolic reconstruction is to develop a genome-scale metabolic model (GEMM), that is to say, an in-silico representation of the metabolic network. A GEMM can be used to generate hypotheses about the metabolic capabilities of a network (which can be experimentally tested) through the theoretical framework known as constraint-based modeling (CBM). In this context, the main purpose of this Thesis is to address the methodological problems of reconstruction and refinement of GEMMs, with emphasis on their application to the study of bacterial endosymbionts. On one hand this work attempts to develop new approaches to detect and resolve structural inconsistencies using endosymbiotic bacteria as the case of study. These approaches will be also used to analyze the consistency of metabolic databases. Finally, the problem of the metabolic complementation exhibited by endosymbiotic bacteria, will also be addressed.

Objectives

This thesis aims to address the following specific objectives:

1. Development of methodological approaches for the detection and resolution of structural inconsistencies in genome-scale metabolic models, through the use of constraint-based modeling.
2. Detection and resolution of inconsistencies in metabolic data bases, through reconstruction and analysis of a metamodel of bacterial metabolism.
3. Application of the proposed approaches to the curation and analysis of a genome-scale metabolic model of *Blattabacterium* Bge, cockroach endosymbiont.
4. Reconstruction of a genome-scale compartmentalized model of the endosymbiont consortium of cedar aphid, formed by bacteria *B. aphidicola* BCc and *S. symbiotica* SCc, to study the phenomenon of metabolic complementation.
5. Perform a kinetic study in a simple model, to identify conditions that may favor the emergence of metabolic complementation, with emphasis on: i) the role of allosteric regulation; ii) the effect of permeability of intermediaries of the route; and iii) the cost associated with the synthesis of enzymes.

Results

Using the concept of blocked reaction, a general definition of gap metabolites was introduced, allowing their detection in a straight forward manner. As a consequence, gap metabolites can be identified using their relation to the set of blocked reactions. There also may exist metabolites that although not being detected as "gaps", using the introduced definition, can cause the blockage of certain reactions, and thus we name them pseudo-gaps. A procedure for the detection of the complete set of gap and pseudo-gap metabolites was introduced. Furthermore, the concept of gap graph was introduced. The connected components of this graph can be interpreted as "modules" of the metabolism that becomes unconnected as a consequence of model inconsistencies and thus are referred as the *Unconnected Modules* (UM). The method has been successfully applied to the curation of *iCG238*, the GEMM of *Blattabacterium cuenoti*, endosymbiont of cockroaches, resulting in new version of the model, named *iMP240*.

A large-scale analysis was performed using an already published dataset. The results showed that a large number of reactions (~22%) are blocked in all the models where they are present. To unravel the nature of such inconsistencies, a metamodel was construed by joining the 130 models in a single network. The analysis performed on the metamodel showed that the recurrent inconsistencies found in the models were already present in the metabolic database used during the reconstructions. This also showed that the presence of inconsistencies in a metabolic database can be propagated to the reconstructed models.

The developed methods were used to assess the reconstruction of *iBSCc*, a compartmentalized model of the whole endosymbiotic consortium of the cedar aphid. Simulation using *iBSCc* reproduced the known events of metabolic complementation exhibited by the consortium, and predicted an event not previously described. Moreover, the in-silico analysis of the distributed tryptophan biosynthesis, showed that the design of this complementation exhibits a quasi-optimal yield of the amino acid, using a minimum number of genes. Finally, a simple model to study the kinetic properties of metabolic complementation was introduced. The simulations showed that, under the presence of enzymatic inhibition metabolic complementation can emerge as an optimal strategy to minimize the total enzyme required to satisfy the demand of the synthesized product.

Conclusions

The main conclusions of this PhD dissertation are the following:

1. The definition of gap metabolite and its extension to the cases of pseudo-gaps combined with the definition of blocked reactions, has allowed the introduction of the concept of unconnected module, as well as a simple method to their identification. The application of this novel approach to the curation of the GEMM of *Blattabacterium Bge iCG238*, resulted in a refined version of the model, named *iMP240*.
2. The performed large-scale analysis allowed the detection of recurrent inconsistencies which indicate the presence of gaps at the level of metabolic database. The reconstruction and curation of a meta-model named MM130, showed that the CBM can be used to review and curate of the metabolic databases.

3. The metabolic reconstruction of *B. aphidicola* BCc and *S. symbiotica* SCc, endosymbionts of the cedar aphid, resulted in two GEMM named *iBCc98* and *iSCc236*, respectively. Furthermore, the analysis of these models has resulted in a revision and correction of the corresponding genomic annotations. These results highlight the importance of using models in the study of non-cultivable microorganisms.
4. Combining *iBCc98* and *iSCc236*, a compartmentalized model of the consortium was reconstructed. The model, named *iBSCc*, reproduced the known events of metabolic complementation exhibited by the consortium, and predicted a new event not previously described. Moreover, the in-silico analysis of the distributed tryptophan biosynthesis, showed that the design of this complementation could have been the outcome of a selective process.
5. The kinetic analysis of metabolic complementation showed, that under the presence of enzymatic by-product inhibition, metabolic complementation emerges as the optimal strategy that minimizes the concentration of enzymes. Moreover the physicochemical properties of the intermediates of a pathway can determine the compound exchanged in a complemented pathway.

Resumen

Introducción

La reconstrucción metabólica es el proceso computacional que tiene como objetivo elucidar la red metabólica de un organismo a partir de su genoma anotado y otras fuentes de información. El objetivo final de una reconstrucción es desarrollar un modelo metabólico a escala genómica (GEMM), *i.e.* una representación in-silico de la red metabólica. Un GEMM se puede utilizar para generar y simular el funcionamiento del sistema a través del marco teórico denominado modelado basado en restricciones (CBM). En este contexto, el objetivo principal de esta Tesis es abordar los problemas metodológicos de la reconstrucción y refinado de GEMM, con énfasis en su aplicación al estudio de bacterias endosimbiontes. En el trabajo se pretenden desarrollar nuevas aproximaciones para detectar y resolver inconsistencias estructurales, empleando como casos de estudio metabolismos de bacterias endosimbiontes. Estas aproximaciones también serán empleadas para analizar la consistencia de bases de datos metabólicas. Finalmente, se abordará el problema de la complementación metabólica en bacterias endosimbiontes.

Objetivos

Esta tesis pretende abordar los siguientes objetivos específicos:

1. Desarrollo de aproximaciones metodológicas para la detección y resolución de inconsistencias estructurales en modelos metabólicos reconstruidos a escala genómica, a través del modelado basado en restricciones.
2. Detección y resolución de inconsistencias en bases de datos metabólicas, mediante la reconstrucción y análisis de un metamodelo del metabolismo bacteriano.
3. Aplicación de las aproximaciones propuestas a la revisión y análisis un modelo metabólico de *Blattabacterium* Bge, endosimbionte de la cucaracha.
4. Reconstrucción de un modelo del consorcio endosimbionte del áfido del cedro formado por las bacterias *B. aphidicola* BCc y *S. symbiotica* SCc, para estudiar el fenómeno de la complementación metabólica.
5. Realizar un estudio cinético sobre las condiciones que favorecen la emergencia de complementación metabólica, haciendo énfasis en: i) el papel de la inhibición enzimática; ii) el efecto de permeabilidad de la membrana a los intermediarios de la ruta; y iii) el coste asociado a la síntesis de las enzimas.

Resultados

Utilizando el concepto de reacción bloqueada, se introdujo una definición general de metabolito *gap*, que permite su detección de una manera directa. También pueden existir metabolitos que aunque no se detecten como *gaps*, pueden causar el bloqueo de ciertas

reacciones y por eso se han denominado *pseudo-gaps*. Se introdujo un procedimiento para la detección del conjunto completo de metabolitos *gap* y *pseudo-gap*. Además, se introdujo el concepto de grafo de inconsistencias y de módulos inconexos (UM). Los UM pueden ser interpretados como "módulos" del metabolismo que se encuentran desconectados como consecuencia de inconsistencias. Las aproximaciones se han aplicado con éxito en la revisión y refinado del GEMM de *Blattabacterium cuenoti* iCG238, endosimbionte de las cucarachas. Esto ha resultado en una nueva versión del modelo, denominado iMP240.

Empleando los desarrollos metodológicos se realizó un análisis a gran escala empleando un conjunto de datos ya publicado de 130 GEMM. Los resultados mostraron que un gran número de reacciones (~22%) están bloqueadas en todos los modelos donde están presentes. Para desentrañar la naturaleza de tales inconsistencias, se construyó un metamodelo uniendo los 130 modelos en una sola red. El análisis realizado sobre el metamodelo mostró que las inconsistencias recurrentes encontradas en los modelos, estaban presentes en la base de datos metabólica utilizada durante las reconstrucciones. Esto demostró como el CBM puede ser empleado para revisar y refinar bases de datos metabólicas.

Los métodos desarrollados fueron utilizados en la reconstrucción de *iBSCc*, un GEMM del consorcio endosimbionte del áfido de cedro. La simulación utilizando *iBSCc* reproducen los eventos conocidos de complementación metabólica que dicho consorcio exhibe, y predice un nuevo evento, no descrito previamente. Por otra parte, los análisis *in-silico* de la complementación en la ruta del triptófano, mostró que este diseño exhibe un rendimiento cuasi-óptimo de producción del aminoácido, con un conjunto mínimo de genes. Finalmente, se realizó un estudio de las propiedades cinéticas de la complementación metabólica. Las simulaciones mostraron que bajo la presencia de inhibición enzimática la complementación metabólica puede emerger como una estrategia óptima para minimizar la enzima total requerida para satisfacer la demanda del producto sintetizado.

Conclusiones

Las principales conclusiones del estudio realizado y que configuran la presente Tesis doctoral son las siguientes:

1. La definición de metabolito *gap*, así como su extensión al caso de los *pseudo-gaps*, combinada con la definición previa de reacción bloqueada, ha permitido introducir el concepto de módulo inconexo así como un método para su cálculo. La aplicación de estas aproximaciones a la revisión del modelo de *Blattabacterium* Bge iCG238, ha dado como lugar a una versión refinada del modelo denominada iMP240.
2. El análisis a gran escala realizado, ha permitido detectar la presencia de inconsistencias recurrentes presentes en bases de datos metabólicas. La reconstrucción y revisado de un metamodelo ha permitido mostrar como el CBM puede ser empleado en la revisión y refinado de bases de datos metabólicas.
3. La reconstrucción metabólica de *B. aphidicola* BCc y *S. symbiotica* SCc, endosimbiontes del áfido del cedro, dio como resultado los respectivos modelos metabólicos denominados, *iBCc98* y *iSCc236*. El análisis de dichos modelos ha

permitido la revisión y corrección de las anotaciones genómicas correspondientes. Esto muestra la importancia del uso de modelos en el estudio de capacidades metabólicas de microorganismo no cultivables.

4. Combinando *iBCc98* e *iSCc236* se reconstruyó un modelo compartimentado del consorcio. El modelo denominado *iBSCc* reproduce los eventos de complementación descritos entre las bacterias, así como un nuevo caso no descrito previamente. Además, el análisis *in-silico* realizado sobre la biosíntesis distribuida del triptófano, muestra que la fijación de dicho evento pudo haber sido consecuencia de un proceso selectivo.
5. El desarrollo y de un modelo simple para el análisis cinético de la complementación, muestra que en presencia de inhibición enzimática, la complementación metabólica es la estrategia óptima que minimiza la concentración enzima necesaria para satisfacer la demanda del producto.

Capítulo 1

1 Introducción

El metabolismo, definido como conjunto de procesos de transporte, síntesis y degradación de moléculas que ocurre dentro una célula, es posiblemente la red biológica más estudiada, debido a su papel fundamental en el funcionamiento de la vida (Morowitz and Smith, 2007). En su estructura interna el metabolismo está compuesto por un conjunto de metabolitos interconectados a través de las reacciones bioquímicas (en su mayoría catalizadas por enzimas), lo que da lugar a una compleja red de transformaciones. Esta compleja red de flujos define las capacidades de una célula para crecer en distintos ambientes y, en el caso de microorganismos como las bacterias, el estado metabólico define en buena medida el fenotipo de la célula (White et al., 2012). Las redes metabólicas de los distintos grupos de organismos comparten muchas características comunes, *e.g.* el uso del ATP como “moneda” de intercambio energético, debido al origen común de la vida (Morowitz et al., 2000). Sin embargo, cada forma de vida posee un cierto repertorio de actividades enzimáticas particulares codificadas en su genoma que le permiten sobrevivir en una cierta gama de condiciones ambientales. Por ejemplo, algunos microorganismos de vida libre presentan metabolismos “versátiles” que les permiten crecer en un amplio rango condiciones ambientales, mientras que otros exhiben capacidades más limitadas y requieren condiciones muy específicas, como es el caso de los microorganismos endosimbiontes adaptados a la vida intracelular (Parter et al., 2007). De esta manera, el estudio de la red metabólica de un organismo puede brindar información sobre su estilo de vida, y puede también ser de gran utilidad para comprender aspectos sobre la organización y funcionamiento del sistema, que facilite el desarrollo de aplicaciones en los campos de la biotecnología y la biomedicina (Oberhardt et al., 2009).

Tradicionalmente, las redes metabólicas se suelen describir en términos de subsistemas o módulos funcionales, comúnmente referidos como rutas o vías metabólicas. En general, una ruta metabólica se refiere a un conjunto coherente de actividades enzimáticas que permiten la transformación de un cierto sustrato en un producto, y que puede presentar un patrón común en la regulación de los genes que codifican las actividades enzimáticas involucradas (Overbeek et al., 2005). Sin embargo, la visión del metabolismo como una colección de rutas metabólicas es un tanto simplista debido a que las rutas metabólicas se encuentran interconectadas unas con otras, dando lugar a una red compleja. Debido a la complejidad exhibida por los sistemas metabólicos, el uso de modelos es esencial para comprender la relación entre la estructura de la red y las

capacidades bioquímicas de un organismo. En biología, se han empleado modelos biofísicos durante décadas, pero su alcance y escala (en término del número de variables que se incluyen) han estado limitados por la escasa disponibilidad de datos experimentales (Heinrich and Schuster, 1996). Sin embargo, en las últimas décadas, la biología y en particular la biología molecular han sufrido una revolución que tuvo su inicio con el comienzo de los proyectos de secuenciación de genomas, y que alcanzó su cénit con el desarrollo de las tecnologías *high-throughput*, que dieron lugar a la era de las ómicas (Hartwell et al., 1999). Estas tecnologías permiten caracterizar y cuantificar, de forma masiva, los *pools* de las diversas moléculas biológicas así como las interacciones entre las mismas (Evans, 2000). La organización y análisis de los datos moleculares producidos por las distintas ómicas ha permitido crear catálogos detallados sobre los componentes moleculares de los seres vivos (Karp, 1998). La disponibilidad de esta información trajo consigo la posibilidad de reconstruir las distintas redes moleculares que existen dentro de la células (Palsson, 1997). En particular, la disponibilidad de genomas anotados para un gran número de organismos ha permitido reconstruir sus correspondientes redes metabólicas, incluyendo la modelización de metabolismos de microorganismos no cultivables a partir de la información genómica o metagenómica (Francke et al., 2005; Notebaart et al., 2006). Esto ha motivado el desarrollo de formalismos y herramientas para la reconstrucción y el análisis de las distintas redes biológicas (Barabási and Oltvai, 2004; Feist et al., 2009).

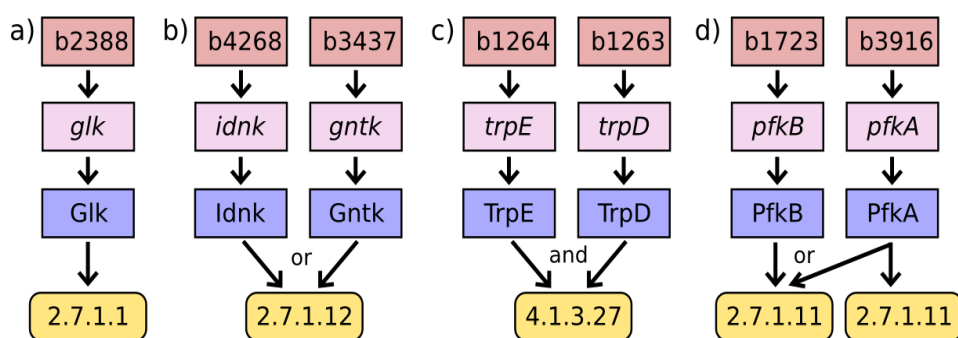


Figura 1.1. Distintos tipos de asociación GPR. Ejemplos de relaciones entre genes, proteínas y actividades enzimáticas. En a) la relación entre gen-actividad es uno a uno; en b) dos genes codifican la misma actividad enzimática de forma que basta con la presencia de uno para que se pueda catalizar EC 2.7.1.12 (operador *or*); en c) *TrpE* y *TrpD* forman un complejo, y se precisan ambos genes para que ocurra 4.1.3.27 (operador *and*); en d) dos isoenzimas pueden catalizar 2.7.1.11 (operador *or*), pero además una de ellas (*PfkA*) puede catalizar una segunda reacción ella sola, la reacción 2.7.1.11 que en este caso tiene dos instancias de reacciones bioquímicas diferentes. Las de arriba a abajo fila indican: el nombre los loci, genes, proteínas y las actividades enzimáticas. Los ejemplos corresponden a las tablas GPR inferidas para *E. coli* K12 MG1655 presentes en el ejemplo propuesto por Thiele and Palsson (2010).

El proceso de reconstruir una red metabólica implica añadir un nuevo nivel a la anotación del genoma, en el cual las distintas actividades enzimáticas codificadas en un genoma se ponen en el contexto de módulos funcionales, *i.e.* rutas metabólicas (Reed et al., 2006a; Stein, 2001). Para reconstruir una red metabólica a partir de un genoma anotado, primero se identifican todos los elementos codificantes del genoma que se encuentran anotados como actividades enzimáticas¹. La relación entre reacciones y genes no es necesariamente de uno a uno debido a la existencia de

¹ A los genes identificados como enzimas se les asocia identificadores de tipo Gene Ontology (GO) o números del Enzyme Commission (EC), que relacionan al gen con una actividad enzimática.

enzimas oligoméricas y de isoenzimas (figura 1.1). La información necesaria para inferir las asociaciones entre genes y reacciones se extrae de la anotación del genoma, y se suelen representar mediante un conjunto de reglas denominadas tablas gen-proteína-reacción (GPR), que relacionan genes con reacciones bioquímicas, mediante operadores lógicos que dan cuenta de la presencia de isoenzimas y complejos (ver figura 1.1). Las actividades enzimáticas derivadas de las asociaciones GPR, son mapeadas en instancias de reacciones bioquímicas concretas y éstas son proyectadas sobre una colección de rutas metabólicas, definida en una base de datos metabólica de referencia². A continuación se emplea un algoritmo de predicción para determinar qué rutas se encuentran codificadas en el genoma (Dale et al., 2010; Romero et al., 2005). El resultado de la reconstrucción es una base de datos que asocia el genoma y sus anotaciones con el conjunto de rutas metabólicas predichas para el organismo. En el modelo relacional empleado en MetaCyc/BioCyc estas bases de datos se denominan *Pathway Genome Data Base* (PGDB).

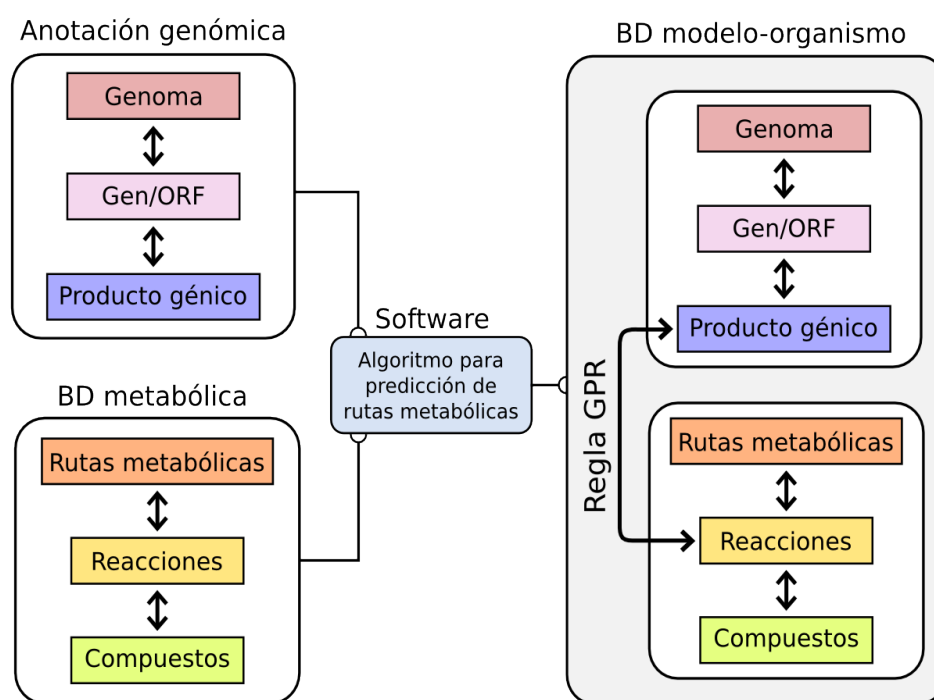


Figura 1.2. Diagrama del modelo relacional de una PGDB. La figura representa la construcción de una PGDB a partir del genoma anotado de un organismo, una base de datos metabólica y un algoritmo para la predicción de rutas metabólicas. ORF: marco abierto de lectura (del inglés, Open Reading Frame).

El diagrama de la figura 1.2 resume el proceso de integración del genoma anotado con una base de datos metabólica que da como resultado una PGDB del organismo. Diversas herramientas permiten construir PGDBs (o esquemas alternativos), a partir de un genoma. Algunos ejemplos incluyen *PathwayTools* (Karp et al., 2015), *ModelSEED* (Henry et al., 2010) o *metaSHARK* (Pinney et al., 2005). Si bien en muchos casos, el objetivo de una reconstrucción metabólica es crear un PGDB de un organismo específico, en la mayoría de ellos el objetivo final es desarrollar un modelo metabólico reconstruido a escala genómica (Genome-Scale Metabolic Model, GEMM)

² Las bases de datos metabólicas contienen información sobre enzimas, reacciones bioquímicas y las rutas metabólicas descritas en la bibliografía y que cubren la mayor parte de la bioquímica conocida, e.g. MetaCyc (Caspi et al., 2014), KEGG (Wixon and Kell, 2000) o RAST-SEED (Overbeek et al., 2014).

para un organismo concreto. Un GEMM es un modelo computacional de la red metabólica de un organismo que permite realizar simulaciones y generar predicciones sobre las capacidades de dicho metabolismo (Overbeek et al., 2014). Un GEMM consta de tres componentes fundamentales: i) la matriz estequiométrica, que define la estructura de la red de reacciones bioquímicas; ii) la tabla de asociaciones GPR derivadas de la anotación genómica del organismo; y iii) una ecuación de biomasa o crecimiento que describe los componentes moleculares que conforman la biomasa del organismo en las proporciones aproximadas en que se encuentran en la célula (Santos et al., 2011; Thiele and Palsson, 2010). Tanto la matriz estequiométrica como las tablas GPR se extraen de un PGDB, mientras que la definición de la ecuación de biomasa implica listar los componentes de las macromoléculas, y las proporciones en que se encuentran en la célula³.

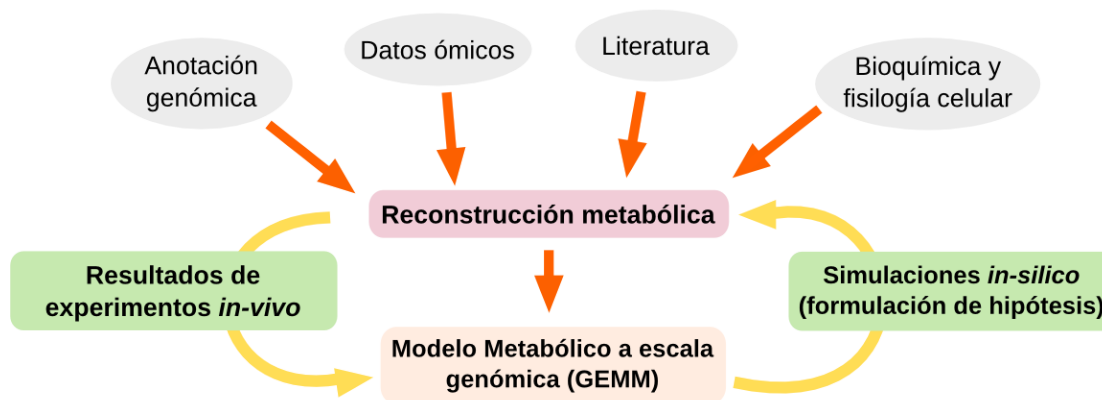


Figura 1.3. Diagrama del proceso iterativo de reconstrucción de un GEMM. El diagrama representa una iteración o ciclo de reconstrucción de un GEMM.

El marco teórico empleado en el análisis de GEMMs se denomina modelado basado en restricciones (Constraint-based Modeling, CBM) (Price et al., 2004). La idea que subyace al CBM es que el comportamiento de un sistema biológico se encuentra limitado o restringido por diversos factores (*e.g.* las leyes de conservación). Estas restricciones se pueden representar con ecuaciones y emplearse para confinar o acotar el espacio de posibles estados, o fenotipos que puede adoptar el sistema (Schellenberger et al., 2011b)⁴. Debido a su potencial aplicación en diversos campos de investigación, el número de modelos metabólicos reconstruidos en forma manual, así como los métodos para su análisis, han crecido de forma ininterrumpida (Lewis et al., 2012; Monk et al., 2014). A su vez, se ha desarrollado una amplia diversidad de métodos computacionales para estudiar problemas que van desde la evolución del metabolismo, al diseño y optimización de organismos de interés industrial (Feist and Palsson, 2008; Ruppin et al., 2010). La reconstrucción de un GEMM es un proceso iterativo en el que las predicciones realizadas por el modelo se pueden contrastar con datos experimentales; las predicciones incorrectas son empleadas para revisar y refinar la formulación del modelo (ver figura 1.3). El nivel de validación más elemental consiste en verificar que el modelo predice la formación de biomasa dado un cierto medio de crecimiento. Si bien existe un desarrollo constante de métodos para automatizar la reconstrucción de GEMMs, *e.g.* Model SEED (Overbeek et al., 2014), TIGER (Jensen et al., 2011), estos suelen dar como resultado un primer borrador de la red, que corresponde a la primera iteración del proceso de reconstrucción.

3 Ver apartado A.1.4 del apéndice A, para más detalles sobre la formulación de la ecuación de biomasa

4 Para una breve introducción a los fundamentos del CBM ver el apéndice A.

Por lo general, en la primera iteración es raro que un GEMM permita predecir crecimiento. La razón más frecuente por la que un modelo no alcanza un nivel tan elemental de validación, es debido a que el modelo es incompleto, *i.e.* carece de una o más reacciones esenciales para producir algún componente de la biomasa. Este tipo de inconsistencia se deben a que la anotación de un genoma es inherentemente incompleta y suele contener errores (Osterman and Overbeek, 2003; Valencia, 2005). Además, más allá del gran avance impulsado por las ómicas, el conocimiento de la bioquímica dista mucho de ser completo y la lista de componentes para la mayoría de los organismos es aún muy incompleta, debido al problema dual de las proteínas de función desconocida y las enzimas “huérfanas”⁵. Como consecuencia, al reconstruir un GEMM todos los genes metabólicos cuyas funciones no han sido predichas, o fueron predichas en forma incorrecta, conducirán a la ausencia de reacciones en la definición del modelo (Orth and Palsson, 2010). Las reacciones ausentes en un GEMM se manifiestan como agujeros o “*gaps*” en la estructura de la red, y representan inconsistencias estructurales que afectaran las predicciones del modelo. Esta clase de inconsistencias puede encontrarse incluso en las propias bases de datos metabólicas empleadas para reconstruir GEMMs (Ponce-de-León et al., 2015).

Una de las aplicaciones del CBM es la detección de inconsistencias así como la identificación de posibles soluciones para las mismas, *e.g.* identificar reacciones candidatas para rellenar los *gaps* (Kumar et al., 2007; Reed et al., 2006b). Si bien tales aproximaciones funcionan adecuadamente al ser aplicadas sobre GEMMs de organismos caracterizados experimentalmente o filogenéticamente próximos a organismos modelo, el problema del completado de *gaps* o *gap-filling* se vuelve más complejo en organismos pertenecientes a grupos poco caracterizados. Además, la mayor parte de los microorganismos requieren de la presencia de otros para poder crecer dando lugar a la formación de comunidades (Faust and Raes, 2012; Gil and Peretó, 2015), *i.e.* redes complejas de interacciones nutricionales en la que que diversos metabolismos se interconectan mediante el intercambio de compuestos o *cross-feeding* (Seth and Taga, 2014). Un caso extremo es la formación de consorcios endosimbiontes donde los metabolismos pueden exhibir un alto grado de complementación (Martínez-Cano et al., 2015; McCutcheon and von Dohlen, 2011). En tales escenarios, el problema del completado de *gaps* o *gap-filling* se vuelve muy complejo debido a que muchas funciones metabólicas están distribuidas entre distintos miembros. Como consecuencia, la presencia de rutas metabólicas incompletas, *i.e.* interrumpidas en algún punto o la presencia *gaps*, pueden no ser producto de errores en la anotación, sino el resultado de la complementación metabólica establecida con otro organismo. Como consecuencia, reconstruir y revisar modelos metabólicos con tales características es una tarea compleja, en la que la representación visual de las inconsistencias puede ser de gran ayuda para desentrañar las complejas interacciones que pueden formar diversos organismos, así como para proponer explicaciones o hipótesis que permitan: por un lado resolver las inconsistencias del modelo; y por otro entender las condiciones que pueden favorecer fenómenos como el *cross-feeding* o la complementación metabólica.

5 Las proteínas de función desconocida son secuencias inferidas a partir de un genoma que no han sido caracterizadas experimentalmente (y lo mismo con las secuencias homólogas en otros genomas) y representan la mitad de las proteínas de los genomas procariotas secuenciados (Niehaus et al., 2015). Las enzimas huérfanas son aquellas actividades caracterizadas experimentalmente para las cuales no se ha reportado un gen asociado, y corresponden, aproximadamente, a un tercio de las actividades descritas (Hanson et al., 2010).

A lo largo de esta Tesis se buscan abordar los problemas metodológicos de la reconstrucción y refinamiento de GEMMs, con énfasis en su aplicación al estudio de las bacterias endosimbiontes. En este sentido, por un lado se busca desarrollar nuevas aproximaciones para la detección y resolución de inconsistencias en la formulación de modelos metabólicos empleando como marco teórico el modelado basado en restricciones. Por otra parte, se trata de responder a algunas preguntas sobre el establecimiento y evolución de bacterias endosimbiontes asociadas a insectos. Estos sistemas suelen exhibir patrones de complementación metabólica, lo que implica que ciertas rutas metabólicas se encuentran distribuidas entre dos o más miembros de la asociación. Este fenómeno plantea una dificultad adicional al problema de la reconstrucción metabólica, ya que muchas funciones aparentemente ausentes podrían no representar errores de anotación, sino que pueden encontrarse codificadas en el genoma de otro miembro de la asociación. A su vez, la detección de eventos de complementación metabólica plantea diversas preguntas como, por ejemplo, si la complementación supone algún tipo de ventaja adaptativa que sea beneficiosa para el sistema huésped-simbionte. Además, la distribución o división de una ruta metabólica entre dos endosimbiontes supone ciertos problemas biofísicos, como son el transporte de intermediarios y los efectos cinéticos que esto conlleva. Como consecuencia, algunos de estos problemas biológicos son también abordados en esta Tesis, ya que se encuentran estrechamente asociados al problema de la reconstrucción de redes metabólicas de organismos y consorcios endosimbiontes. De esta forma, su estudio detenido podría permitir la identificación de algún patrón sobre la emergencia de la complementación metabólica que pueda facilitar la tarea de la reconstrucción de redes en esta clase de sistemas.

La presente Tesis se encuentra estructurada en capítulos, que si bien siguen un orden natural de lectura, estos son, hasta cierto punto, auto-contenidos. Como consecuencia, cada capítulo incluye su introducción correspondiente, los materiales y métodos específicos, y una discusión específica de los resultados presentados. Finalmente se discuten, de forma integrada y en una presentación menos técnica, los principales resultados de la Tesis. Para terminar se incluyen cuatro apéndices: el primero corresponde a una breve introducción al CBM; el segundo incluye las referencias a todo el material suplementario; el tercero corresponde a las herramientas y recursos computacionales empleados en el desarrollo de esta Tesis; y el cuarto lista las publicaciones que han dado lugar a esta Tesis.

Capítulo 2

2 Objetivos

En la presente Tesis se buscan cubrir dos objetivos generales. Por una parte se pretende llevar adelante el desarrollo de metodologías, centradas en el modelado basado en restricciones, para la reconstrucción y análisis de modelos metabólicos a escala genómica. Por otra parte, se buscará aplicar dichas aproximaciones en la reconstrucción y estudio de modelos metabólicos de bacterias endosimbiontes de insectos, haciendo especial énfasis en el fenómeno de la complementación metabólica. De acuerdo con este planteamiento se proponen los siguientes objetivos concretos:

1. Desarrollo de aproximaciones metodológicas para la detección y resolución de inconsistencias estructurales en modelos metabólicos reconstruidos a escala genómica, empleando el modelado basado en restricciones.
2. Detección y resolución de inconsistencias en bases de datos metabólicas, mediante la reconstrucción y análisis de un metamodelo del metabolismo bacteriano.
3. Aplicación de las aproximaciones propuestas a la revisión y análisis un modelo metabólico reconstruido a escala genómica de *Blattabacterium* Bge, endosimbionte de la cucaracha.
4. Reconstrucción de un modelo compartimentado a escala genómica del consorcio endosimbionte del áfido del cedro formado por las bacterias *B. aphidicola* BCc y *S. symbiotica* SCc, para estudiar el fenómeno de la complementación metabólica.
5. Realizar un estudio cinético en un modelo simple, de las condiciones que favorecen la emergencia de complementación metabólica, haciendo énfasis en: i) el papel de la inhibición enzimática; ii) el efecto de permeabilidad de la membrana a los intermediarios de la ruta; y iii) el coste asociado a la síntesis de las enzimas.

Capítulo 3

3 Detección de inconsistencias y refinado de modelos metabólicos

3.1 Introducción

Estudios de secuenciación y análisis de genomas a gran escala, han revelado que entre el 30-40% de las proteínas codificadas en genomas bacterianos no tiene una función conocida (Galperin and Koonin, 2004; Osterman and Overbeek, 2003). Esta proporción aumenta en forma dramática si se consideran los genomas de organismos superiores como plantas o animales (Niehaus et al., 2015). Además, muchas de las funciones “conocidas” carecen de la evidencia soportada por una caracterización experimental; incluso en el caso de organismo modelo extensamente estudiados como *Esherichia coli*, solo hay información experimental para el 54% de los productos génicos (Frishman, 2007). Por su parte, las enzimas huérfanas representan aproximadamente un tercio de los de las actividades con número EC asignado (Hanson et al., 2010). El problema de las partes faltantes del catálogo es un problema abierto que, naturalmente, se encuentra estrechamente relacionado con la reconstrucción de modelos metabólicos ya que como consecuencia de la presencia de actividades metabólicas huérfanas, así como genes de función desconocida, los modelos metabólicos contiene por lo general errores e inconsistencias en su formulación (Orth and Palsson, 2010). La formulación de un modelo puede contener errores de distinta naturaleza, sin embargo, en este capítulo nos centraremos principalmente en aquellas inconsistencias que son consecuencia de errores en la anotación del genoma, o de inconsistencias presentes en las bases de datos metabólicas. Si bien los errores de anotación pueden deberse a distintas causas, estos pueden ser agrupados en dos tipos, de acuerdo con su impacto en un modelo metabólico: 1) ausencia de actividades enzimáticas y/o transportadores que puede deberse a la no anotación, y/o a la asignación funcional incorrecta de uno o más genes; y 2) la asignación incorrecta de actividades funcionales a uno o más genes. Estos errores a nivel de la anotación se propagan al modelo a través de la formulación de la tabla de asociación Gen-Proteína-Reacción (GPR), uno de los componentes que definen al modelo. Los errores en una tabla GPR conducirán en última instancia a la omisión de determinadas reacciones así como a la inclusión errónea de reacciones no presentes en el metabolismo del organismo que se verán reflejados en la definición de la matriz estequiométrica.

En último término, las inconsistencias en un modelo pueden afectar las capacidades predictivas del mismo, dando lugar a falsos positivos y/o negativos al contrastar las predicciones *in-silico* con los resultados experimentales *in-vivo*. Por ejemplo, al contrastar una predicción *in-silico* sobre el efecto que tiene el *knockout* de un determinado gen con el resultado experimental podría ocurrir que el modelo prediga que el gen es letal mientras que el experimento indica que el mutante es viable. Es importante tener en cuenta que los errores en las predicciones pueden también deberse a limitaciones intrínsecas del modelado basado en restricciones, *e.g.* no considerar restricciones impuesta por la regulación y/o la cinética de las reacciones (Chandrasekaran and Price, 2010; Smallbone et al., 2010). Sin embargo, este tipo de problemas son más difíciles de resolver ya que pueden depender de muchos factores y en la mayoría de los casos no hay disponibilidad de datos experimentales sobre regulación y/o cinética a escala genómica, de forma que esta información no es considerada en el presente estudio.

Por otra parte, las inconsistencias que se deben a una formulación incorrecta del modelo pueden ser detectadas y solucionadas empleando el modelado basado en restricciones en lo que se conoce como refinado o curado de un modelo (Kumar et al., 2007; Lewis et al., 2012; Thiele and Palsson, 2010). Si bien, la utilización de resultados experimentales contra los que contrastar las predicciones del modelo es clave en la validación y refinado del mismo, muchas inconsistencias se reflejan en la propia estructura de la red y por lo tanto se pueden detectar mediante el análisis de la matriz estequiométrica. Por su naturaleza, este tipo de inconsistencias pueden ser referidas como estructurales, y como se verá en detalle en las secciones siguientes, estas se manifiestan en los modelos a través de la presencia de metabolitos *gap* y reacciones bloqueadas. El problema de la detección de metabolitos *gap* así como reacciones bloqueadas es referido comúnmente como búsqueda de *gaps* (*gapfinding*) (Brooks et al., 2012; Kumar et al., 2007) o análisis de consistencia. La identificación y resolución de *gaps* es un paso crítico en el proceso de reconstrucción de un modelo metabólico a escala genómica (Henry et al., 2010; Karp et al., 2007; Latendresse et al., 2012; Thiele and Palsson, 2010).

3.2 Detección de reacciones bloqueadas en redes metabólicas

Una de las aplicaciones del modelado basado en restricciones es la detección de inconsistencias en un modelo metabólico (Burgard et al., 2004; Kumar et al., 2007; Lewis et al., 2012; Schuster et al., 2000; Thiele and Palsson, 2010). Como se ha dicho, durante la reconstrucción de un modelo metabólico puede suceder que por distintas causas ciertas actividades enzimáticas no estén incluidas en la anotación del genoma; esto resultará en que el modelo no incluirá determinadas reacciones. A su vez, la ausencia de dichas reacción puede conducir a la aparición de metabolitos *deadend* (del inglés, “callejones sin salida”) (Orth and Palsson, 2010). Algo similar puede ocurrir si durante la construcción del modelo se añaden por error reacciones que no tienen lugar en el metabolismo del organismo. En tales casos, es bastante probable que las reacciones añadidas por error se encuentren desconectadas del resto de la red ya que pueden incluir metabolitos que no forman parte de la red metabólica del organismo y que por tanto serán *deadends*.

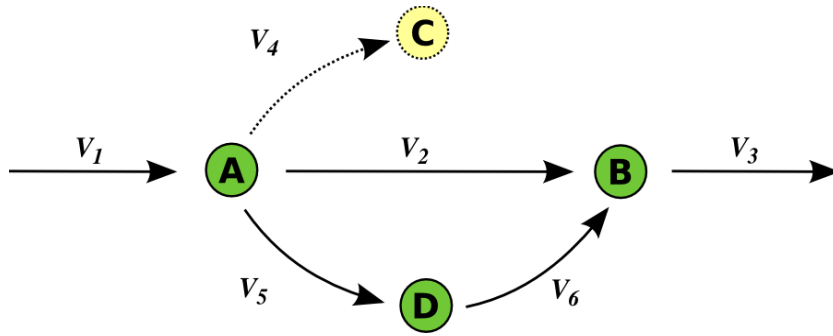


Figura 3.1. Inconsistencias en modelos metabólicos. Representación esquemática de un modelo metabólico que contiene inconsistencias en forma de reacciones bloqueadas y metabolitos deadend. Las flechas indican reacciones y los nodos metabolitos. El nodo amarillo (C) representa un metabolito deadend y como consecuencia la reacción v_4 (flecha punteada) es una reacción bloqueada.

Un *deadend* es un metabolito de la red que sólo es producido/consumido por las reacciones de la red. Como consecuencia, el flujo estacionario a través de todas las reacciones en las que participa deberá ser cero debido a la restricción de balance de masas (ver Apéndice A). De no ser así, el metabolito se acumularía de forma indefinida, o se consumiría completamente, violando la condición estacionario impuesto por el CBM. Por lo tanto, la presencia de metabolitos *deadend* conducirá a que algunas reacciones de la red se bloqueen, *i.e.* que el valor de su flujo se vea restringido a cero. Como se muestra en el ejemplo de la figura 3.1 la ausencia de reacciones que consuman el metabolito C, lo convierten en un *deadend*, y como consecuencia la reacción v_4 se encontrará bloqueada. Dado que es teóricamente posible detectar inconsistencias estructurales mediante el análisis de la matriz estequiométrica, un paso clave durante la reconstrucción de un modelo es verificar si el modelo incluye *gaps*, y en tal caso resolverlos. En este sentido, dado un modelo metabólico con su correspondiente matriz estequiométrica N y su conjunto de reacciones irreversibles, el primer paso para la detección y resolución de inconsistencia estructurales es evaluar si la red es o no consistente. En términos formales una red metabólica se dice que es consistente si todas sus reacciones son consistentes, *i.e.* que cada una de sus reacciones se encuentra activa en al menos un modo de flujo $v \in F$ donde F es el espacio de flujos¹:

Definición 3.1: una reacción j en una red N se define como consistente si se encuentra activa en al menos un modo de flujo, *i.e.* que $\exists v \in F$ tal que $v_j \neq 0$.

En forma complementaria llamaremos reacciones bloqueadas a todas aquellas reacciones que no sean consistentes (Acuña et al., 2009; Burgard et al., 2004):

Definición 3.2: una reacción j en una red N se define como bloqueada si no se encuentra activa en ningún modo de flujo, *i.e.* no existe ningún modo flujo $v \in F$ tal que $v_j \neq 0$.

En base a la definición 3.2 es posible definir el conjunto de reacciones bloqueadas $J_{Blocked}$ de la siguiente forma:

$$J_{Blocked} = \{j \in J : v_j = 0, \forall v \in F\} \quad (\text{Ec. 3.1})$$

- 1 Con F se denota el espacio flujos que incluye todas las distribuciones de flujos compatibles con el estado estacionario y las restricciones de irreversibilidad (ver apéndice A).
- 2 Desde un punto de vista computacional la condición $v_j = 0$ suele planearse como $|v_j| \leq \varepsilon$ para algún umbral positivo ε suficientemente pequeño. Esto es debido al límite de la precisión de un ordenador.

En caso de que la red no sea consistente el problema se centra en la identificación de aquellas reacciones que se encuentren bloqueadas así como las causas del bloqueo. Dada la importancia que tiene el poder identificar inconsistencias en la formulación de un modelo, se han propuestos diversos métodos para la detección de reacciones bloqueadas. Una primera aproximación, se basa en el cálculo de una matriz \mathbf{K} que sea una base generadora, o núcleo del espacio nulo de \mathbf{N} (Heinrich and Schuster, 1998). Encontrar una base \mathbf{K} del espacio nulo equivale a resolver el sistema $\mathbf{N}\cdot\mathbf{v}=0$, y este cálculo se puede realizar de forma eficiente empleando métodos provenientes del álgebra lineal, como por ejemplo la eliminación Gauss-Jordan. En tal caso, las columnas de \mathbf{K} representan soluciones no triviales del sistemas de ecuaciones $\mathbf{N}\cdot\mathbf{v}=0$ que son linealmente independientes y cuya combinación lineal permite construir cualquier solución (ver apartado A.1.2 del apéndice A). Dado que cada fila de la matriz \mathbf{K} representa una reacción de la red, de acuerdo con la ecuación 3.1, aquellas filas cuyas entradas sean todas iguales a cero se corresponden con reacciones bloqueadas. Esto se debe a que si \mathbf{K} es una base del espacio de soluciones y en todos los vectores de dicha base la entrada correspondiente a una velocidad es cero, entonces no es posible construir una solución del sistema en cual dicha variable sea distinta de cero.

Si en los sistemas metabólicos todas las reacciones fuesen reversibles, la aproximación planteada sería completamente válida y permitiría la detección del conjunto completo de reacciones bloqueadas³. Sin embargo, los sistemas metabólicos suelen presentar muchas reacciones, que en condiciones fisiológicas, pueden considerarse irreversibles. Como consecuencia, si bien las columnas de \mathbf{K} son compatibles con la restricción del estado estacionario ($\mathbf{N}\cdot\vec{\mathbf{v}}=0$), estas no son necesariamente compatibles con las restricciones de signo impuestas sobre las reacciones irreversibles, *i.e.* $v_i \geq 0 \forall i \in J_{Irrrev}$ (ver apéndice A). Por tanto los vectores de una base del espacio de soluciones \mathbf{K} no necesariamente se corresponden con modos de flujo, ya que al calcular dicha base solo se considera la restricción de balance de masas impuesto por la estequiometría y se relajan las restricciones de irreversibilidad (*i.e.* se considera que todas las reacciones del sistema son reversibles). Esta simplificación puede dar como resultado, que ciertas reacciones no sean detectadas como bloqueadas si solo se considera la estequiometría y el balance de masas, pero que sí lo sean al considerar la termodinámica.

En la figura 3.2 se muestra un ejemplo muy simple en el cual la detección de reacciones bloqueadas se ve alterada en función de si se consideran o no las restricciones sobre la irreversibilidad de las reacciones. En efecto, la consideración de la irreversibilidad y la dirección en que están definidas las reacciones pone de manifiesto la presencia de una inconsistencia en la formulación de la red. Ahora bien, dado que la relajación reversible del sistema incluye todos los posibles estados del sistema original (incluyendo aquellos que pueden violar las restricciones termodinámicas) si una reacción se detecta bloqueada en el sistema reversible esta necesariamente se encontrará bloqueada en el sistema original⁴. El recíproco, sin embargo, no es necesariamente cierto como puede verse en el ejemplo que se presenta en la figura 3.2. Es decir que al considerar las irreversibilidades pueden aparecer reacciones bloqueadas que no lo estén en el caso reversible.

3 Si todas las reacciones fuesen reversibles, el *kernel* o núcleo de la matriz \mathbf{N} sería una base generadora del espacio \mathbf{F} , y la dimensión del espacio sería igual al número de columnas menos el rango de la matriz.

4 Al añadir nuevas restricciones solo es posible que soluciones del problema original sean incompatibles con las nuevas restricciones, entonces las soluciones del nuevo espacio son también factible en problema original.

Como consecuencia, el conjunto de reacciones bloqueadas detectadas mediante el análisis del espacio nulo de N es un subconjunto de $J_{Blocked}$ y solo en casos muy particulares ambos conjuntos coinciden.

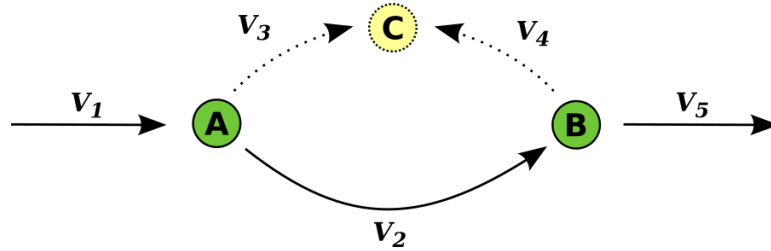


Figura 3.2. Detección de inconsistencias mediante análisis del espacio nulo de N . Si todas las reacciones son irreversibles el metabolito C es un deadend y las reacciones v_4 y v_5 se encuentran bloqueadas. Luego, el único modo de flujo es $v = [1, 1, 1, 0, 0]^T$. Sin embargo, el espacio nulo de N incluye dos elementos en la base, e.g $k^1 = [1, 1, 1, 0, 0]^T$, y $k^2 = [1, 0, 0, 1, -1]^T$ y cada reacción se encuentra presente en al menos un vector, lo que indicaría que no hay reacciones bloqueadas. Sin embargo, k_2 no es un modo de flujo ya que v_5 tiene un valor negativo y esto viola la irreversibilidad.

Dadas las limitaciones del álgebra lineal para incluir restricciones de desigualdad, el desarrollo de métodos para la detección de reacciones bloqueadas se ha basado principalmente en el uso de Programación Lineal (LP) (Acuña et al., 2009; Vlassis et al., 2014) y Programación Entera-Mixta (MILP) (Kumar et al., 2007). La detección del conjunto de reacciones bloqueadas puede realizarse de forma directa iterando sobre todas las reacciones de la red y calculando el valor mínimo y máximo para el flujo de cada reacción mediante la formulación de resolución de un problema lineal sujeto a las restricciones de estado estacionario (ecuación A.3) e irreversibilidades (ecuación A.4). Esta misma idea es la base del Análisis de Variabilidad de Flujos (FVA) (Mahadevan and Schilling, 2003). La única diferencia es que la restricción de óptimo o *cuasi-óptimo* ($v_{obj} \geq v_{obj}^* \cdot \lambda$) impuesta en el FVA no es incluida en la detección de reacciones bloqueadas (ver apartado A.2.2 del apéndice A):

$$\begin{aligned}
 & \text{Min/Max: } v_j \forall j \in J \\
 & \text{s.t.} \\
 & N \cdot v = 0 \\
 & \beta_j \leq v_j \leq \alpha_j \forall j \in J
 \end{aligned}
 \tag{LP 3.1}$$

Después de iterar sobre cada reacción maximizando y minimizando su flujo, aquellas cuyo flujo máximo y mínimo sean cero se encontrarán bloqueadas (Burgard et al., 2004). La mayor desventaja del procedimiento basado en LP 3.1 es que iterar sobre todas las reacciones de la red puede ser muy ineficiente desde el punto de vista computacional. El número total de problemas de optimización a resolver se encuentra entre n (para el caso en que todas las reacciones sean irreversibles) y $2n$ (en el caso extremo en que todas las reacciones fuesen reversibles) siendo n el número total de reacciones de la red. Como consecuencia, el uso de este algoritmo puede volverse prohibitivo en el casos de redes con muchas reacciones como puede ser el caso de la reconstrucción metabólica en humanos (Thiele et al., 2013). Una versión mejorada del algoritmo antes descrito, denominada *fastFVA* (Gudmundsson and Thiele, 2010), consigue acelerar el cálculo empleando una técnica de programación lineal denominada *warm-starts*⁵. Si bien *fastFVA* disminuye considerablemente los

5 La resolución de problemas de programación lineal se puede dividir en dos fases: 1) encontrar una solución factible; 2) encontrar la solución óptima a partir de la solución factible previamente identificada.

tiempos de cálculo, el número de problemas de optimización a resolver sigue siendo el mismo. Como consecuencia, se han propuesto diversos métodos que reducen sensiblemente el número total de problemas de optimización a resolver y por lo tanto el tiempo de cómputo empleado para detectar el conjunto de reacciones bloqueadas.

La idea detrás de estos métodos es que en cada iteración se evalué la consistencia del mayor número posible de reacciones, de forma que se reduzca el número de iteraciones necesarias para evaluar la consistencia de todas las reacciones. Dado que encontrar un modo de flujo en que una reacción tenga una velocidad distinta a cero es condición suficiente para determinar que la reacción es consistente, al resolver un problema lineal, *e.g.* maximizar el flujo en una reacción determinada, se puede evaluar esta condición sobre todas las reacciones de la solución calculada. Luego aquellas reacciones que tengan un flujo distinto de cero indican que son consistentes y no hace falta evaluarlas. Por ejemplo, si en el modelo de la [figura 3.2](#) se maximiza el flujo a través de v_1 (imponiendo una cota superior para que la solución esté acotada), el modo óptimo que se obtiene tiene activas tanto a v_2 como a v_3 . De esta manera con la resolución de un único problema de optimización se prueba simultáneamente la consistencia de múltiples reacciones. La idea de evaluar la consistencia de múltiples reacciones al resolver una única instancia del problema fue también utilizada en el desarrollo de un algoritmo iterativo denominado *CheckModelConsistency* (Jerby et al., 2010). Este busca evaluar la consistencia de múltiples reacciones en cada iteración empleando una función objetivo que maximiza/minimiza la suma total de flujo sobre un subconjunto de reacciones C :

$$\begin{aligned}
 & \text{Min/Max: } \sum_{j \in C} v_j \\
 & \text{s.t.} \\
 & N \cdot v = 0 \\
 & \beta_j \leq v_j \leq \alpha_j \quad \forall j \in C \subseteq J
 \end{aligned}
 \tag{LP 3.2}$$

Inicialmente C contiene todas las reacciones de la red, y en cada iteración se van quitando las reacciones consistentes, de manera que van quedando las reacciones cuya consistencia no ha sido evaluada. Cuando C no se puede reducir más, se minimiza la suma de flujos sobre las reacciones reversibles presentes en C . Finalmente, la consistencia de las reacciones que queden en C debe ser evaluada una a una. De esta forma se identifican las reacciones que no son consistentes. Los métodos basados en LP descritos previamente son exhaustivos, en el sentido de que permiten determinar la consistencia de la red y, en caso de que esta no sea consistente permiten identificar el conjunto completo de reacciones bloqueadas $J_{Blocked}$. Si bien *CheckModelConsistency* es una implementación más eficiente que la basada puramente en FVA, recientemente se han propuesto dos métodos (muy similares en su formulación) denominados *FastConsistencyCheck* (Vlassis et al., 2014) y *OnePrune* (Dreyfuss et al., 2013), que permiten reducir de forma significativa el número de iteraciones respecto a los métodos previamente discutidos, y en consecuencia se reduce el tiempo total del cálculo. La principal innovación introducida por estos métodos es definir una función objetivo que permita maximizar el número total de reacciones activas sobre un subconjunto de reacciones $C \subseteq J$. De esta forma, en cada iteración, con la resolución de un único problema de

Si las modificaciones en la formulación del problema no afectan a la estructura del mismo (*e.g.* re-definir la función objetivo), es posible emplear una solución previamente calculada para inicializar la búsqueda de una nueva solución que sea óptima frente un objetivo alternativo (*warm-starts*) lo que reduce en forma sensible el tiempo de cálculo.

optimización se evalúa la consistencia del mayor número de reacciones posibles, lo que reduce sensiblemente el número de iteraciones, *i.e.* el número total de problemas de optimización a resolver. El problema de optimización que se formula en cada iteración de *FastConsistencyCheck* es el siguiente:

$$\begin{aligned}
 & \max \sum_{i \in C} z_i \\
 & \text{s.t.} \\
 & N \cdot v = 0 \\
 & \beta_j \leq v_j \leq \alpha_j \quad \forall j \in J \\
 & z_j \leq v_j \quad \forall j \in C \subseteq J \\
 & z_i \in [0, \varepsilon]
 \end{aligned}
 \tag{LP 3.3}$$

La novedad en la formulación planteada en LP 3.3 es la introducción de un conjunto de variables auxiliar $z_j \quad \forall j \in C \subseteq J$ que son empleadas para forzar flujo a través de todas las reacciones de la red. Estas variables están acotadas entre cero y un umbral ε lo suficientemente pequeño. Si el valor de ε es lo suficientemente pequeño, maximizar el sumatorio de las variables z_i es equivalente a maximizar el número de elementos distintos de cero (Vlassis et al., 2014). Dado que hay restricciones que vinculan a cada variable auxiliar z_i con su correspondiente reacción v_i , al maximizar la suma de las variables z_i se maximiza el número de reacciones activas. De esta forma se garantiza que en cada iteración se evalúa la consistencia del mayor número posible de reacciones.

Con el objetivo de evaluar de forma comparativa el desempeño computacional de cada método, los cuatro algoritmos para la detección de reacciones bloqueadas fueron implementados en lenguaje Python (ver Apéndice C). Para realizar la evaluación se escogieron dos conjuntos de modelos con tamaños entre ~ 100 y 10.000 reacciones sobre los que aplicar los métodos. Los conjuntos de datos empleados fueron SEED130 que incluye los 130 modelos reconstruido de forma automática, previamente publicados (Henry et al., 2010) (ver apartado 6.2.1 para una descripción detallada) y BiGG76 que incluye los 76 modelos metabólicos reconstruidos y revisados de forma manual, disponibles en la base de datos de modelos metabólicos BiGG (Schellenberger et al., 2010) (accesible a través del enlace: <http://bigg.ucsd.edu/models>). Cada uno de los métodos fue evaluado empleando ambos conjuntos de datos, *i.e.* sobre un total de 206 modelos metabólicos. Finalmente, dado que la detección de reacciones bloqueadas basada en FVA (LP 3.1) es el único método que evalúa, de forma independiente, la consistencia de cada reacción, el resultado obtenido con este método fue empleado como referencia para determinar el número total de falsos positivos y negativos presentado, en caso de que los hubiera, por los otros tres métodos.

La figura 3.3 resume los resultados de la evaluación comparativa de los cuatro métodos (FVA, Análisis del espacio nulo de N , *CheckModelConsistency* y *FastConsistencyCheck*) aplicados sobre los dos conjuntos de datos (SEED130 y BiGG76). Los gráficos indican que los tiempos de cálculo respecto al tamaño del problema escalan de forma similar, independientemente del conjunto de datos empleados. Esto quiere decir que el comportamiento de cada método es consistente y no depende en gran medida de la estructura de los datos. Los resultados muestran de forma clara que el algoritmo con mejor desempeño es *FastConsistencyCheck*. Por otra parte, el desempeño más bajo lo presentó el método basado en el análisis del espacio nulo de la matriz estequiométrica. Esto

puede deberse a que las bibliotecas empleadas para calcular el espacio nulo son *software* libre mientras que el *solver* empleado en la resolución de programas lineales es una implementación comercial muy optimizada (ver Apéndice C). Para todos los casos, la complejidad algorítmica es de tipo polinómica respecto al tamaño del problema y para el caso de los métodos basado en programación lineal la principal diferencia se encuentra en la cantidad de iteraciones o problemas lineales que se deben resolver. Por otra parte, la precisión y la sensibilidad en la detección del conjunto de reacciones bloqueadas fueron también evaluadas para cada método. En este contexto, un falso positivo indica una reacción no detectada como bloqueada que si lo está, mientras que un falso negativo se corresponde al caso de clasificar una reacción como bloqueada cuando no lo está. Los métodos basados en LP (*CheckModelConsistency* y *FastConsistencyCheck*) no generaron ni falsos positivos ni negativos. En todos los casos el conjunto de reacciones bloqueadas predichas coincidió en un 100% con el resultado de referencia obtenido empleando FVA. Sin embargo, el método basado en el análisis del espacio nulo presentó un promedio de ~8% y ~23% de falsos positivos con SEED130 y BiGG76, respectivamente.

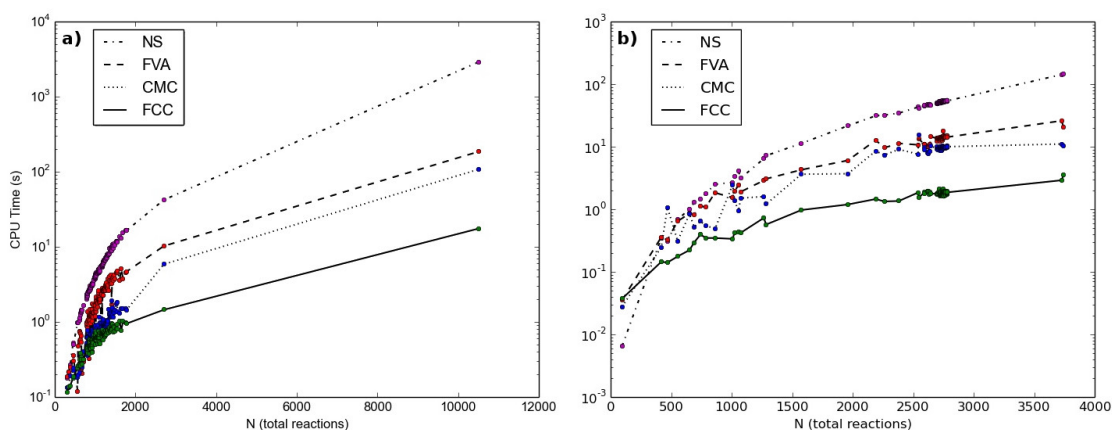


Figura 3.3. Resultados de la evaluación comparativa de métodos para la detección de reacciones bloqueadas. Las curvas representan el tiempo de ejecución de cada método aplicado sobre modelo de tamaños diferentes. El eje x corresponde al número de reacciones de cada modelo y en el eje y el tiempo de cálculo representado en escala logarítmica. En a) y b) se presentan los resultados de la evaluación de los 4 métodos empleando SEED130 y BiGG76, respectivamente. Las leyendas se corresponden: NS a NullSpace; FVA a Flux Variability Analysis; CMC CheckModelConsistency; FCC FastConsistencyCheck.

Como fue discutido previamente en esta sección, la diferencia respecto a los métodos basado en LP se debe a que al emplear un método basado en álgebra lineal se consideran las restricciones de irreversibilidad (ver el ejemplo de la figura 3.2). Esto mismo también explica la gran diferencia en el total de falsos positivos encontrados entre los dos conjuntos: cuanto menor sea el número de reacciones irreversibles en un modelo menor debería de ser el número de falsos positivos, y en el caso extremo en que todas las reacciones fuesen reversibles, el análisis del espacio debería identificar todas las reacciones bloqueadas sin ambigüedad. En el caso del conjunto de datos SEED130 los modelos tienen en promedio ~59% de reacciones reversibles mientras que para el caso de BiGG76 el promedio de reacciones reversibles es del ~29% lo que explica la diferencia en los falsos positivos entre los conjuntos de datos al realizar la búsqueda de reacciones bloqueadas empleando el método basado en análisis del espacio nulo de la matriz estequiométrica. De los resultados obtenidos en la comparación de métodos se puede concluir que cualquiera de los

algoritmos basados en LP identifica en forma exacta el conjunto de reacciones bloqueadas. A su vez, de los métodos evaluados el más eficiente, en términos del tiempo de cálculo es *FastConsistencyCheck*. Esto es debido a que es el algoritmo que emplea el menor número de iteración para evaluar la consistencia de cada reacción de la red. Por otra parte, las reacciones que se identifican como bloqueadas empleando los métodos basados en LP pero que se detectan como consistentes al analizar el espacio nulo de la matriz (*i.e.* al relajar las restricciones de irreversibilidad), puede indicar o bien que alguna de estas reacciones ha sido definida como irreversible cuando en realidad no lo es, o que la reacción es irreversible pero el sentido de la misma ha sido definido de forma incorrecta (figura 3.2).

3.3 Inconsistencias termodinámicas en redes estequiométricas

El FVA no solo se introdujo para detectar reacciones bloqueadas sino que fue empleado para mostrar que el espacio de flujos puede ser no acotado (Mahadevan and Schilling, 2003; Müller and Bockmayr, 2013a). Esto significa que algunas reacciones pueden tomar valores de flujos arbitrariamente grandes⁶. Este tipo de situaciones suelen ocurrir cuando la estructura de una red contiene ciclos internos balanceados, que dan lugar a la presencia de modos de flujo elemental cíclicos. Los ciclos están formados por conjuntos de reacciones que al operar en forma conjunta no producen ni consumen de forma neta de ninguna de las especies involucradas (Gagneur and Klamt, 2004; Schilling et al., 2000). La presencia de tales ciclos puede dar lugar a que distribuciones de flujo óptimas, obtenidas con FBA, pueden contener ciclos internos con flujo neto mayor que cero. Esto implica una violación de la segunda ley de la termodinámica. Esto tiene como consecuencia adicional que el espacio de soluciones óptimas sea degenerado para una cierta función objetivo (Lee et al., 2000; Mahadevan and Schilling, 2003; Müller and Bockmayr, 2013b). Este tipo de distribuciones o modos de flujo, si bien compatibles con el estado estacionario, e incluso con las restricciones de signo impuestas sobre las reacciones irreversibles no corresponden a situaciones reales, ya que incurren en la violación de la segunda ley de la termodinámica. La presencia de modos cíclicos en el espacio de soluciones asociados a una red es una consecuencia de que en el análisis estequiométrico no se considere el balance termodinámico sobre la red de reacciones, *i.e.* no hay ninguna restricción que excluya este tipo de situaciones.

En muchas situaciones la irreversibilidad de una reacción es evaluada en ciertas condiciones experimentales que pueden no ajustar a las condiciones fisiológicas de un determinado organismo y en tal caso esta pueda operar de forma reversible. Para los casos en que no se dispone de información experimental sobre la reversibilidad de una reacción se han desarrollado métodos computacionales como el método de contribución por grupos (Jankowski et al., 2008), que permiten estimar la reversibilidad de las. En la sección previa se mostró cómo el considerar como irreversibles ciertas reacciones puede poner de manifiesto inconsistencias en el modelo. En forma equivalente, al considerar el balance energético que del sistema es posible identificar otro tipo de inconsistencias. A diferencia con la restricción de balance de masas (ecuación A.3) o las restricciones de irreversibilidad (ecuación A.4), la incorporación de restricciones sobre la conservación de la energía resulta problemática debido a que las ecuaciones que representan estas

6 En la práctica tan grande como la cota que se imponga sobre el valor de los flujos de dichas reacciones.

restricciones son no-lineales y/o no-convexas (Flamholz et al., 2012; Henry et al., 2007; Jankowski et al., 2008). Desde un punto de vista práctico, el problema que conlleva no considerar el balance termodinámico es que las distribuciones de flujo que se obtengan, si bien compatibles con el balance de masas, pueden ser no factibles de un punto de vista termodinámico. Las restricciones termodinámicas en los flujos de las reacciones de un sistema pueden plantearse de la siguiente manera. Dado un sistema con un conjunto I de metabolitos y J de reacciones, un modo de flujo v es termodinámicamente factible si existen un conjunto de potenciales μ_i que cumplan (Beard et al., 2004):

$$\Delta \mu_j = \sum_{i \in I} \mu_i \cdot N_{ij} \quad \forall j \in J_{Int} \quad (\text{Ec. 3.2})$$

$$\Delta \mu_j \cdot v_j < 0 \quad \vee \quad v_j = 0 \quad \forall j \in J_{Int} \quad (\text{Ec. 3.3})$$

donde J_{Int} corresponde al conjunto de reacciones internas del sistema. La ecuación 3.2 permite calcular el incremento en el potencial químico de cada reacción a partir de los potenciales químicos de reactivos y productos ponderados por el coeficiente estequiométrico correspondiente⁷. Por su parte, la ecuación 3.3 introduce una restricción que vincula el incremento de potencial con el signo en el flujo de cada reacción. Esta restricción “fuerza” que el flujo debe tener el sentido inverso al incremento de potencial, de forma que al ocurrir la reacción haya una disminución en el potencial, o que el flujo neto en la reacción sea cero⁸

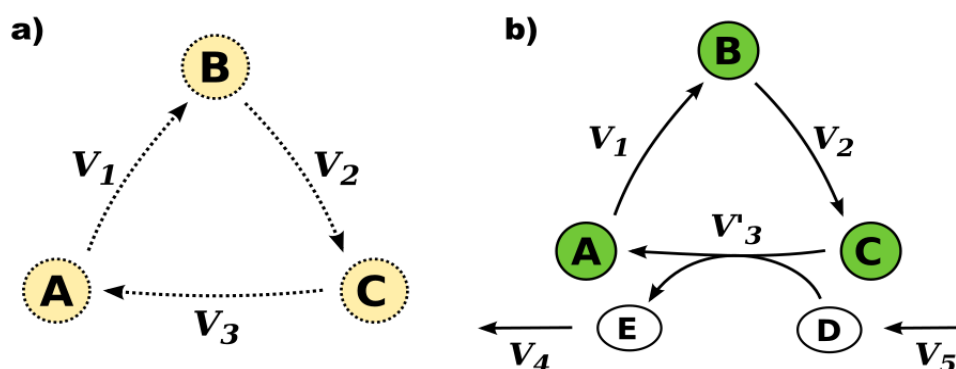


Figura 3.4. Inconsistencias termodinámicas y reacciones pseudo-bloqueadas. El diagrama en a) representa un ciclo metabólico cerrado, cuyo único modo de flujo estacionario corresponde a una vía extremas de tipo III. En b) se representa el mismo sistema que en a) pero v_3 se encuentra acoplada a un proceso externo que consume energía y en consecuencia el ciclo metabólico es termodinámicamente factible, si el potencial entre D y E es mayor que aquel entre A y C, en cuyo caso el modo de flujo se corresponde con la definición bioquímica tradicional de ciclo fútil.

En la figura 3.4 se representan dos casos de ciclos que pueden tener lugar en un modelo metabólico. Si bien desde el punto de vista estequiométrico ambos ciclos son consistentes, *i.e.* que para ambos existe un modo de flujo compatible con el estado estacionario, el ciclo presentado en figura 3.4a no es factible desde el punto de vista termodinámico. Esto se debe a que la suma de potenciales en un ciclo cerrado debe ser igual a cero, como consecuencia de la segunda ley de la termodinámica. De acuerdo con la ecuación 3.3, para que exista flujo en una reacción su

7 En realidad, no es otra cosa que la afinidad química de la reacción cambiada de signo.

8 Dicho de otro modo, la afinidad química de la reacción debe tener el mismo signo que el del proceso que impulsa.

incremento de potencial debe ser diferente de cero. Esta restricción puede interpretarse en forma análoga a la segunda ley de Kirchoff para circuitos eléctricos (Price et al., 2002). Considerando las reacciones del ejemplo de la [figura 3.4](#), de acuerdo con las ecuaciones 3.2 y 3.3 para que haya flujo que convierta A en B ($v_1 > 0$) el potencial de A debe ser mayor que el de B, *i.e.* $\mu_A > \mu_B$; por otra parte, si se considera flujo de B a C ($v_2 > 0$) la relación entre los potenciales químicos de ambos compuestos debe cumplir $\mu_B > \mu_C$. Finalmente, aplicando la propiedad transitiva se obtiene la siguiente relación $\mu_A > \mu_C$ que al considerar la [ecuación 3.3](#) implica $v_3 < 0$, es decir que no es posible que exista flujo cíclico. Como consecuencia, el único estado estacionario compatible con el balance termodinámico es el equilibrio, en el cual el flujo neto a través de cada reacción es igual a cero. Contrariamente, en el caso presentado en la [figura 3.4b](#) el ciclo se encuentra acoplado a una fuente externa de energía. Si la diferencia de potencial entre D y E es mayor entre A y C el modo de flujo a través del ciclo será termodinámicamente factible y podrá existir un flujo mayor que cero en todas las reacciones.

Dado que en ausencia de restricciones adicionales, las reacciones que participan en ciclos cerrados pueden tomar valores de flujos extremos, este tipo de artefactos puede afectar la interpretación de las predicciones realizadas empleando un modelo metabólico, *e.g.* al comparar los valores de flujos predichos con los niveles de expresión de los genes asociados (Lewis et al., 2010). Se han propuesto distintas aproximaciones para incorporar restricciones termodinámicas en GEMM. Sin embargo, estas implican la resolución de problemas computacionalmente demandantes. Como consecuencia su aplicación a redes reconstruidas a escala genómica se ve limitada. Algunos ejemplos incluyen la formulación de un problema de optimización global continua no-convexa (Beard et al., 2002; Fleming et al., 2010; Nagrath et al., 2007), la resolución de un problema NP-Complejo (Yang et al., 2005), así como la resolución de un problema de programación entera mixta (Schellenberger et al., 2011a). La incorporación de tales restricciones permite excluir los ciclos internos al computar soluciones óptimas (Beard et al., 2002; Fleming et al., 2012). De forma adicional, se han propuesto métodos específicos para la enumeración de ciclos incompatibles con el balance termodinámico así como para la eliminación los ciclos presentes en soluciones óptimas, sin afectar el valor de la función objetivo.

Si bien se ha prestado gran atención a la detección de los ciclos, así como al problema de eliminar su efecto al computar distribuciones de flujo óptimas, se ha puesto menos atención en el hecho de que la mera existencia de ciclos internos puede también indicar la presencia de inconsistencias en la formulación del modelo. En este sentido, la atención se ha centrado en el problema de computar una distribución o modo de flujo que sea consistente desde un punto de vista termodinámico (Lewis et al., 2010; Müller and Bockmayr, 2013a). Sin embargo, si la atención se centra en la consistencia de las reacciones, es posible plantear el siguiente problema: dada una reacción consistente en términos estequiométricos, ¿es la reacción termodinámicamente consistente? O dicho de otra forma, si además de considerar el balance de masas se toma en cuenta el balance energético de la red ¿existe algún modo de flujo compatible con el balance energético de la red en el que la reacción tenga flujo distinto de cero?

En la sección previa se introdujo una definición de reacción bloqueada y se describieron métodos para su identificación. Si bien la [definición 3.1](#) permite la detección de las reacciones inconsistentes de forma muy eficiente, empleando métodos basados en programación lineal, esta definición no

considera la existencia de modos de flujo cíclicos que violen la segunda ley de la termodinámica. Esto implica que al evaluar la consistencia de una reacción particular solo se considera el flujo de la reacción y no la estructura del modo de flujo en el que se encuentra embebida. Una reacción puede encontrarse activa si existe al menos un modo de flujo para el cual la reacción tiene un flujo distinto de cero. Sin embargo, alguna reacción sólo está activa en modos cíclicos, esto indica la presencia de una inconsistencia en la formulación del modelo. Esto se debe a que tales modos carecen de sentido físico

En la [figura 3.5](#) se representa un diagrama compuesto por cuatro reacciones irreversibles y dos metabolitos. Los posibles estados del sistema compatibles con el estado estacionario y con la restricción de irreversibilidad sobre las reacciones pueden descomponerse en un total de dos modos de flujo elementales representados en las [figuras 3.5bc](#) (ver [Apéndice A](#)). El caso planteado en [figura 3.5b](#) se corresponde con un modo de flujo que importa A lo convierte en B y lo excreta o elimina del sistema. Por otra parte, [figura 3.5c](#) representa un modo de flujo cíclico en el que no hay intercambio de materia ni energía con el exterior. En efecto, por tratarse de un ciclo la masa total de los metabolitos involucrados se conserva, y dado que el ciclo es cerrado y no se encuentra acoplado a ningún proceso que transfiera energía para que haya flujo, el único estado compatible con la termodinámica es el equilibrio en el que todos los flujos son iguales a cero.

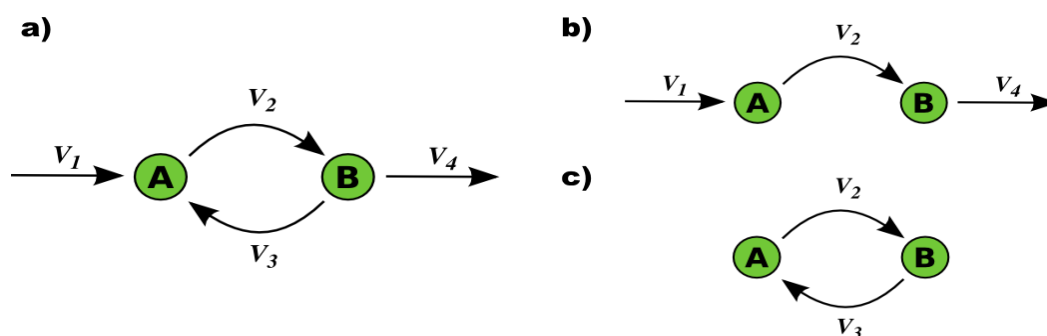


Figura 3.5. Modos de flujo cíclicos y reacciones pseudo-bloqueadas. El diagrama en a) representa un sistema metabólico compuesto por cuatro reacciones irreversibles y dos metabolitos. En b) se representan los modos de flujo elementales asociados al sistema presentado en a).

En el caso de la reacción v_2 dado que esta está activa en un modo compatible no solo con las restricciones de estado estacionario e irreversibilidad, sino que además es compatible con las ecuaciones de balance energético (ecuaciones 3.2 y 3.3) esta reacción es consistente en un sentido más amplio que el planteado en la [definición 3.1](#) ([figura 3.5b](#)). Sin embargo, la reacción v_3 solo se encuentra activa en un modo elemental cíclico que viola el balance energético. Por consiguiente la reacción v_3 no es consistente desde un punto de vista termodinámico. Además en este caso es muy fácil ver la incongruencia termodinámica ya que v_2 y v_3 no pueden ser ambas positivas, por la cuestión de los potenciales, luego una de ellas tiene que ser cero. No puede ser v_2 , porque entonces daría lugar a *deadend*, luego v_3 tiene que ser cero. De acuerdo con las consideraciones acerca de la termodinámica, las definiciones de reacción consistente/bloqueada pueden ser extendidas o generalizadas de forma que consideren el balance energético que impone la segunda ley de la termodinámica. De esta manera, aquellas reacciones que solo se encuentran activas en modos de flujo que son termodinámicamente no factibles, *i.e.* que solo se encuentran activas en modos de flujo cíclicos, estarán bloqueadas o serán inconsistentes desde un punto de vista termodinámico.

Definición 3.3: Una reacción j en una red N con un conjunto de reacciones irreversibles J_{Irr} se dice termodinámicamente consistente si y sólo si existe al menos un modo de flujo elemental $\vec{e} \in F$ tal que $e_j \neq 0$ y \vec{e} es un modo de flujo elemental no cíclico.

De la definición se desprende que las reacciones termodinámicamente no consistentes, que para simplificar serán referidas como *t-bloqueadas*, serán aquellas reacciones que solo se encuentran activas en modos de flujo cíclicos. Luego el conjunto de reacciones candidatas a encontrarse *t-bloqueadas* se reduce al conjunto de reacciones que pueden encontrarse activas en al menos un ciclo interno. Este conjunto de reacciones que denotaremos como J_{loop} puede identificarse de forma sencilla, mediante la detección de reacciones bloqueadas en el sistema cerrado, *i.e.* forzando a cero el valor de todos los flujos de intercambio (Wright and Wagner, 2008). Aquellas reacciones que se encuentren activas en el sistema cerrado son capaces de formar ciclos internos. Es importante destacar que si bien las reacciones *t-bloqueadas* (de haberlas) deberán estar contenidas en J_{loop} , la afirmación contraria puede no ser válida, es decir que una reacción que pueda formar ciclos internos no necesariamente se encuentra *t-bloqueada* (ver el caso de v_2 en la figura 3.5). Una forma directa para detectar reacciones *t-bloqueadas* puede realizarse empleando el análisis de modos de flujo elementales:

- 1) Se obtiene la matriz de modos de flujo elementales E
- 2) Se inspeccionan las filas y se eliminan aquellas cuyas entradas sean todas cero ya que se corresponden con reacciones estequiométricamente bloqueadas.
- 3) Se identifican los modos elementales que forman ciclos internos, *i.e.* los modos que no tengan activo ningún flujo de intercambio, y se eliminan dichas columnas.
- 4) Se inspeccionan las filas de la matriz obtenidas en 3). Aquellas filas cuyas entradas sean todas cero se corresponden a reacciones que solo participan en ciclos internos.

De acuerdo con la definición 3.3 el conjunto de reacciones obtenido en 4) son las reacciones *t-bloqueadas*. El problema con esta aproximación es que el número de modos de flujo elemental crece de forma exponencial con el tamaño de la red y por lo tanto el cálculo de modos elementales es computacionalmente intratable. Esta aproximación, en principio, simple y directa, no es aplicable a la mayoría de los GEMM debido a la explosión combinatoria asociada al cálculo de modos elementales. Como se ha mencionado previamente en esta sección, distintos autores han propuesto diversas formulaciones que permiten introducir las restricciones termodinámicas impuestas por la ecuación 3.3. En particular, Schelleberg y colaboradores han introducido una forma de formular las restricciones termodinámicas planteadas en las ecuaciones 3.2 y 3.3 en términos de programación entera mixta (Schellenberger et al., 2011a). Dichas restricciones se aplican únicamente sobre aquellas reacciones J_{Loop} .

$$\begin{aligned}
 -C a_i + (1 - a_i) &\leq g_i \leq -a_i + C(1 - a_i) \\
 -C(1 - a_i) &\leq v_i \leq C a_i \\
 K_{Int} \cdot \vec{g} &= 0 \\
 a_i &\in \{0, 1\} \quad \forall i \in J_{Loop} \\
 g_i &\in \mathbb{R} \quad \forall i \in J_{Loop}
 \end{aligned}
 \tag{Ec. 3.4}$$

Donde K_{int} se corresponde con la base del espacio nulo del sistema cerrado, es decir una base generadora del espacio de ciclos internos. En conjunto, estas restricciones impiden que las soluciones contengan ciclos internos. La mayor desventaja de esta formulación es que las restricciones implican la necesidad de introducir variables binarias (enteras) por lo que el problema pasa de ser puramente lineal a ser entero-mixto, y la resolución de problemas de este tipo es computacionalmente costosa. Sin embargo, si el número de variables enteras no es muy grande la resolución de problemas de optimización que incluyan estas restricciones es abordable y los tiempos de cálculo no son excesivamente grandes. La principal ventaja de la aproximación es que permiten aplicar distintas estrategias de optimización, *e.g.* FBA (Varma and Palsson, 1994), MOMA (Segrè et al., 2002), ROOM (Shlomi et al., 2005), etc, mientras que garantizan la consistencia termodinámica de la solución calculada (ver apartados A.2.1 y A.2.5 del apéndice A). En particular, estas restricciones pueden combinarse con cualquiera de los métodos de detección de reacciones bloqueadas basados en LP. A partir de esta combinación es posible plantear el siguiente procedimiento para identificar reacciones *t-bloqueadas*:

- 1) Se identifica el conjunto de reacciones estequiométricamente bloqueadas empleando alguno de los algoritmos basados en LP (*e.g.* *FastConsistencyCheck*).
- 2) Se elimina las reacciones bloqueadas del modelo.
- 3) Se realiza la detección de las reacciones que formen ciclos J_{loop} Empleando FVA.
- 4) Se incorporan las restricciones representada por la [ecuación 3.4](#).
- 5) Se evalúa la consistencia de las reacciones del subconjunto de reacciones J_{loop} .

Aquellas reacciones que sean identificadas como bloqueadas, lo están como consecuencia de las restricciones que supone la [ecuación 3.4](#) y de acuerdo con la [definición 3.3](#) son reacciones termodinámicamente inconsistentes, ya que solo se encuentran activas en ciclos internos.

Con el objetivo de evaluar el procedimiento propuesto éste fue implementado y aplicado sobre tres versiones del modelo metabólico de *E. coli* K12 *MG1655*: *iJR904* (Reed et al., 2003), *iAF1260* (Feist et al., 2007a) e *iJO1366* (Orth et al., 2011). Las tres versiones del GEMM de *E. coli* contienen reacciones *t-bloqueadas* siendo la versión más reciente la que contiene el mayor número con un total de cinco. En la [figura 3.6](#) se presentan los tres ciclos internos en los que participan las cinco reacciones *t-bloqueadas* que fueron detectadas en *iJO1366*. La validez de los resultados fue evaluada corroborando que la única distribución en la que las cinco reacciones *t-bloqueadas* están activas se corresponden únicamente con uno de los tres ciclos internos presentados en la [figura 3.6](#). De acuerdo a la base de datos EcoCyc (Karp et al., 2007) las reacciones *t-bloqueadas* EC 2.3.1.222 y EC 2.7.2.15 del caso presentado en [figura 3.6a](#), son los pasos finales de la vía de degradación de la L-treonina I (ver [figura suplementaria S.3.1](#)). Esta permite la degradación de la L-treonina en propanoato, generando ATP (mediante fosforilación a nivel de sustrato) y permitiendo el crecimiento de la bacteria empleando L-treonina como única fuente de carbono (Luginbuhl et al., 1974). Por su parte la reacción EC 6.2.1.17 que completa el ciclo de la [figura 3.6a](#) es el primer paso del *ciclo I del ácido 2-metilcitrato* (ver [figura suplementaria S.3.1b](#) en el Apéndice B). Esta ruta permite a *E. coli* emplear el propanoato como fuente única de carbono y energía (Brock et al., 2002; Textor et al., 1997). La inconsistencia surge en que el propanoil-CoA es un intermediario en

la vía de degradación de la L-treonina I y también lo es del ciclo I del ácido 2-metilcitrato. Dado que en el *iJO1366* el propanoato solo puede ser consumido por EC 6.2.1.17, esto implica que si EC 2.3.1.222 y EC 2.7.2.15 están activas, EC 6.2.1.17 también debe estarlo o en caso contrario el propanoato se acumularía con efectos negativos ya que es un compuesto tóxico para la célula (Horswill et al., 2001).

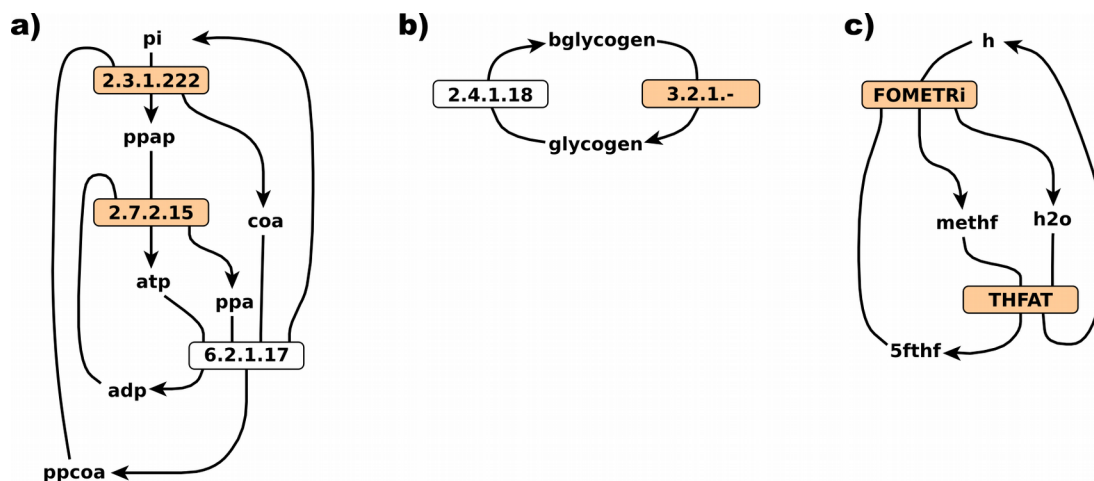


Figura 3.6. Modos de flujo cíclicos y reacciones t-bloqueadas en *iJO1366*. En la figura se representan los tres modos cíclicos que incluyen las cinco reacciones t-bloqueadas detectadas en *iJO1366*. Los rectángulos color anaranjado denotan las reacciones t-bloqueadas mientras que las blancas son reacciones consistentes que además de formar parte en ciclos pueden encontrarse activas en modos elementales no cíclicos. Las reacciones t-bloqueadas presentes en a) fueron detectadas en las tres versiones del GEMM de *E. coli*. Las reacciones t-bloqueadas presentes en b) fueron detectadas como tal en *iAF1260* y en *iJO1366*. Las reacciones t-bloqueadas presentes en c) solo se detectaron en *iJO1366*. Las abreviaciones de los metabolitos son las siguientes: fosfato (*pi*), propanoato (*ppa*), propanoil-fosfato (*ppap*), propanoil-CoC (*ppcoa*), glucógeno ramificado (*bglycogen*), glucógeno (*glycogen*), H^+ (*h*) *N*⁵-formil-tetrahidrofolato ciclo-ligasa (*5fthf*), 5,10-meteniltetrahidrofolato (*methf*).

Pero si debido a que forman un ciclo cerrado la presencia de las tres reacciones activas en simultáneo no es termodinámicamente factible, el papel de las enzimas EC 2.3.1.222 y EC 2.7.2.15 quedaría incierto ya que tal como se encuentra formulada esta vía, tanto en el *iJO1366* como en el EcoCyc las enzimas EC 2.3.1.222 y EC 2.7.2.15 no podrían funcionar. Una posibilidad sería que el propanoato pudiese ser excretado por la célula lo que permitiría el funcionamiento de la degradación de la L-treonina a propanoato y este último, generado por EC 2.7.2.15, podría ser luego eliminado de la célula. En efecto, si se realiza dicha modificación en *iJO1366*, las reacciones EC 2.3.1.222 y EC 2.7.2.15 dejan de ser inconsistentes desde un punto de vista termodinámico y pasan a desempeñar una función plausible, de acuerdo con el modelo. En cualquier caso, para comprender el rol fisiológico de dichas actividades se requeriría realizar experimentos combinando condiciones de crecimiento con *knockouts* específicos. Cabe destacar que la detección de inconsistencia de este ejemplo ha permitido generar una posible hipótesis sobre la función de dos actividades enzimáticas que indican una inconsistencia.

Los casos presentados en la figuras 3.6b-c se corresponden con artefactos. En concreto, las reacciones presentadas en figura 3.6b realizan la interconversión de glucógeno ramificado en glucógeno y viceversa. Los metabolitos como el glucógeno no pueden ser tratados en forma adecuada en el análisis estequiométrico ya que para modelar un polímero se debería incluir en el

modelo un metabolito para cada posible grado de polimerización, lo cual no es abordable. En general, la inclusión de reacciones iterativas involucradas en la producción y degradación de polímeros puede ser problemática y conducir a artefactos. Como consecuencia, los modelos estequiométricos en general no suelen incluir reacciones iterativas excepto para el caso de la síntesis de componentes de la membrana, *e.g.* lipopolisacáridos, peptidoglicano (para un ejemplo concreto ver la [apartado 4.3.1](#)). El caso de la [figura 3.6c](#) está compuesto por dos reacciones con identificadores THFAT y FOMETRi, las cuales no tienen un número EC asociado en el modelo *iJOR1366*. El ciclo interno se forma debido a que ambas son la misma reacción pero definidas en sentidos opuesto; además el N⁵-formil-tetrahidrofolato solo participa en estas dos reacciones. Claramente, esta situación representa un artefacto. De acuerdo con la tabla de asociaciones GPR del modelo *iJOR1366*, la reacción THFAT tiene asociado el gen b2551, que codifica la serina hidroximetiltransferasa. La definición de la reacción en el modelo coincide con la de EcoCyc, donde además se indica que la reacción no se encuentra asociada a ninguna vía metabólica descrita en la literatura. Por su parte, FOMETRi se encuentra codificada por b2912 y cataliza la reacción inversa. Sin embargo al consultar EcoCyc, el producto de b2912, la N⁵-formil-tetrahidrofolato ciclo-ligasa cataliza la reacción EC 6.3.3.2 que participa en la vía de síntesis del N¹⁰-formil-tetrahidrofolate:



A diferencia con FOMETRi, la definición de EC 6.3.3.2 implica consumo de ATP. Luego si se corrige la formulación de FOMETRi y se añade el ATP como un sustrato, el ciclo presentado en la [figura 3.6c](#) dejaría de ser un ciclo interno, termodinámicamente no factible, para convertirse en un ciclo fútil que disipa energía, lo que posiblemente también represente un artefacto en la formulación del modelo. Las reacciones *t-bloqueadas* de la [figura 3.6a](#) se encuentran presentes en las tres GEMMs; las reacciones *t-bloqueadas* de la [figura 3.6b](#) se detectaron en *iAF1260* y en *iJO1366* y finalmente las presentes en la [figura 3.6c](#) se identificaron únicamente en *iJO1366*. Dado que los tres modelos han sido publicados como actualizaciones de una reconstrucción inicial (Edwards and Palsson, 2000), este resultado muestra que las inconsistencias termodinámicas presentes en las versiones más antiguas persisten en las versiones más recientes. Si bien cinco reacciones *t-bloqueadas* en un modelo como *iJO1366* que contiene ~2000 reacciones es un número muy bajo, es importante notar que dicho modelo se corresponde a la última versión del GEMM del organismo modelo posiblemente más estudiado. Por tanto, es esperable que la presencia de reacciones *t-bloqueadas* en GEMM de organismos menos estudiados sea mayor.

Después de validar el método para detectar reacciones *t-bloqueadas*, este fue aplicado para detectar este tipo de inconsistencias en distintos GEMMs, empleando los conjuntos de datos SEED130 y BiGG76, antes descritos. Los resultados del análisis, resumidos en la [tabla 3.1](#), indican que en ambos conjuntos de datos existen diversos modelos que contienen reacciones *t-bloqueadas*. En particular, en ambos conjuntos de datos prácticamente todos los modelos contienen al menos una reacción *t-bloqueada*. Además, al analizar los modelos de BiGG76 se identificadas 55 reacciones que se encuentran bloqueadas en el 100% de los modelos en que se encuentran presentes. Esto podría sugerir que estas inconsistencias se encuentran presentes en la propia base de datos metabólica empleada para la reconstrucción de los modelos. Los resultados del análisis indican que prácticamente la mayoría de los modelos contiene reacciones *t-bloqueadas* (~97% en

ambos casos). Sin embargo, los modelos del BiGG76 contienen muchas menos reacciones *t-bloqueadas* que las que se encuentran en los modelos del SEED130. En efecto, al comparar los modelos con el mayor número de reacciones *t-bloqueadas* en cada conjunto de datos, se ve que el modelo del SEED130 correspondiente a *B. xenovorans* LB400 (Seed36873.1) contiene 562 reacciones *t-bloqueadas* de un total de 1425 (~40%) mientras que el el modelo RECON1 solo exhibe 26 de un total de 3275 (~0.8%). Además, el número total de reacciones que son detectadas como *t-bloqueadas* en al menos un modelo del SEED130 es considerablemente mayor que las que se encuentran en los modelos del BiGG76.

Tabla 3.1. Resultados del detección de reacciones *t-bloqueadas* en BiGG76 y SEED130. La primera fila indica el número total de reacciones que fueron detectadas como *t-bloqueadas* en al menos un GEMM; la segunda fila refiere al número de reacciones que se encuentran *t-bloqueadas* en todos los casos en que están presentes; la tercera fila indica el máximo número de modelos en que una reacción fue detectada como *t-bloqueadas*; la cuarta fila denota el total de modelos que contienen al menos una reacción *t-bloqueada*; y la quinta fila indica el GEMM con mayor número de reacciones *t-bloqueadas*. Las dos últimas columnas corresponden a los resultados obtenidos con cada conjunto de datos.

Descripción	BiGG76	SEED130
Nº total de <i>t-bloqueadas</i>	100	648
Nº siempre <i>t-bloqueadas</i>	55	0
Nº máximo de GEMMs que ocurre una misma <i>t-bloqueada</i>	58	95
Nº de GEMM con reacciones <i>t-bloqueadas</i>	74	126
Mediana en el total de reacciones <i>t-bloqueadas</i> por GEMM	5	7
Nº máximo de <i>t-bloqueadas</i> en un único GEMM	26 *	562 **

* *Homo sapiens* RECON 1

** *Burkholderia xenovorans* LB400 (ID: Seed36873.1)

Si bien los modelos en ambos conjuntos de datos presentan inconsistencias termodinámicas, es claro que las mismas se reducen en gran medida en los modelos reconstruidos y revisados en forma manual, mientras que la presencia de estas inconsistencias es mucho mayor en los modelos reconstruidos en forma automática. En el caso de los modelos reconstruidos en forma automática es posible que los algoritmos de *gap-filling* basados en optimización y empleados sin supervisión resuelvan las inconsistencias estequiométricas, *i.e.* reacciones bloqueadas como consecuencia de la presencia de metabolitos deadends, cerrando ciclos, ya que tales soluciones solo implican incorporar un número reducido de reacciones. En el caso de los modelos reconstruidos manualmente, el problema de las reacciones *t-bloqueadas* es considerablemente menor aunque, como se ha visto, un análisis detallado puede conducir a revisar el papel fisiológico de algunas enzimas. En cualquier caso, la definición de reacción *t-bloqueada* extiende el concepto de reacción bloqueada más comúnmente empleado y el procedimiento propuesto a partir de dicha definición permite su detección.

3.4 Definición y detección de metabolitos *gap*

Si bien el balance termodinámico de la red puede causar el bloqueo de reacciones, la causa más común son las inconsistencias estequiométricas. En efecto, una reacción bloqueada suele ser la consecuencia de que algún metabolito de la red se encuentra únicamente producido o consumido. En tales casos, la restricción de balance de masas forzará a cero el flujo de todas reacciones involucradas. Por tal motivo, dichos metabolitos suelen ser referidos como *deadends* ya que, tal como se comentó en la [sección 3.1](#), representan un “callejón sin salida” para el flujo de las reacciones que lo producen/consumen. La presencia de metabolitos *deadend* y de reacciones bloqueadas puede indicar la ausencia de una o más actividades enzimáticas (o transportes) que deberían estar presente en la red, o la presencia de una o más reacciones añadidas al modelo por error. Desde un punto de vista topológico, los metabolitos *deadend* se pueden agrupar en dos clases: i) los metabolitos que solo son consumidos por reacciones (transportes) de la red (metabolito A en el ejemplo de la [figura 3.7a](#)) y se denominan *metabolito terminal no producido* (RNP, *Root-Non-Produced*); y los metabolitos que solo son producidos por reacciones del sistema (ver metabolito H del ejemplo en la [figura 3.7b](#)) y que por tanto se denominan *metabolito terminal no consumido* (RNC, *Root-Non-Consumed*).

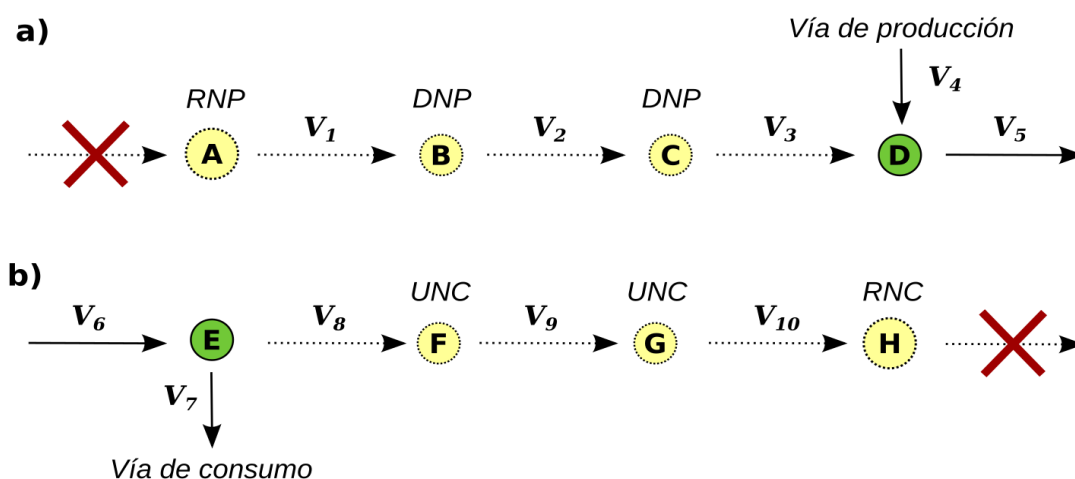


Figura 3.7. Descripción de las clases de metabolito *gap*. Representación de las cuatro clases de metabolitos *gap* en modelos metabólicos. Las cruces rojas indican reacciones ausentes en el modelo. Las flechas punteadas y continuas representan reacciones bloqueadas y no-bloqueadas, respectivamente. Los círculos amarillos y verdes corresponden a metabolitos *gaps* y no-*gaps*, respectivamente. Los *gap* están etiquetados con su clase correspondiente. En a) la ausencia de una o más causa que el metabolito A sea un metabolito RNP y que esta inconsistencia se propague hacia abajo generando nuevos metabolitos *gap* de tipo DNP y nuevas reacciones bloqueadas. En b) se muestra el caso simétrico en el cual la presencia de un metabolito RNC se propaga corriente arriba.

La detección de RNP y RNC puede realizarse mediante la inspección de las filas de la matriz estequiométrica (Kumar et al., 2007). En el caso en que todas las reacciones estén definidas como irreversibles, aquellas filas cuyos coeficientes distintos de cero sean todos del mismo signo corresponden a metabolitos *deadend*. Es decir que las filas con todos sus coeficientes no nulos del mismo signo (*i.e.* todos negativos o todos positivos), se corresponden con metabolitos RNC y RNP respectivamente, siempre y cuando las reacciones en las que participa sean irreversibles. En el caso de que el sistema incluya reacciones reversibles, debe tenerse en cuenta que la reacción puede tanto

consumir como producir el metabolito y en tal caso el signo del coeficiente no es informativo ya que no permite la detección de *deadends* (excepto en los casos de que el metabolito participe en una única reacción). De acuerdo con lo anterior, la presencia de metabolitos *deadend* producirá que determinadas reacciones se encuentren bloqueadas. En el caso más simple, cualquier reacción conectada a un metabolito *deadend* se encontrará bloqueada. Dado que el flujo de las reacciones bloqueadas se encuentra restringido a cero en el estado estacionario, estas no contribuyen al balance de ningún metabolito y por tanto las reacciones pueden ser tratadas como si no estuvieran presentes en el sistema. Luego, si se omiten las reacciones bloqueadas, quitando las respectivas columnas de la matriz estequiométrica y se vuelven a inspeccionar las filas de la matriz estequiométrica, es posible que nuevos metabolitos sean detectados como *deadends*. La detección de nuevos *deadend* traerá a su vez la identificación de nuevas reacciones bloqueadas. Este proceso iterativo permite ver como las inconsistencias en forma de metabolitos *deadend* se pueden propagar a través del modelo.

De la misma forma en que se identifican dos clases de metabolitos *deadend*, *i.e.* RNP y RNC, las inconsistencias causadas por estos se pueden propagar corriente abajo o corriente arriba, respectivamente (Kumar et al., 2007). En el caso de la [figura 3.7a](#), el metabolito A es un *deadend* de tipo RNP debido a la ausencia de reacciones que puedan producirlo. Como consecuencia, este efecto se propaga corriente abajo: i) la reacción v_1 se bloquea debido a que A es un RNP; ii) al considerar el bloqueo de v_1 el metabolito B es detectado como un nuevo *deadend*; y iii) si B causa el bloqueo de nuevas reacciones se repite el procedimiento. Así, los metabolitos que se detectan como *deadends* como consecuencia de que una inconsistencia se propague corriente abajo se denominan DNP (*Downstream-Non-Produced*, [figura 3.7a](#)). En forma simétrica, en el caso de un metabolito RNC la inconsistencia se propagará corriente arriba causando nuevos *deadends* denominados UNC (*Upstream-Non-Consumed*, [figura 3.7b](#)). En términos generales los metabolitos *gap* pueden definirse como aquellos nodos de la red los cuales no pueden ser atravesados por un flujo estacionario. Mientras que la identificación de metabolitos RNP y RNC puede realizarse de forma directa examinando las filas de la matriz estequiométrica, este procedimiento no puede aplicarse para la detección de los metabolitos UNC y DNP. Esto se debe a que la presencia de reacciones reversibles puede dificultar decidir si un metabolito es *deadend* terminal. Si bien quitando las reacciones que se detectan bloqueadas en forma iterativa es posible identificar nuevos metabolitos *gap*, existe un forma más directa para detectar el conjunto completo de metabolitos *gap* en un GEMM. Empleando la definición de reacción bloqueada previamente introducida ([definición 3.2](#)) o su extensión ([definición 3.3](#)) es posible dar una definición de metabolito *gap* que además permite su detección en forma directa (Ponce-de-Leon et al., 2013).

Definición 3.4: *dada una red metabólica con una matriz estequiométrica N , un metabolito $i \in I$ es un *gap* si y solo si, todas las reacciones en las que participa (ya sea como reactivo o como producto) se encuentran bloqueadas.*

La [definición 3.4](#) implica que si un metabolito es *gap*, ningún modo de flujo pasará a través de él. Por otra parte, si definimos el conjunto de reacciones en las que participa un metabolito i de la siguiente manera:

$$\sigma(i) = \{j \in J : N_{ij} \neq 0\} \quad \forall i \in I \quad (\text{Ec. 3.5})$$

Entonces, un metabolito *gap* queda definido mediante la siguiente expresión:

$$I_{Gap} = \{i \in I : \sigma(i) \subseteq J_{Blocked}\} \quad (\text{Ec. 3.6})$$

Entonces, dado el conjunto de reacciones bloqueadas $J_{Blocked}$ la detección de metabolitos *gap* se puede realizar en forma directa aplicando las ecuaciones 3.5 y 3.6 sobre cada metabolito. Además, el conjunto de reacciones bloqueadas puede incluir a las reacciones detectadas como termodinámicamente bloqueadas lo que a su vez puede permitir la identificación de metabolitos *gap* adicionales.

La definición de metabolito *gap* propuesta no hace distinción entre las diferentes clases de *gap* previamente definidas (*i.e.* RNP, RNC, UNP, DNC). Si bien no hay un procedimiento general que permita clasificar los metabolitos *gap* de acuerdo a las cuatro clases, en algunos casos es posible identificar casos que se correspondan a RNP y RNC. Como se menciona al inicio de la sección, los metabolitos *gap* RNP y RNC pueden detectarse mediante la inspección de las filas de la matriz estequiométrica y teniendo en cuenta la reversibilidad de las reacciones. Por ejemplo, si para el caso de un metabolito *gap*, en todas las reacciones en que participa lo hace como sustrato y todas estas reacciones son irreversibles el metabolito será RNP. Un razonamiento similar puede aplicarse de forma simétrica para la clasificación de metabolitos RNC que se corresponden a metabolitos que solo son producidos por las reacciones de la red. De esta forma una vez identificados todos los metabolitos *gap*, es posible hacer una clasificación de los mismo en base a su participación en reacciones, como reactivo y/o producto.

3.5 Pools conservados y el problema de los metabolitos *pseudo-gap*

Como se ha comentado, los metabolitos *gap* pueden ser definidos considerando la relación entre el conjunto de reacciones en las que cada metabolito participa y el conjunto reacciones bloqueadas $J_{Blocked}$ presentes en la red. Si un metabolito participa únicamente en reacciones bloqueadas es un *gap*. Sin embargo, dentro de la red pueden existir ciertos metabolitos que si bien participan en reacciones consistentes (*i.e.* no bloqueadas) y de acuerdo a la [definición 3.4](#) no son *gaps*, estos pueden causar el bloqueo de algunas reacciones de la red. Como consecuencia, la detección de tales metabolitos es clave para poder resolver determinadas inconsistencias. En el ejemplo ilustrado en la [figura 3.8](#), el par de metabolitos D y D* representan una coenzima y su forma activada o conjugada, respectivamente. En las situaciones en las que siempre que el metabolito D es producido por una reacción D* es también consumido y recíprocamente, el par de metabolitos definen un *pool* conservado. Como consecuencia ninguno de los dos metabolitos puede ser producido o consumido netamente por el sistema y esto implicará que la reacción v_4 se encuentre bloqueada. Al igual que con los *deadend*, la inconsistencia puede propagarse corriente arriba/abajo bloqueando otras reacciones (*e.g.* $v_1 - v_3$). Aquellos metabolitos que como D y D* participen en reacciones consistentes pero que por formar parte de un *pool* conservado son la causa del bloqueo de una o más reacciones han sido denominados metabolitos *pseudo-gap* (Ponce-de-Leon et al., 2013). La idea de *pseudo-gap* está directamente relacionada con las denominadas relaciones de conservación de un sistema; estas son subconjuntos o *pools* de metabolitos conectados a través de reacciones de forma tal que la masa total de las especies involucradas en el *pool* se conserva dentro del sistema (Heinrich and Schuster, 1996; Sauro and Ingalls, 2004). Por ejemplo, si el efecto de las

reacciones (lentas) de síntesis y degradación de la coenzima menaquinona son despreciadas, la forma oxidada y reducida de la vitamina cumplen la siguiente relación:



Es decir que la masa total de ambas formas de la coenzima no varía con el tiempo. Los *pools* conservados suelen por lo general corresponder a las distintas formas o estado de una coenzima o cofactor (*e.g.* NAD⁺/NADH o ATP/ADP/AMP) que son empleados por la célula para transferir distintas propiedades químicas (*e.g.* electrones, grupos químicos) dentro de la red. Como consecuencia, en general siempre que se consume una de las formas de la coenzima su conjugada es producida (excepto en los casos en que se considere su síntesis y/o degradación) y por lo tanto la masa total de ambas formas se mantiene constante en forma independiente a las tasa a la que son intercambiados.

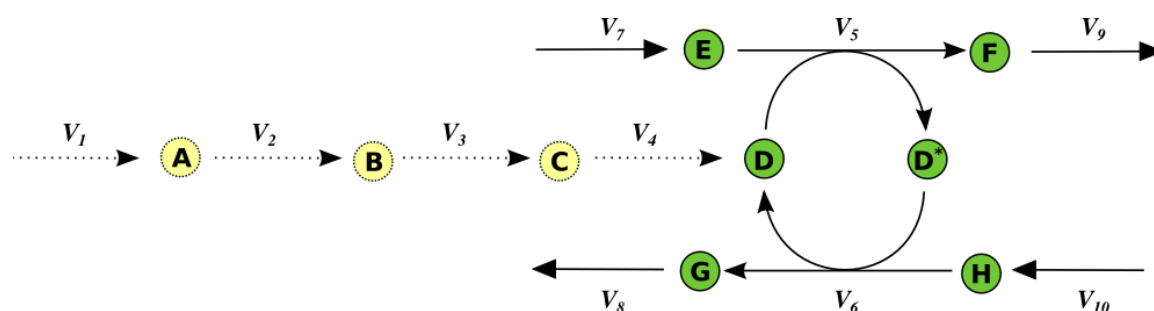


Figura 3.8. Metabolitos pseudo-gaps. En el sistema, la presencia de una relación de conservación entre los metabolitos D y D* tiene como consecuencia el bloqueo de cuatro reacciones ($v_1 - v_4$). Dado que D participa en dos reacciones activas (v_5 y v_6) este no es clasificado como gap. Sin embargo, por ser la causa directa del bloqueo de v_4 es denominado un pseudo-gap. Las flechas punteadas y continuas representan reacciones bloqueadas y no-bloqueadas, respectivamente mientras que círculos amarillos y verdes corresponden a metabolitos gaps y no-gaps, respectivamente.

Si bien cuando se modela una sección concreta del metabolismo (*e.g.* la glicólisis) las coenzimas y cofactores, así como otros metabolitos ubicuos (*e.g.* H₂O, CO₂) suelen considerarse externos al sistema (Schuster et al., 2000), los modelos reconstruidos a escala genómica suelen incluir las vías de síntesis de cofactores y coenzimas que se encuentren codificadas en el genoma del organismo, y por tanto se considera el balance de masas de estos metabolitos. De esta manera, cuando el modelo incluye la ruta de síntesis de un determinado cofactor o coenzima se debe también incluir flujo de degradación o drenado de dicho metabolito. En caso contrario, la coenzima se convertirá en un *pseudo-gap* y causará el bloqueo de las reacciones involucradas en su ruta de síntesis, debido a la restricción que surge de la relación de conservación entre la coenzima y su forma conjugada. En el diagrama de la [figura 3.8](#) las reacciones ($v_1 - v_4$) representan la ruta de síntesis de la coenzima D, mientras que las reacciones v_6 y v_5 representan procesos de consumo y regeneración de D, respectivamente. Desde un punto de vista algebraico, las relaciones de conservación se corresponden a subconjuntos de filas de la matriz estequiométrica que son linealmente dependientes (Heinrich and Schuster, 1996; Palsson, 2006). Las relaciones de conservación pueden identificarse computando una base del espacio nulo izquierdo de N, *i.e.* el espacio nulo de la transpuesta de N⁹. Sin embargo, la elección de una base del espacio nulo es arbitraria y en muchas

9 También denominado espacio de filas en contraposición al espacio nulo derecho de la matriz que está definido sobre las columnas.

situaciones, la interpretación de *pools* conservados en términos de vectores generadores puede no ser intuitiva (Heinrich and Schuster, 1998, 1996).

Las relaciones de conservación pueden enumerarse de forma exhaustiva calculando la base convexa el espacio nulo izquierdo N (Famili and Palsson, 2003; Nikolaev et al., 2005). Las distintas relaciones de conservación representadas en los elementos de la base, pueden ser clasificadas en tres clases, en forma análoga a la clasificación propuesta para los modos elementales y vías extremas (Schilling et al., 2000). Dicha clasificación está basada en un agrupamiento de metabolitos como primarios o secundarios: los primarios corresponden a los metabolitos cuya estructura sufre cambios a lo largo de las transformaciones en el esqueleto de carbono (e.g. la glucosa a lo largo de la glicólisis); los metabolitos secundarios se corresponden a las coenzimas o cofactores (e.g. ATP, NAD, etc). En base a esta agrupación de metabolitos, los tipos de *pools* corresponden a: Tipo A son los *pools* de compuestos únicamente por metabolitos primarios; Tipo B son aquellos *pools* que contienen tanto metabolitos primarios como secundarios; y Tipo C son los *pools* compuestos únicamente por metabolitos secundarios (Famili and Palsson, 2003). De acuerdo con esta clasificación, en las situaciones en que un modelo contenga *pseudo-gaps*, estos deberían formar parte de *pools* conservados de tipo C. Sin embargo, es posible que otros metabolitos que no sean *pseudo-gaps* también se encuentren presentes en relaciones de conservación de tipo C.

La detección de metabolitos *pseudo-gap* no puede realizarse empleando las ecuaciones 3.5 y 3.6, debido a que estos participan en reacciones activas. No obstante, una posible aproximación para realizar la detección de *pseudo-gap* es la identificación de *pools* conservados que surgen del análisis de las relaciones de conservación presentes en el sistema (Ponce-de-Leon et al., 2013). Sin embargo, es posible identificar los metabolitos *pseudo-gap* de una forma más directa mediante el uso de variables auxiliares y programación lineal. El procedimiento propuesto consiste en clasificar cada metabolito i de acuerdo a si el sistema es capaz de producir y/o consumir el metabolito en forma neta en los siguiente cuatro tipos: 1) consumir el metabolito i pero no producirlo; 2) producir el metabolito i pero no consumirlo; 3) consumir y producir el metabolito i ; 4) el sistema no puede producir ni consumir el metabolito i . La clasificación se realiza iterando sobre cada metabolito interno que no haya sido clasificado como *gap*. En cada iteración se selecciona un metabolito i al cual se le añade un flujo auxiliar b_i que permita el libre consumo del metabolito. El flujo auxiliar b_i se maximiza y si el valor óptimo es mayor que cero, el metabolito es producido por el sistema. En forma simétrica b_i es definido de forma tal que produzca el metabolito i y nuevamente se maximiza el valor a través de b_i si el valor óptimo es mayor a cero el sistema puede consumir al metabolito i . De esta forma, aquellos metabolitos que no puedan ser ni consumidos ni producidos por el sistema serán, o bien metabolitos que incluyen algún motivo conservado (ver el ejemplo del ACP en la apartado 4.3.1) o *pseudo-gaps*. En base a la definición de metabolito *gap* (definición 3.4) y al procedimiento planteado se plantea la siguiente definición:

Definición 3.5: dada una red metabólica con una matriz estequiométrica N , un metabolito $i \in I$ es un *pseudo-gap* si y solo si, i no es un *gap* de acuerdo con la definición 3.4 y el sistema no es capaz de producir ni consumir en forma neta al metabolito i .

Una forma de resolver esta clase de inconsistencias en modelos a escala genómica, consiste en incluir alguna de las formas del cofactor en la ecuación de biomasa o incluir un flujo de

intercambio que pueda drenar el metabolito fuera del sistema. Desde un punto de vista biológico esto se justifica en el hecho que: i) los cofactores o coenzimas, por estables que sean químicamente, tenderán a degradarse; y ii) cuando un organismo se encuentra en fase de crecimiento las concentraciones de los metabolitos comenzarán a diluirse como consecuencia del aumento en el volumen celular. En ambas situaciones, la regeneración del cofactor implicará: o bien síntesis *de-novo*, o que sea importado del ambiente. Dado que las formas alternativas del cofactor o coenzima conforman un *pool*, el flujo de consumo o degradación puede añadirse a cualquiera de las formas. Al igual que con las inconsistencias termodinámicas, la idea de *pseudo-gap* extiende el concepto de metabolito *gap* previamente definido. Como se verá en la próxima sección el considerar los *pseudo-gaps* puede facilitar en buena medida la identificación de la causa última de muchas inconsistencias que pueden presentar los modelos metabólicos.

3.6 El grafo de inconsistencias y los módulos inconexos

Al analizar los conjuntos de reacciones bloqueadas y metabolitos *gap* detectados en un modelo metabólico, es claro de que ambos conjuntos se encuentran estrechamente relacionados, *e.g.* algunas reacciones bloqueadas pueden encontrarse directamente conectadas a un metabolito *deadend* RNP/RNC. Sin embargo, pueden existir reacciones bloqueadas que no estén directamente conectadas a un metabolito de tipo RNP/RNC. Esto se debe a que las reacciones bloqueadas pueden encontrarse conectadas entre sí a través de metabolitos *gap* de tipo DNP/UNC, como es el caso de la reacciones v_2 que si bien no está directamente conectada con al metabolito A, que es un *deadend* de tipo RNP, se encuentra indirectamente conectada a través de un metabolito B de tipo DNC y la reacción también bloqueada v_1 (ver figura 3.9a). Algo similar ocurre cuando la causa de la inconsistencia es producida por un metabolito RNC (*e.g.* el caso de la reacción v_8 en el ejemplo de la figura 3.9b).

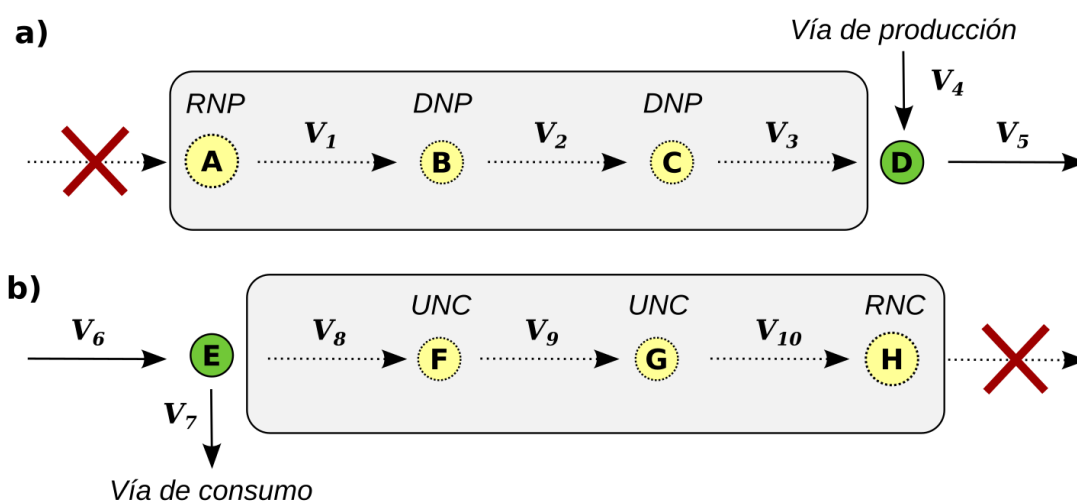


Figura 3.9. Representación de metabolitos *gap* y reacciones bloqueadas agrupadas en módulos inconexos. El esquema metabólico es el mismo que se presenta en la figura 3.7, pero en este caso las reacciones bloqueadas y metabolitos *gap* han sido agrupados en módulos inconexos.

Dado que es posible identificar en forma sistemática la relación entre reacciones bloqueadas y metabolitos *gap* se podrían detectar aquellas subredes o módulos del metabolismo que han sido reconstruidos en forma inconsistente. Una aproximación para buscar las relaciones entre reacciones bloqueadas y metabolitos *gap* sería detectar *clusters* o agrupamientos de reacciones bloqueadas y metabolitos *gap*, en la estructura de la red, ya sea en forma manual o empleando algún algoritmo de agrupamiento. Para esto, primero es necesario proyectar la estructura de la red metabólica en el grafo bipartito correspondiente, $G=(V, E)$ donde el conjunto de vértices V es la unión del conjunto de reacciones y el conjunto de metabolitos, *i.e.* $V=I \cup J$, y donde las aristas entre pares reacción-metabolito se establecen siempre que un cierto metabolito participe en una reacción, *i.e.* $E=\{(i, j) \forall i \in I, j \in J: N_{ij} \neq 0\}$. Por ejemplo, en la [figura 3.10](#) se representa el grafo bipartito correspondiente al modelo metabólico *iCG238* de la bacteria *Blattabacterium* Bge, endosimbionte de la cucaracha *Blattella germanica* (González-Domenech et al., 2012). En el esquema las 69 reacciones identificadas como bloqueadas y los 58 metabolitos detectados como *gaps* están coloreados de rojo y amarillo, respectivamente.

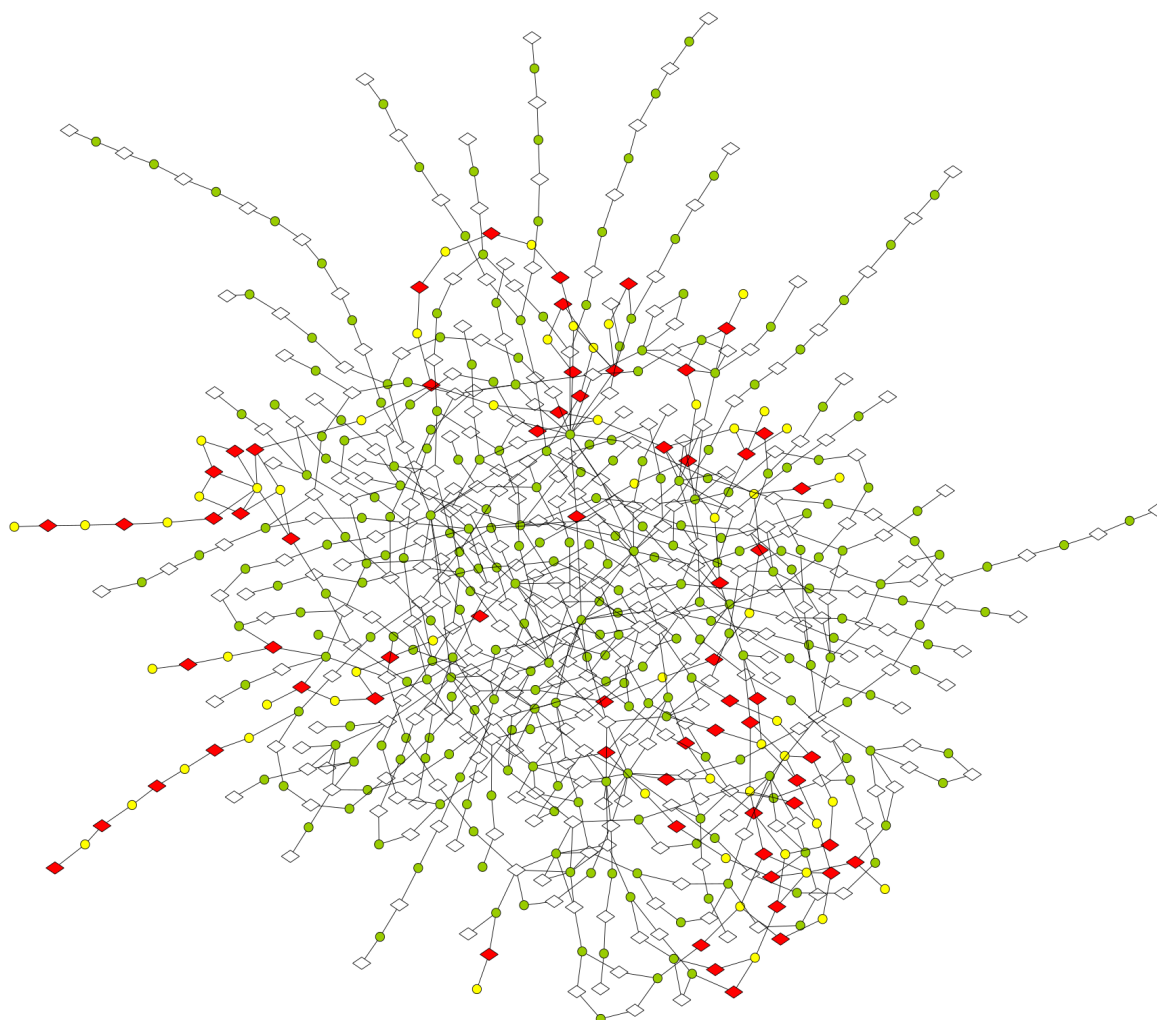


Figura 3.10. Grafo metabólico del modelo *Blattabacterium* Bge *iCG238*. Representación del GEMM de *Blattabacterium* Bge *iCG238* mediante un grafo bipartito. Los nodos en forma de diamantes rojos y blancos representan a las reacciones bloqueadas y no bloqueadas respectivamente. Los nodos en forma de círculos representan metabolitos (verdes) y metabolitos *gap* (amarillos). Para facilitar la visualización los siguiente metabolitos fueron eliminados: H^+ , H_2O , CO_2 , ATP/ADP , P_i , $NAD^+/NADH$.

La inspección visual de la red representada en la [figura 3.10](#) permite detectar, por ejemplo algunas cadenas de reacciones bloqueadas conectadas a través de metabolitos *gap*. Sin embargo, dada la complejidad que presenta la estructura de una red metabólica, incluso tratándose de un metabolismo reducido, la identificación en forma individual y automática de aquellos módulos o secciones de la red que son inconsistentes no resulta *a priori* trivial. De forma que no es clara que metabolitos son la causa última de las inconsistencias. La complejidad de la representación se debe en buena medida a la presencia de metabolitos que, como el ATP, los protones o el agua, son muy ubicuos, y en consecuencia interconectan vías metabólicas diferentes. La gran conectividad de estos metabolitos es lo que en última instancia dificulta la visualización de la estructura de la red. Una aproximación bastante común para lidiar con este problema es eliminar de la red los metabolitos muy conexos (Jeong et al., 2000; Ma and Zeng, 2003; Wagner and Fell, 2001). Sin embargo, esta estrategia no resulta muy apropiada ya que en ciertos casos las inconsistencias en la red pueden estar relacionadas con dichos metabolitos, como se ha reportado en la reconstrucción del modelo metabólico de *Portiera aleyrodidarum* endosimbionte de la mosca blanca (Calle-Espinosa et al., 2016). Además, como se verá a continuación, es posible detectar las causas y relaciones de entre reacciones bloqueadas y metabolitos *gap*, en forma directa. Si bien al realizar la proyección de la red en un grafo bipartito hay una pérdida de información, la relación de conexión entre reacciones y metabolitos se conserva¹⁰. Como consecuencia de lo anterior tiene que existir una forma directa de establecer la relación entre metabolitos *gap* y reacciones bloqueadas.

La forma más natural es considerar la subred o subgrafo que incluya únicamente los elementos inconsistentes, *i.e.* las reacciones bloqueadas y los metabolitos *gap*. En general, considerar cualquier subgrafo equivale a definir un par de subconjuntos $\hat{\mathbf{I}} \subseteq \mathbf{I}$ y $\hat{\mathbf{J}} \subseteq \mathbf{J}$ de metabolitos y reacciones, respectivamente. De esta forma solo se preservan las aristas existentes entre los metabolitos y las reacciones de los subconjuntos $\hat{\mathbf{I}}$ y $\hat{\mathbf{J}}$ definidos. En particular, es posible considerar el subgrafo donde el subconjunto de vértices se corresponda con los elementos inconsistentes de la red y de esta forma definimos el grafo de inconsistencias.

Definición 3.6: *dado un modelo metabólico inconsistente con un conjunto de metabolitos \mathbf{I}_{gap} y reacciones bloqueadas $\mathbf{J}_{Blocked}$, el grafo de inconsistencias se define como un subgrafo del grafo bipartito que induce la red, donde el conjunto de nodos se define mediante la unión del conjunto de reacciones bloqueadas y metabolitos \mathbf{gap} : $V = \mathbf{I}_{gap} \cup \mathbf{J}_{Blocked}$.*

Si bien en su primera formulación, el grafo de inconsistencias solo incluía metabolitos *gap* y reacciones bloqueadas estequiométricamente, es importante destacar que el grafo de inconsistencias se puede extender incluyendo metabolitos identificados como *pseudo-gaps* y reacciones *t-bloqueadas*. De esta forma el grafo contiene la estructura de todas las inconsistencias detectadas en el modelo. Una vez que el grafo de inconsistencias ha sido construido es posible descomponer el mismo en componentes conexas como se muestra en el ejemplo de la [figura 3.11](#). La inspección del grafo de consistencias de la [figura 3.11](#) permite ver cómo las reacciones bloqueadas y los metabolitos *gap* suelen encontrarse agrupados en módulos o componentes independientes unos de otros. En este contexto cada componente conexo del grafo de

¹⁰ Siempre que se realiza una proyección de una red metabólica en un grafo se pierde la información de la estequiometría y las restricciones impuestas por esta. Sin embargo, en muchas situaciones es posible que al realizar la proyección, ciertas propiedades estructurales de interés para responder a una pregunta, se conserven.

inconsistencias se puede interpretar como un módulo del metabolismo que en el modelo se encuentra desconectado del resto de la red como consecuencia de la presencia de inconsistencias en red (e.g. metabolitos *gap* RNP/RNC y/o metabolitos *pseudo-gaps*). Por dichos motivos los componentes conexos del grafo de inconsistencias son referidos como módulos inconexos (UM, *Unconnected Module*) (Ponce-de-Leon et al., 2013).

Definición 3.7: dado un modelo metabólico inconsistente, un **módulo inconexo** se define como un componente conexo del grafo de inconsistencias (*definición 3.6*).

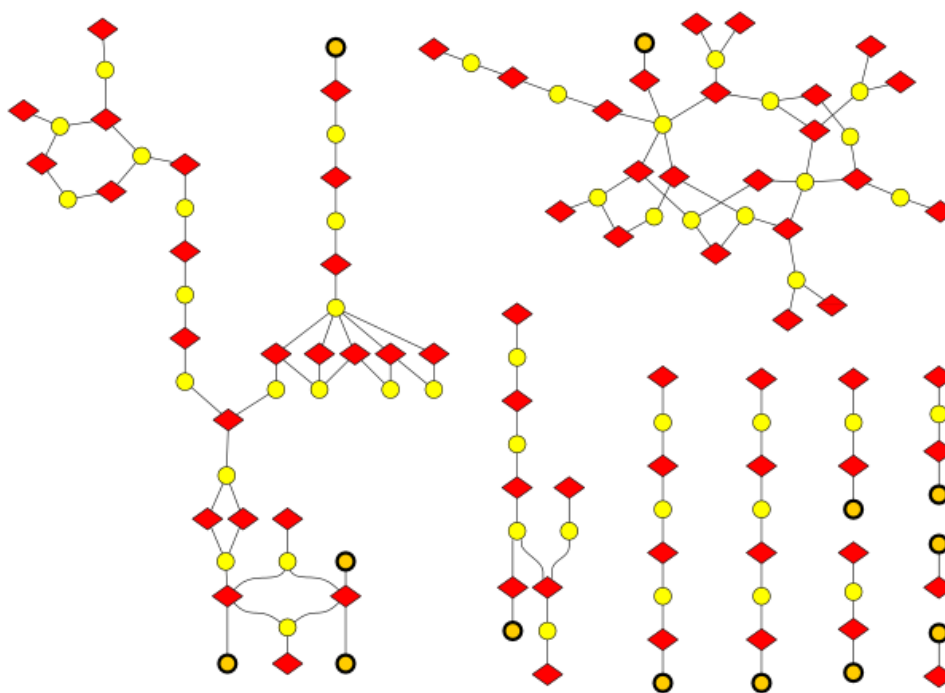


Figura 3.11. Grafo de inconsistencias del modelo *Blattabacterium Bge iCG238*. La figura representa el grafo de inconsistencias de *iCG238*. Cada uno de los diez subgrafo aislado (o componentes) se corresponde con un módulo inconexo. Los nodos en forma de diamante rojos representan las reacciones bloqueadas y los círculos amarillos representan los metabolitos *gap*; los círculos anaranjados con el contorno más marcado indican los metabolitos *gap* clasificados como RNP o RNC.

La resolución manual de cada UM implica modificar la formulación del modelo, ya sea agregando nuevos elementos, como quitando aquellos que representan artefactos. De esta manera, el conjunto de operaciones elementales que se pueden aplicar para resolver inconsistencias en modelos metabólicos han sido discutidas en distintas publicaciones, y son las mismas que se pueden aplicar para resolver UMs individuales (Kumar et al., 2007). Estas operaciones incluye a: 1) incorporación al modelo de reacciones y/o transportadores no presentes; 2) incorporación de flujos de intercambio que operen a modo de fuente o sumidero para el caso de aquellos metabolitos *deadend* cuyo origen o destino bioquímico sea desconocido; 3) incorporar en la ecuación de biomasa aquellos metabolitos RNC que sean productos finales del alguna ruta metabólica de biosíntesis y que se consideren esenciales para el crecimiento del organismo; 4) modificación del sentido en que se encuentran definidas ciertas reacciones irreversibles; 5) eliminación de ciertas reacciones y metabolitos del modelo en caso de que sean un artefacto. La estrategia de resolución

de inconsistencias mediante la identificación y análisis de UMs puede además ser extendida o combinada con métodos de *gap-filling*. Esto puede permitir el realizar un *gap-filling* de forma iterativa, *i.e.* calculando de forma automática las modificaciones necesarias para resolver un cierto UM. Con esta información, un experto puede decidir *a posteriori* cuales son los UM que tienen un correlato biológico y deben ser resueltos o completados, y cuales representan alguna clase de artefacto y deben ser eliminados del modelo. Por último, una de las principales ventajas que este análisis presenta frente a otras estrategias de resolución de inconsistencias es que es un cálculo que no requiere el ajuste de parámetros y el costo computacional es virtualmente nulo.

3.7 Metacoupling y su aplicación al *gap-filling*

Si se considera una red global o metamodelo del metabolismo, que incluya todas las reacciones bioquímicas y procesos de transporte conocidos y que pueden tener lugar en algún organismo, entonces es posible concebir el metabolismo de cualquier organismo como un subconjunto de esta red global. En este sentido, analizar las propiedades estructurales de un metamodelo puede dar información sobre la organización de las rutas metabólicas a una escala global. En el contexto de la reconstrucción metabólica, hemos encontrado que las relaciones de acoplamiento de un metamodelo pueden ser empleadas para identificar de forma directa actividades enzimáticas que permitan completar ciertos *gaps* presentes en la reconstrucción metabólica de un organismo. En este apartado se plantea y discute la posibilidad de emplear las relaciones de acoplamiento de un metamodelo para asistir el proceso de reconstrucción a escala genómica de modelos metabólicos, así como a la resolución de inconsistencias. Concretamente, se mostrará como las relaciones de acoplamiento estequiométricas inferidas a partir del análisis de acoplamiento de flujos (ver [apartado A.2.3](#) en el apéndice A) pueden ser empleadas para detectar conjuntos de reacciones que dependen de otras reacciones para poder funcionar, y que por tanto si una está presente en el organismo, es altamente probable que también lo esté la otra. Por ejemplo, en el caso de una ruta metabólica lineal todas las actividades involucradas conforman un conjunto de acoplamiento completo (FCS, *full coupling set*), es decir que la ausencia de una única reacción (cualquier de ellas) es suficiente para que la ruta no pueda sostener un flujo estacionario distinto de cero. Entonces, si al reconstruir dicha ruta en organismo particular, se tiene conocimiento que esta conforma un FCS, la resolución de *gaps* en dicha ruta sería, en principio, trivial, ya que bastaría con identificar que reacciones del FCS no están presentes en el modelo para deducir las reacciones que completarán la ruta.

Si consideramos que el modelo de un organismo particular puede ser tratado como una subred del metamodelo, entonces las relaciones de acoplamiento del metamodelo deberán estar presentes en los modelos individuales. Esta consideraciones surgen como consecuencia de la propiedades matemáticas de las relaciones de acoplamiento, previamente estudiadas por Marashi y Bockmayr (Marashi et al., 2012; Marashi and Bockmayr, 2011). Los autores han analizado cómo son afectadas las relaciones de acoplamiento entre las reacciones de una red, cuando se añaden nuevas reacciones al sistema. Concretamente, probaron que dado un modelo consistente la relación de acoplamiento entre cualquier par de reacciones de la red o bien se conservará o bien se relajará

cuando la red se extiende mediante el añadido de nuevas reacciones¹¹. En concreto, dada una red N consistente, uno de los teoremas demostrado por los autores plantean todos los escenarios de posible cambio en la relación de acoplamiento de cualquier par de reacciones i y j , cuando N es expandida añadiendo nuevas reacciones. A continuación se plantea el enunciado de dicho teorema¹².

Teorema. sea N una red y N' la red extendida que incluye a N , entonces todos los casos de cambios en las relaciones de acoplamiento que pueden ocurrir entre dos pares de reacciones no bloqueadas son los siguientes¹³:

Caso 1: si $i \Leftrightarrow j$ en N entonces en N' puede ocurrir un de las siguientes posibilidades:

- $i \Leftrightarrow j$
- $i \leftrightarrow j$
- $i \rightarrow j$
- $j \rightarrow i$
- $i \neq j$

Caso 2: si $i \leftrightarrow j$ en N entonces en N' puede ocurrir un de las siguientes posibilidades:

- $i \leftrightarrow j$
- $i \rightarrow j$
- $j \rightarrow i$
- $i \neq j$

Caso 3: si $i \rightarrow j$ en N entonces en N' puede ocurrir un de las siguientes posibilidades:

- $i \rightarrow j$
- $i \neq j$

Si bien los autores emplean el resultado para advertir que las relaciones de acoplamiento calculadas sobre un modelo pueden cambiar si el modelo se encuentra incompleto debido, por ejemplo, a errores en la anotación funcional, los mismos resultados sirven para probar que las relaciones de acoplamiento de un modelo consistente no se verán modificados o se volverán más estrictas en cualquier subred, *i.e.* en cualquier modelo derivado del metamodelo. El único caso donde el acoplamiento entre dos reacciones puede volverse más estricto ocurre cuando una reacción i se encuentra bloqueada en N , y se desbloquea en N' como consecuencia de añadir una nueva reacción j . Este escenario está por fuera de los supuestos del teorema, pero puede ocurrir como se verá en el capítulo sobre construcción y análisis de metamodelos (ver [capítulo 6](#)). En base a los argumentos expuestos, se deduce que las relaciones de acoplamiento elucidadas a nivel de un metamodelo, o en general, de cualquier base de datos metabólica pueden ser empleadas para asistir el proceso de reconstrucción metabólica. Los principales resultados derivados por los autores establecen que dada una red consistente, la relación de acoplamiento entre cualquier par de reacciones tenderá a relajarse (o permanecer igual) como consecuencia de la expansión de la red,

11 Por ejemplo, como consecuencia de añadir nuevas reacciones, puede ocurrir que una relación de acoplamiento completo se convierta en un acoplamiento direccional, o que las reacciones se desacoplen.

12 La demostración del teorema encontrarse en la publicación original (Marashi et al., 2012; Marashi and Bockmayr, 2011).

13 El símbolos indican: \Leftrightarrow acoplamiento completo; \leftrightarrow acoplamiento parcial; \rightarrow acoplamiento direccional; \neq reacciones desacopladas. Para más detalles sobre el FCA consultar el [apartado A.2.3](#) en el apéndice A.

mediante la adición de nuevas reacciones. Consecuentemente, el acoplamiento entre pares de reacciones consistentes (*i.e.* no bloqueadas) no podrá volverse más estricto (*e.g.* pasar de un acoplamiento direccional a un acoplamiento completo) como consecuencia de añadir nuevas reacciones a la red. Dicho de otra manera, si un subconjunto de reacciones conforman un FCS en el metamodelo, cualquier modelo que tengan alguna/as reacción/es pertenecientes a dicho FCS, deberá tener el subconjunto completo, o de lo contrario las reacciones estarán bloqueadas. El único caso posible en que el acoplamiento entre reacciones puede pasar a ser más estricto se produciría cuando una reacción *i* se encuentra bloqueada en una red, y esta se desbloquea como consecuencia de añadir una nueva reacción *j*. En dicho caso la reacción *i* pasaría a estar direccional o completamente acoplada a *j*. Esta situación particular fue considerada durante la construcción del metamodelo al analizar los casos en que ciertas reacciones de desbloquearon de forma espontánea al construir el metamodelo, como se verá más adelante (ver apartado 6.3.2). Por otra parte, cuando una o más reacciones son eliminadas de una red consistente pueden darse tres situaciones con respecto a la relación de acoplamiento entre cualquier par de reacciones *i* y *j*:

1. El acoplamiento entre *i* y *j* no se ve afectado.
2. El acoplamiento entre *i* y *j* se vuelve más estricto.
3. Alguna de las dos reacciones *i* y *j* (o ambas) se bloquean.

Al eliminar reacciones de una red se plantean tres escenarios posibles respecto a las relaciones de acoplamiento de la subred: 1) que la reacción eliminada fuese requerida por otra/s reacciones en cuyo caso dicho conjunto de reacciones se “bloqueará”; 2) que se modifiquen algunas relaciones de acoplamiento volviéndose “más rígidas”; y 3) que las relaciones de acoplamiento no se vean afectadas. Como corolarios se desprende que al eliminar reacciones de una red, las relaciones de acoplamiento no se relajarán en ningún caso. La propuesta es que a partir de las relaciones de acoplamiento de una red global es posible detectar reacciones ausentes en una subred que traen como consecuencia el bloqueo de algunas reacciones del modelo. De acuerdo de lo anterior estas reacciones ausentes pueden ser candidatas para completar los *gaps* de la red. Estos resultados pueden emplearse durante el proceso de reconstrucción y refinamiento de un GEMM. Concretamente, dado un metamodelo consistente con matriz estequiométrica *N* si una cierta reacción *i* se encuentra completa o direccionalmente acoplada a otra reacción *j*, entonces cualquier subred *N'*, correspondiente al modelo de un organismo particular, que incluya la reacción *i* debería, en principio, incluir también la reacción *j*, o de otra forma la reacción *i* se encontrará bloqueada en *N'*. De esta forma, si se conocen las relaciones de acoplamiento presentes en un metamodelo, estas pueden ser empleadas para refinar un modelo de la siguiente forma: después de la primera etapa de la reconstrucción metabólica en la que a partir de la anotación del genoma se construyen las tablas de asociación gen-proteína-reacción. El procedimiento consiste en emplear las relaciones de acoplamiento a nivel del metamodelo para detectar de forma directa las reacciones que debería incluir el modelo. Dado que esta aproximación se basa en la hipótesis de que el metamodelo incluye tanto la bioquímica conocida como la inferida, cualquier descubrimiento de una nueva actividad enzimática modificará la estructura del metamodelo y, por lo tanto, podría afectar las relaciones de acoplamiento previamente establecidas. De esta forma, cada extensión del metamodelo (*e.g.* mediante el añadido de nuevas reacciones, procesos de transporte, o metabolitos) implicaría la actualización de las relaciones de acoplamiento entre las reacciones de la red.

Capítulo 4

4 Revisión y análisis de un modelo metabólico de *Blattabacterium cuenoti*.

4.1 Introducción

Muchos insectos han establecido relaciones de simbiosis con bacterias intracelulares que se han prolongado durante extensos períodos de tiempo (Dale and Moran, 2006; Moya et al., 2008). Las bacterias endosimbiontes que coexisten con insectos suelen tener un rol importante en la expansión de las capacidades o funciones metabólicas del hospedador (Macdonald et al., 2012; MacDonald et al., 2011; Russell et al., 2014). En ciertos casos como los pulgones, las moscas tsetse, o las moscas blancas, los endosimbiontes complementan las capacidades metabólicas de sus respectivos hospedadores, los cuales ocupan nichos ecológicos con dietas deficientes en diversos nutrientes esenciales, como amino ácidos y/o vitaminas. Por su parte, los endosimbiontes que habitan organismos omnívoros como las hormigas carpinteras o las cucarachas, también pueden desempeñar una función importante en el reciclado del nitrógeno (López-Sánchez et al., 2009; Patiño-Navarrete et al., 2014; Sabree et al., 2009; Thomas et al., 2009). En general, durante el establecimiento de una simbiosis, aquellas vías metabólicas que sintetizan compuestos que son esenciales tanto para la bacteria como para el huésped serán mantenidas a pesar del proceso de reducción genómica al que el endosimbionte se encuentra sometido. Sin embargo, muchas de las capacidades metabólicas que son redundantes con el hospedador suelen conducir a la pérdida de actividades enzimáticas, por lo general, en la red del endosimbionte. Esta reducción de la estructura y las capacidades de la red suele conducir a la emergencia de complementaciones metabólicas obligadas entre ambos miembros de la asociación y, que en última instancia dan origen a la simbiosis. En este contexto, cabe mencionar que el caso de las cucarachas y su endosimbionte *Blattabacterium* Bge es quizás uno de los casos más antiguos de simbiosis mutualistas descritos hasta la fecha y en consecuencia su metabolismo se encuentra muy reducido (López-Sánchez et al., 2009; Moran and Bennett, 2014).

Una de las principales consecuencias que la reducción del metabolismo tiene para el endosimbionte es el pasaje de una forma de vida intracelular facultativa a una obligatoria. La vida intracelular obligada, consecuencia de la complementación metabólica entre el endosimbionte y su

hospedador plantea dificultades experimentales a la hora de aislar y caracterizar estos organismos ya que su cultivo en el laboratorio no es abordable. Como consecuencia, la disponibilidad de modelos metabólicos reconstruidos precisos es de especial importancia para el estudio funcional de las bacterias endosimbiontes obligadas (endosimbiontes primarios) dada la dificultad intrínseca a la hora de obtener datos experimentales directos en dichos sistemas. En forma similar, el estudio de patógenos intracelulares obligados, como el caso de los *Mycoplasmas*, plantea dificultades similar y, en efecto, el uso de modelos metabólicos ha probado ser de gran utilidad para, por ejemplo, predecir medios mínimos de cultivo para *Mycoplasma pneumoniae* y *Mycoplasma genitalium* (Suthers et al., 2009; Yus et al., 2009).

El uso de modelos metabólicos ha sido aplicado con éxito en muy diversos sistemas (Feist and Palsson, 2008; McCloskey et al., 2013). En el contexto del estudio de bacterias endosimbiontes, el estudio de *Buchnera aphidicola* BAp (endosimbionte obligado de los pulgones) para evaluar el papel del endosimbionte y su hospedador en la síntesis de aminoácidos esenciales (Macdonald et al., 2012; Russell et al., 2014) así como el grado de fragilidad la red (González-Domenech et al., 2012). Así mismo, el GEMM de *B. aphidicola* BAp ha permitido modelar el proceso evolutivo de reducción genómica que tuvo lugar en dicho organismo (Li et al., 2010; Pál et al., 2006). También se han desarrollado modelos metabólicos para caracterizar los estadios intermedios de la reducción de la red metabólica de *Sodalis glossinidius*, endosimbionte facultativo de la mosca tse-tse (Belda et al., 2012), así como la consecuencias en las capacidades energéticas de la pequeña red metabólica de *Ca. Portiera aleyrodidarum*, endosimbionte de las moscas blancas (Calle-Espinosa et al., 2016). En el caso de *Blattabacterium cuenoti*, la reconstrucción y análisis del modelo metabólico iCG238 ha permitido una mejor comprensión de la sorprendente conservación que la estructura metabólica del endosimbionte ha mantenido durante un extenso período de tiempo, así como el que el papel que dicho metabolismo juega en la gestión del nitrógeno (Patiño-Navarrete et al., 2013). Si bien los métodos automáticos para la revisión de modelos metabólicos reconstruidos a escala genómica son sin lugar a dudas una herramienta fundamental para mejorar la formulación de un modelo, existen muchas situaciones donde la inspección manual por parte de un experto sigue siendo imprescindible. En efecto, este es el caso de la reconstrucción de redes metabólicas de organismos que han estado sometidos a un largo proceso evolutivo de reducción genómica y que codifican metabolismos muy reducidos (e.g. bacterias endosimbiontes objeto de nuestro estudio, así como como diversos parásitos intracelulares). En general, las capacidades metabólicas compartidas entre huésped y hospedador suelen reflejarse en el endosimbionte en forma de vías metabólicas incompletas, i.e. interrumpidas en algún punto, o mediante *gaps* que son detectados cuando se reconstruye el metabolismo del endosimbionte. A su vez, los *gaps* darán lugar a inconsistencias en la formulación del modelo metabólico (Ponce-de-Leon et al., 2013). En la mayoría de los casos, tales *gaps* no son producto de errores en la anotación del genoma de la bacteria, sino que son el resultado de la complementación metabólica establecida con otro organismo (e.g. el huésped u otro endosimbionte). Como consecuencia, reconstruir y revisar modelos metabólicos con tales características es una tarea compleja, donde la representación visual de las inconsistencias puede ser de ayuda para que un experto puede elucidar la naturaleza de las inconsistencias y plantear hipótesis que conlleven a la resolución de las mismas.

4.2 Materiales y Métodos

4.2.1 Modelo metabólico y anotación funcional

El modelo *iCG238* fue obtenido de la base datos Biomodels (Li et al., 2010) y su identificador correspondiente es [MODEL1110130000](#). El modelo contiene 364 metabolitos, de los cuales 317 son citosólicos y 47 extracelulares; 419 reacciones, de las cuales 372 son reacciones intracelulares y las 47 restantes corresponden a flujos de intercambio, asociados a cada uno de los metabolitos extracelulares. El modelo incluye 238 genes que codifican para actividades enzimáticas. Sin embargo no todas las reacciones catalizadas por enzimas tienen asociado un gen/es, sino que existen 43 reacciones y 33 transportadores huérfanos. Durante el proceso de revisión se empleó como referencia la anotación genómica disponible (RefSeq: GCF_000022605.2) (López-Sánchez et al., 2009) y la base de datos modelo-organismo *Blattabacterium* Bge, versión 17.5 ([BSP331104](#)) disponible en el Tier 3 del BioCyc (Caspi et al., 2014). El resumen de las características del genoma se encuentra en la [tabla 4.1](#).

Tabla 4.1 Descripción de la anotación genómica de *Blattabacterium* Bge.

Replicón	Nº Genes	Nº Proteínas	Nº RNAs	Pseudogenes	Longitud (bp)
NC_013454 (Cr.)	626	587	39	1	636.850
NC_015679 (Pl.)	4	4	0	0	4.085
Total:	629	591	39	1	640.935

4.3 Resultados

4.3.1 Análisis de consistencia del modelo *Blattabacterium* Bge *iCG238*

El primer paso para la revisión del modelo de *Blattabacterium* Bge *iCG238* fue un análisis de consistencia para detectar: 1) el conjunto de reacciones estequiométrica y termodinámicamente bloqueadas; 2) el conjunto de metabolitos *gap* y *pseudo-gap*; y 3) el conjunto de módulos inconexos que se definen a partir de las reacciones bloqueadas y los metabolitos *gap*. Los resultados del análisis indican que 68 reacciones de un total de 419 (~16%) se encuentran estequiométricamente bloqueadas para las condiciones de crecimiento más relajadas, es decir aquellas donde todos los metabolitos extracelulares pueden ser consumidos por el sistema. Por otra parte, únicamente se detectaron dos reacciones *t-bloqueadas*. Empleando estos resultados, se identificó el conjunto de metabolitos *gap* compuesto por 60 metabolitos de un total de 364 (~16%). El conjunto de metabolitos *pseudo-gap*, *i.e.* los metabolitos que si bien participan en reacciones activas no pueden ser producidos ni consumidos en forma neta por el sistema, fueron identificados empleando la aproximación propuesta en la [sección 3.5](#), e incluye un total de 17 compuestos.

El conjunto de metabolitos *pseudo-gap* se encuentra listados en la [tabla 4.2](#), donde han sido agrupados con respecto al *pool* conservado que estos definen. Por ejemplo, el menaquinol y la menaquinona son ambos *pseudo-gaps* y conforman un *pool* conservado. Es importante destacar que 9 de los 17 metabolitos *pseudo-gap* detectados son intermediarios en la ruta de síntesis de lípidos de membrana, y como consecuencia se encuentran unidos a la proteína transportadora de acilos (ACP, *acyl carrier protein*). Dado que el ACP no es en sí un metabolito secundario sino una

proteína, ni su síntesis ni su degradación se encuentran incluidas en el modelo, ya que este no incluye tales procesos. Como consecuencia cualquier compuesto que tenga un ACP unido no puede ser consumido/producido en forma neta por el sistema ya que de ser así, se violaría la conservación del motivo ACP. Sin embargo, esto no ocurre ni con el producto inicial ni con el producto final de la ruta de síntesis de lípidos de membrana, debido a que ninguno de los compuestos se encuentra ligados al ACP. El ACP como tal es un motivo conservado que no ni producido ni consumido netamente por el sistema y esto ocurrirá a cualquier compuesto que tenga un ACP unido. Esto demuestra que es el motivo ACP la causa de que los 9 metabolitos intermediarios de la ruta, sean identificados como *pseudo-gaps*. Por esta razón, el *pool* de metabolitos que queda definido a causa del ACP fue excluido del análisis ya que no representan una inconsistencia del modelo, sino que son una consecuencia de que los procesos de síntesis y degradación del ACP están fuera del alcance del modelo.

Tabla 4.2 Metabolitos pseudo-gaps agrupados pools conservados.

Pools conservados

mql8, mqn8

trdox, trdrd

uaagmda, uagmda

udcpdp, udcpp

tdeACP, octeACP, ACP, malACP, myrsACP, acACP, actACP, palmACP, hdeACP

La abreviación de los metabolitos se encuentra en la [tabla suplementaria S.4.1](#).

Las inconsistencias detectadas fueron empleadas para determinar el conjunto de módulos inconexos a través de la construcción del grafo de inconsistencias. En el grafo de inconsistencias se incluyeron además las 2 reacciones *t-bloqueadas* y los 8 metabolitos *pseudo-gaps*, no relacionados con el ACP, en forma adicional a las 68 reacciones bloqueadas y a los 58 metabolitos *gap* presentes en la publicación original (Ponce-de-Leon et al., 2013). A partir de dicho grafo se calcularon los módulos inconexos, como se describe en la [sección 3.6](#). La descripción de los UMs identificados en *iCG238* se encuentra resumida en la [tabla 4.3](#), que está ordenada en forma decreciente de acuerdo al tamaño de cada UM (en número de nodos). Las inconsistencias se encuentran agrupadas en un total de 16 UMs, de los cuales 6 están compuestos por un único metabolito que corresponden a *pseudo-gaps*. Este hecho indica que algunos de los metabolitos identificados *pseudo-gap* se encuentran formando parte de alguno de los UMs¹. El análisis individual de cada UM permitió identificar los metabolitos *gap* que corresponden a las clases RNP y RNC. Sin embargo, la discriminación entre los DNP y UNC no es posible en aquellos casos donde el UM contiene tanto RNPs como RNCs. Dado que la presencia de un RNP implica que la inconsistencia se propague corriente abajo y en el caso de un RNC esta se propaga corriente abajo, la presencia simultánea de metabolitos *gap* terminales de ambas clases trae como consecuencia una ambigüedad a la hora de clasificar los metabolitos *gap* de la UM que no son un terminal sino que son consecuencia de la propagación de inconsistencias tanto corriente arriba, como corriente abajo.

¹ Los pseudogaps suelen formar conjuntos de dos o más compuestos y por lo general suelen agruparse en UMs independientes como muestra el ejemplo de la [figura 3.8](#).

Al examinar el conjunto de reacciones involucradas en cada UM, se encontró que, en la mayoría de los casos, cada UM se superpone en buena medida con algún subsistema bioquímico, o ruta metabólica, previamente descrito en la literatura. Dado que por lo general cada reacción del modelo está asociada a un único subsistema bioquímico, cada UM fue asociada al subsistema mayormente representado en la UM, en término de la frecuencia de reacciones de la UM asociadas al mismo. Por ejemplo, todas las reacciones que forman parte de la UM1 constituyen la ruta de biosíntesis del menaquinol, luego este UM representa una reconstrucción inconsistente de dicha ruta. En el otro extremo ciertas UMs pueden estar compuestas por una única reacción o metabolito, como son los casos de los UMs 11-16. Para simplificar el análisis de cada UM, estos fueron representados y dibujados como grafos, de forma independiente. Después, cada grafo fue inspeccionado visualmente para detectar las posibles causas que producen la inconsistencia de todos los elementos involucrados en la UM. Por último, se realizó un proceso de *gap-filling* manual de forma que las modificaciones introducidas en el modelo permitan resolver las inconsistencias. A continuación se consideran uno a uno los casos y problemáticas planteadas por los distintos UMs, así como la forma de resolución de cada una.

Tabla 4.3 Descripción de los Módulos Inconexos detectados en iCG238. La tabla incluye la descripción del conjunto de UMs presentes en iCG238. El caso de las 5 UM compuesta por un único metabolito *pseudo-gap* y numeradas 12-17 fueron agrupadas en una única fila. La abreviación de los metabolitos se encuentra en la [tabla S.4.1](#).

UM	Sub-sistema asociado	Reacciones	Metabolitos	RNP	RNC
1	Biosíntesis de menaquinol	23	22	mev, 2ombzl	2ommb1, mql8, udcpdp
2	Ruta de recuperación de nucleótidos	22	15	-	hxn, xan, r1p, 2dr1p, thym, ura
3	Biosíntesis piridoxal-5-fosfato	7	6	-	4hthr
4	Biosíntesis de lipopolisacárido	4	4	-	u3hga
5	Biosíntesis de sirohemo	4	4	-	uppg3
6	Metabolismo de arginina y prolina	2	2	-	1pyr5c
7	Transporte extracelular (Fe ²⁺)	2	2	-	Fe ²⁺
8	Transporte extracelular (K ⁺)	2	2	-	K ⁺
9	Superóxido dismutasa	1	1	O ₂ ⁻	-
10	ACP Sintasa	1	1	apoACP	-
11-16	--	0	6	-	-
Total		70	68	4	15

UM1 – Ruta de biosíntesis de menaquinol

En primer lugar se realizó el análisis del UM de mayor tamaño que se encuentra conformado por 21 reacciones y 22 metabolitos, de los cuales dos (mql8, undecaprenil-pp) corresponden a metabolitos *pseudo-gaps*. En la [figura 4.1](#) se presenta en forma de grafo bi-partito el UM1, el cual se corresponde, en buena medida, con la ruta de síntesis del menaquinol. Este es un compuesto lipofílico de baja masa molecular, presente en muchas especies de bacterias y que funciona como transportador de electrones de la cadena respiratoria. La inspección detallada de la [figura 4.1](#)

muestra que el UM1 está compuesto por dos ramas que convergen en la síntesis de desmetilmenaquinol-8 (DMQ) a través de la actividad EC 2.5.1.74. Las ramas corresponden, por un lado, a la síntesis de 1,4-dihidroxi-2-naftoato partir del isocorismato, y por el otro a la síntesis del trans-octaprenil-difosfato. A su vez, de los 22 metabolitos presentes en el UM, se detectan dos RNP (mevalonato y 2-octaprenil-6-metoxi-1,4-benzoquinol) y tres RNC (2-octaprenil-3-metil-6-metoxi-1,4-benzoquinol, 2-desmetil-menaquinona-8 y undecaprenil-difosfato), mientras que el resto de los metabolitos se corresponden a DNC o UNP. Si en el modelo se permite la entrada libre de los RNP así como la salida de los RNC mediante la incorporación de las fuentes y sumideros correspondientes, todas las reacciones presentes en la UM dejan de estar bloqueadas y por ende los metabolitos dejan de ser *gap*. Esto indica claramente que dichos metabolitos son la causa última de las inconsistencias que representa el UM1 y como consecuencia un análisis detallado de los mismos contendrá la clave para resolver dicho UM.

La actividad 2.1.1.201 cataliza la conversión de 2-octaprenil-6-metoxi-1,4-benzoquinol en 2-octaprenil-3-metil-6-metoxi-1,4-benzoquinol, los cuales corresponden a un RNP y un RNC, respectivamente. Dicha actividad está asignada al locus BLBBGE_110, el cual ha sido anotado como un homólogo de *ubiE* y que codifica para una enzima, la C-metiltransferasa, la cual participa en la biosíntesis de la ubiquinona (Q). A su vez, *ubiE* participa en la biosíntesis del menaquinol (MQ) catalizando la conversión de 2-desmetil-menaquinol en menaquinol (EC 2.1.1.163) (Lee et al., 1997). Ahora bien, mientras que el genoma de *Blattabacterium* Bge posee los genes putativos que codifican todas las enzimas involucradas en la biosíntesis de MQ, este no contiene genes anotados que codifiquen para las actividades enzimáticas de la biosíntesis de Q, a excepción de *ubiE*. Por esta razón, el 2-octaprenil-6-metoxi-1,4-benzoquinol y el 2-octaprenil-3-metil-6-metoxi-1,4-benzoquinol son detectados en el modelo como metabolitos *gap* de tipo RNP y RNC respectivamente. Como consecuencia, la actividad EC 2.1.1.201 carece de contexto biológico en la red metabólica de *Blattabacterium*, y por esta razón no debe ser incluida en el modelo metabólico de dicho organismo. A su vez, al quitar del modelo la actividad EC 2.1.1.201 se quitan también dos de los metabolitos involucrados conflictivos. Después de quitar la actividad EC 2.1.1.201, el UM solo contiene un metabolito RNP que corresponde al mevalonato. El mevalonato es un intermediario en la ruta de síntesis del isopentil di-fosfato, el cual a su vez es un precursor involucrado en la biosíntesis de MQ y de DMQ. Esta ruta, referida como la ruta del mevalonato, se encuentra presente tanto en organismos eucariota como en algunas bacterias (Kuzuyama, 2002). La ruta incluye siete actividades enzimáticas de las cuales las tres primeras permiten la formación de mevalonato a partir de acetyl-CoA. Se realizó una inspección de la versión más reciente de la anotación del genoma, así como de la reconstrucción metabólica disponible en el BioCyc correspondiente a *Blacttabacterium* Bge para buscar genes candidatos que codifiquen las tres primeras enzimas de la ruta del mevalonato que se encuentran ausentes en el modelo. Sin embargo, no se encontró ninguno de los genes que codificaran para las actividades ausentes, lo que indicaría una posible pérdida parcial de dicha vía. Una consecuencia de esta posible pérdida sería que el mevalonato debería ser importado por la bacteria desde el ambiente en el que vive, *i.e.* las células del insecto hospedador en la cuales se encuentran alojadas. Esta situación sugiere un caso potencial de complementación metabólica entre la bacteria endosimbionte y su hospedador.

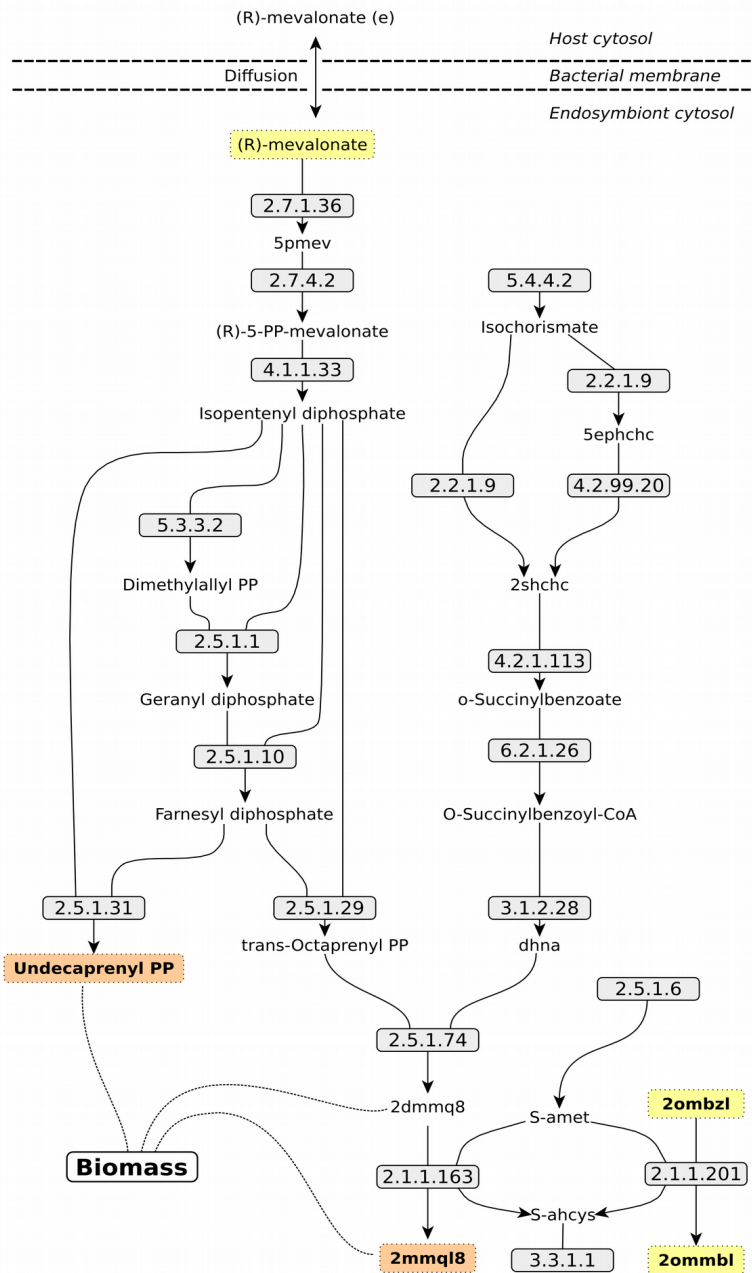


Figura 4.1 Representación esquemática del UMI. Los rectángulos grises representan reacciones, con sus correspondientes números EC. Los metabolitos están representados por sus abreviaciones: los rectángulos de borde punteado corresponden a RNP y RNC. El color amarillo indica que el metabolito es un gap mientras que rojo denota que el metabolito es un pseudo-gap. Se indican los metabolitos que fueron conectados a la ecuación de biomasa así como la complementación metabólica propuesta con el hospedador. La abreviación de los metabolitos se encuentra en la [tabla S.4.1](#).

Un argumento a favor de esta hipótesis se basa en el hecho de que la ruta del mevalonato tiene un papel esencial en el metabolismo de los insectos, ya que a través de esta ruta se sintetiza el precursor de la hormona juvenil (JH) (Rivera-Perez et al., 2014; Yi et al., 2006). A su vez, evidencias experimentales sugieren que la ruta se encuentra activa en el cuerpo graso de *B. germanica*, hospedador del *Blattabacterium* Bge (Bellés et al., 2005). Sobre el problema del transporte del mevalonato desde el exterior, experimentos de nutrición en insectos, en los que el mevalonato es empleado como precursor de la JH en el *copora allata*, sugieren la posibilidad de

que dicho metabolito pueda difundir hasta alcanzar el citosol del endosimbionte (ver (Bellés et al., 2005) para una revisión detallada). Si se supone el escenario de complementación metabólica propuesto donde el endosimbionte importa el mevalonato proporcionado por el hospedador, esto debería reflejarse en el modelo mediante la incorporación de un transporte que permita la importación del mevalonato, así como su correspondiente flujo de intercambio. Teniendo en cuenta esta modificación del modelo, la inspección del UM1 representado en la [figura 4.1](#) indica que el UM aún contiene dos metabolitos RNC: el menaquinol y el undecaprenil-difosfato. Es importante notar que ambos metabolitos fueron detectados como *pseudo-gaps* (previamente referidos como el problema de la coenzima *pseudo-gap* (Ponce-de-Leon et al., 2013)) y que la inclusión de los mismos en el cálculo de los UMs pone de manifiesto el hecho de que a pesar de participar en reacciones no bloqueadas, son la causa del bloqueo de otras reacciones (ver [sección 3.6](#)).

Por su parte el menaquinol es un metabolito que actúa como transportador de electrones en la cadena respiratoria, así como en otras reacciones de óxido-reducción. Cuando tales procesos tienen lugar en el sistema, siempre que se consume una de las formas de la coenzima se produce su conjugada de forma tal que la forma oxidada y reducida de la coenzima forman un *pool* conservado. Como consecuencia, para que pueda haber flujo a través de las reacciones que producen la coenzima en forma neta tiene que existir un flujo de degradación o dilución. El caso de la actividad EC 2.5.1.31 que produce undecaprenil-difosfato es en cierta forma similar, ya que este metabolito opera como una coenzima en la biosíntesis de la mureína y por lo tanto también forma un *pool* conservado. En los modelos a escala genómica, el efecto de la dilución de los *pools* de cofactores y coenzimas como consecuencia del crecimiento celular se suele representar mediante la incorporación de tales metabolitos a la ecuación de biomasa del modelo (Santos et al., 2011; Thiele and Palsson, 2010). En el caso del modelo *iCG238*, una vez que se permite la entrada de mevalonato dentro del endosimbionte y tanto el menaquinol como el undecaprenil-difosfato son incorporados en la ecuación de biomasa, el UM1 se resuelve completamente, lo que significa que todas las reacciones involucradas dejan de estar bloqueadas. Para los dos nuevos metabolitos incorporados en la biomasa del modelo, los coeficientes estequiométricos fueron extraídos de la definición de la biomasa del GEMM de *E. coli* *iJO1366* (Orth et al., 2011) y corregidos siguiendo la normalización propuesta por Henry et al. (2010).

UM2 – Ruta de recuperación de nucleótidos

El caso del UM2, compuesto por 22 reacciones y 15 metabolitos *gap*, se corresponde en buena medida a la ruta de recuperación de nucleótidos. En la [figura 4.2](#) se representa el UM2 y como puede verse, muchos de los números EC son compartidos por distintas reacciones. En efecto, las 22 reacciones del UM mapean en un total de 9 números EC indicando que las actividades enzimáticas involucradas en el UM presentan ambigüedad respecto a los sustratos que son capaces de metabolizar. Además de los 15 metabolitos *gap* incluidos en el UM2, 6 han sido clasificados como RNC. De acuerdo con la formulación de las relaciones gene-proteína-reacción de *iCG238*, doce de las reacciones del UM2 son huérfanas en el contexto del modelo. Es decir, que dichas actividades no tienen un gen codificante asociado (ver las reacciones indicadas en violeta en la [figura 4.2](#)).

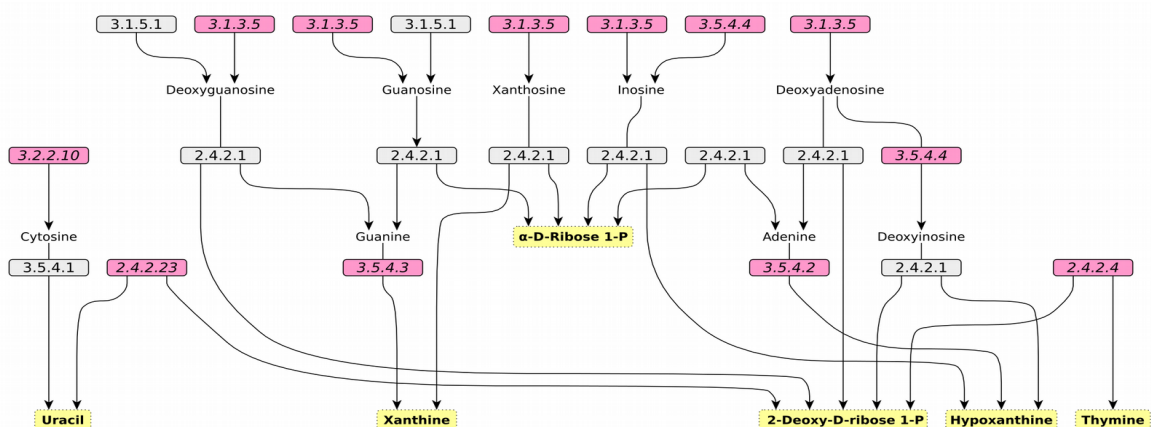


Figure 4.2. Representación esquemática de la UM2. Las actividades incluidas corresponden a la ruta de recuperación de nucleótidos. Las reacciones son representadas con rectángulos con su número EC. Las reacciones coloreadas de rosa son aquellas que no tienen un gen codificante asociado. Aquellos que se encuentran encerrados en rectángulos amarillos de borde punteado corresponden a RNP y RNC.

Las diez reacciones restante son catalizadas por tres enzimas diferentes asociadas en las siguientes formas: siete reacciones tienen asociada la actividad EC 2.4.2.1, que se encuentra asignada al gen BLBBGE_377 el cual se encuentra anotado como purina fosforilasa; dos reacciones se encuentran asociadas a la actividad EC 3.1.5.1, la cual está asignada al gen BLBBGE_612; la reacción restante asignada a EC 3.5.4.1, lleva asociado el gen BLBBGE_353. Debido al gran número de reacciones huérfanas presentes en el UM2 y la presencia de numerosos metabolitos *gap* de tipo RNC, una primer paso en el análisis del UM fue el de quitar todas aquellas reacciones que no tuviesen asignado un gen codificante.

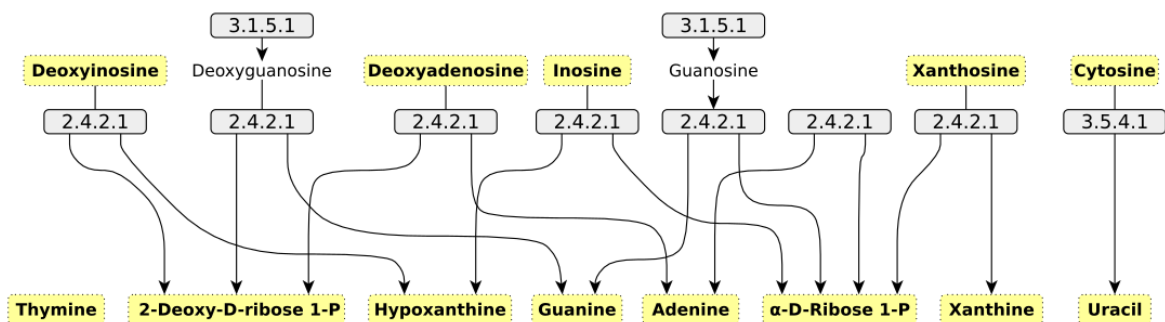


Figura 4.3. Representación esquemática del UM2 sin reacciones huérfanas. Las reacciones y metabolitos son representadas siguiendo el mismo esquema que en la figura 4.1. De acuerdo con la tabla gen-proteína-reacción las actividades presentes en el UM se asocian a los siguientes genes: 3.1.5.1→BLBBGE_612; 2.4.2.1→BLBBGE_377; 3.1.5.1→BLBBGE_353.

El resultado de quitar las reacciones huérfanas se encuentra representado en la figura 4.3. Como puede verse en dicha figura, una vez eliminadas las reacciones huérfanas, la estructura del UM2 se modifica drásticamente. De los 15 metabolitos presentes en el UM, uno queda completamente desconectado y, a excepción de la guanosina y la desoxi-guanosina, los restantes cinco metabolitos pasan a ser RNP y dos se suman a los seis metabolitos RNC que ya estaban presentes en el UM2. Este nuevo esquema pone de manifiesto que la presencia de ciertas actividades en el modelo puede deberse a un artefacto producto de errores en la anotación funcional. Con el propósito de evaluar

dicha posibilidad, se inspeccionó manualmente la anotación de los tres genes implicados en el UM2. En primer lugar, la anotación de BLBBGE_612 indica que es una proteína putativa anotada con dominios trans-membrana perteneciente a la familia HD². Si bien la secuencia presenta homología con dominio extracelular con actividad hidrolasa (7TM-HD), esta familia de proteínas se encuentran asociadas a la transducción de señales (Anantharaman and Aravind, 2003). Considerando que la anotación de BLBBGE_612 no se encuentra respaldada por suficiente evidencia que sostenga que el gen codifica para una enzima que catalice la actividad EC 3.1.5.1, y teniendo en cuenta que las reacciones asociadas a esta actividad se encuentran bloqueadas, es más plausible asumir que tal actividad no tiene lugar en el metabolismo del endosimbionte. Algo similar ocurre con el caso del gen BLBBGE_353 (EC 3.5.4.1). Si bien la secuencia forma parte de la super-familia de las amidohidrolasas dependientes de metal (PF01979.16), que incluye un gran número de enzimas, esta se encuentra anotada como proteína hipotética y no posee homólogos cercanos con la actividad enzimática EC 3.5.4.1 asignada. En forma complementaria, la reacción asociada a EC 3.5.4.1 (la desaminación de la citosina para dar uracilo) se encuentra bloqueada ya que tanto su sustrato la citosina como su producto el uracilo son *deadends* en el modelo *iCG238*.

Por otra parte, la asignación de la actividad EC 2.4.2.1 al gen BLBBGE_377 se encuentra respaldada por la presencia de genes homólogos. La secuencia del gen hipotético BLBBGE_377 se encuentra anotado por similitud como una nucleósido de purina fosforilasa. Sin embargo, la actividad EC 2.4.2.1 presenta muy poca especificidad de sustrato e incluye un total de siete reacciones (en el modelo *iCG238*), las cuales se encuentran todas bloqueadas por estar conectadas en forma directa con metabolitos RNP y/o con RNC (ver [figura 4.3](#)). El escenario que plantea el UM2 es complejo debido a que parte de la información de la anotación es ambigua, a que algunas de las enzimas son muy poco específicas, y al hecho de que el modelo corresponde a una bacteria endosimbionte. Como consecuencia, la resolución del rompecabezas metabólico que plantea esta situación requeriría de información experimental adicional, la cual no se encuentra disponible de momento. Como consecuencia, ante una situación semejante, las reacciones y metabolitos que conforman el UM2 son eliminados del modelo y la función de los tres genes mencionados queda en suspenso hasta que nueva información experimental permita arrojar más luz sobre este problema.

UM3 – Ruta biosíntesis del piridoxal-5-fosfato

Las siete reacciones que componen el UM3 se corresponden con la vía de síntesis del piridoxal-5-fosfato a partir de eritrosa-4-fosfato (ver [figura 4.4](#)). El piridoxal-5-fosfato o vitamina B6 es un cofactor esencial de un gran número de enzimas, predominantes en la síntesis y catabolismo de aminoácidos. Es, además, uno de los cofactores más versátiles y participa en reacciones de transaminación y decarboxilación, entre otras. La inspección visual de la UM3 indica la presencia de dos metabolitos RNC: 4-hidroxi-L-treonina y piridoxal-5-fosfato ([figura 4.4a](#)). Sin embargo, la ruta tal como se encuentra definida en *iCG238* contiene errores en su formulación respecto a las definiciones presentes en las bases de datos KEGG ([ko00750](#)) y MetaCyc ([pyridoxal 5'-phosphate biosynthesis I](#)). Por ejemplo, la reacción anotada como 4HTHRS no se corresponde con ninguna de las actividades que definen la ruta. En tales casos, la reacción fue eliminada. Adicionalmente, algunos números EC no se correspondían con las definiciones de las reacciones y, en tales casos,

2 HD es un dominio de proteínas muy conservado conservado en bacterias, arqueas y eucariotas, que se encuentra presente en una superfamilia de enzimas con actividad fosfohidrolasa (predicha o conocida).

las reacciones fueron formuladas de acuerdo a su correcta definición bioquímica. Una vez corregida la formulación de la ruta, el UM3 adopta la forma que representa la figura 4.4b. En este nuevo contexto el UM solo presenta un metabolito deadend de tipo RNC que corresponde el producto final de la ruta, es decir, al piridoxal-5-fosfato. Como ya fue discutido en la sección 3.5 y en la discusión sobre la resolución de la UM1, siempre que los cofactores no sean consumidos de forma neta por las reacciones del sistema, la resolución de las inconsistencias implicará la incorporación de tales metabolitos a la ecuación de biomasa del modelo. Dada la importancia particular que el piridoxal-5-fosfato tiene en el metabolismo, aún más en el caso de un endosimbionte que produce aminoácidos, la incorporación del cofactor en la ecuación de biomasa está más que justificada.

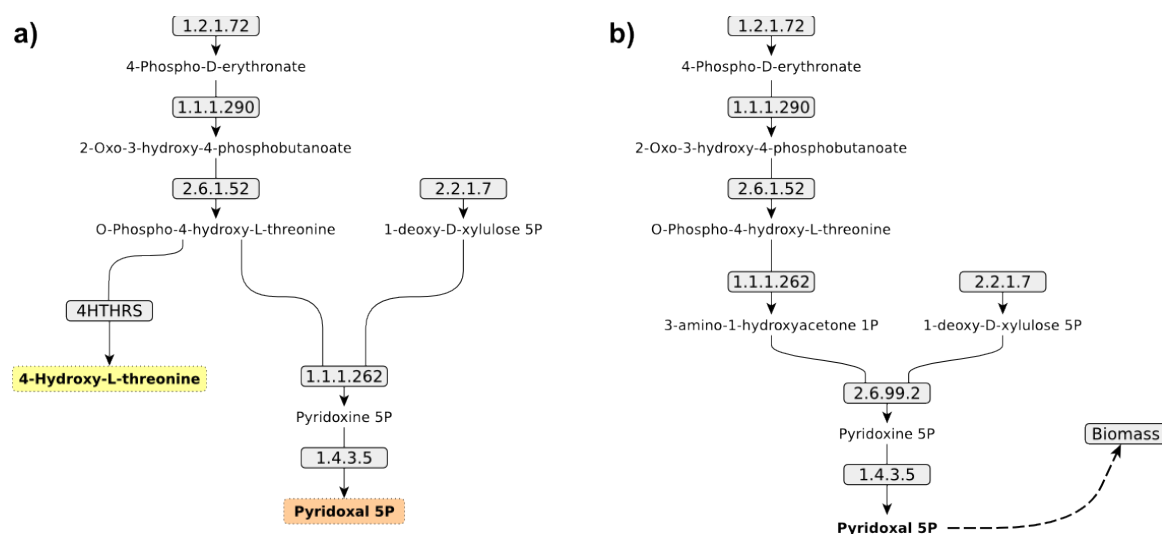


Figura 4.4. Ruta de síntesis de piridoxal-5-fosfato. En a) se representa la UM tal como fue identificada en el modelo iCG238. En b) se representa el UM tras corregir la formulación incorrecta de las reacciones 4HTHRS y EC 1.1.1.262. En el caso b) el único deadend es el producto final de la vía que al conectarse a la biomasa desbloquea el UM. Las reacciones están representadas por rectángulos grises con el número EC correspondiente. Los metabolitos encerrados en rectángulos de borde punteado corresponden a RNC. El color amarillo indica un gap mientras que anaranjado denota un pseudo-gap. La abreviación de los metabolitos se encuentra en la tabla S.4.1.

En el presente caso, dada la falta de información experimental sobre la composición del endosimbionte, se ha empleado como referencia el coeficiente estequiométrico del cofactor tal y como está definido en la ecuación de biomasa del modelo metabólico de *E. coli* iJO1260. Además, dado que los animales, y en particular los insectos, carecen de la ruta de biosíntesis del piridoxal-5-fosfato, estos deben obtenerlo por otros medios, e.g. la dieta, para poder sobrevivir (T. Tanaka et al., 2005). Como consecuencia, una hipótesis plausible es que *Blattabacterium* Bge provea al huésped de piridoxal-5-fosfato, como ha sido propuesto para el caso de otros endosimbiontes asociados a insectos (Michalkova et al., 2014; Nikoh et al., 2014) y también a eucariotas unicelulares (Klein et al., 2013). En cualquier caso, desde el punto de vista de la revisión del modelo, tanto la incorporación del cofactor a la biomasa como la adición de un transportador que permitan excretar el piridoxal producido, son suficientes para el desbloqueo completo de la UM.

UM4 – Ruta de biosíntesis de lipopolisacáridos

En este caso, el UM se encuentra compuesto por una cadena lineal de cuatro reacciones que se encuentran involucradas en la síntesis de distintos lípidos de membrana (figura 4.5a). La biosíntesis de lípidos, al igual que su degradación, es un proceso iterativo donde en cada iteración de la fase de elongación se repiten las siguientes cuatro reacciones: condensación, reducción, deshidratación y reducción. En el primer ciclo se condensan un acetil-ACP y un malonil-ACP, produciéndose una molécula de cuatro carbonos. Luego, con cada nueva iteración se añadirá un malonil-ACP, *i.e.* dos nuevos átomos de carbono. El producto final del proceso es siempre ácido palmítico, un ácido graso saturado de 16 carbonos. Como se puede ver en la figura 4.5, las dos primeras reacciones (C120SN, KAS16) forman parte de la ruta de síntesis del palmitato. Sin embargo, tal y como están definidas, estas no son reacciones bioquímicas sino que son reacciones condensadas, *i.e.* la reacción neta o global que resulta de una cadena lineal de reacciones. La reacción etiquetada como C120SN se corresponde con la condensación de las 19 reacciones que permiten la formación de dodecanoil-ACP a partir de la condensación sucesiva de cuatro moléculas de malonil-ACP al acetoacetil-ACP; por su parte KAS16 es la condensación de las actividades EC 2.3.1.41 y EC 1.1.1.100. Por otra parte, las siguientes dos reacciones del UM, EC 2.3.1.129 y EC 3.5.1.18, corresponden a los primeros pasos de la ruta de biosíntesis del lípido IV_A. Para facilitar la interpretación, el UM4 fue reformulado mediante el desagrupado de las reacciones C120SN y KAS16 en las actividades correspondientes (ver figura 4.6).

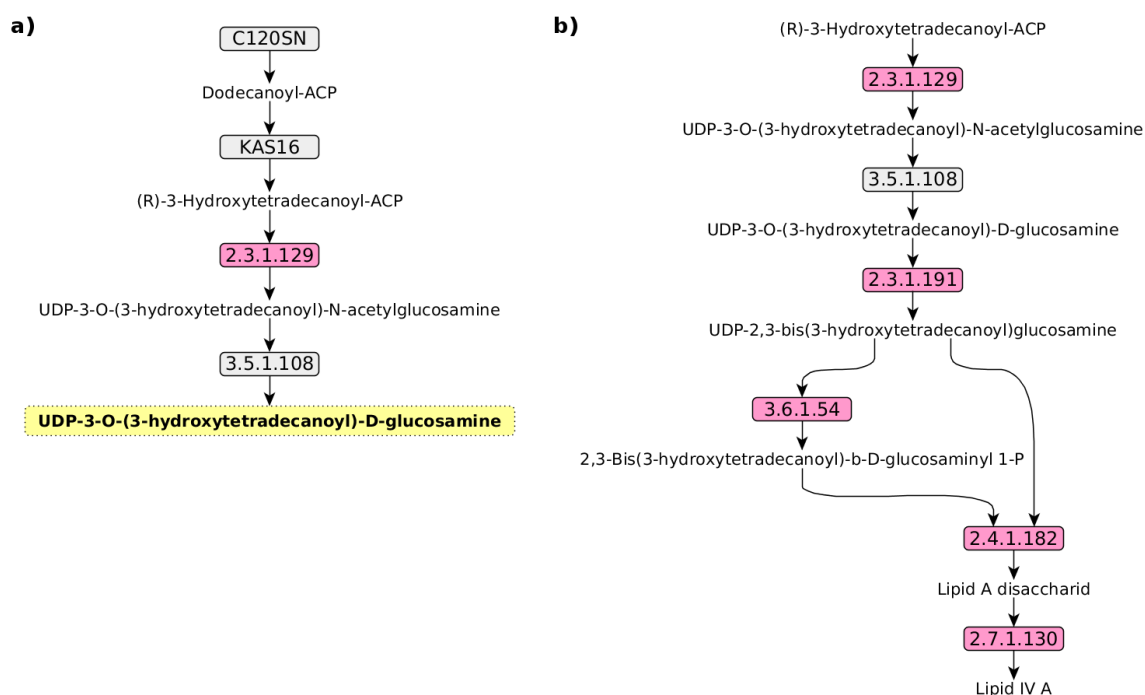


Figura 4.5. Representación del UM4 correspondiente a una porción de la ruta de síntesis lípidos de membrana. En a) UM identificada en el modelo iCG238. Las reacciones etiquetadas como C120SN y KAS16, corresponden a cadenas lineales de reacciones que han sido condensadas en la reacción neta. El UM incluye un único metabolito deadend de tipo RNC. En b) ruta de síntesis de lípido IVA en donde las dos primeras reacciones se corresponden con las finales del UM4. Los rectángulos de color representan las reacciones: los grises representan actividades con gen codificante; los rosa se son reacciones localmente huérfanas. Los rectángulos de color amarillo son metabolitos RNC.

De acuerdo con la definición de la ruta de síntesis del lípido IV_A (ver figura 4.5b) el modelo no incluye las cuatro reacciones que completan la biosíntesis del lípido IV_A. Además, el análisis del grafo muestra que: 1) las actividades 4.2.1.58 y 4.2.1.59 no tienen un gen asociado en el modelo de forma que son reacciones huérfanas; 2) que todas las actividades de la ruta de síntesis del lípido IV_A que se incluyen *iCG238* son huérfanas excepto por la actividad EC 3.5.1.108 que tiene asociado un gen codificante (BLBBGE_037), anotado como bifuncional, que se trata de una UDP-3-O-[3-hidroximiristoil]-N-acetilglucosamine-deacetilasa/(3R)-hidroximiristoil-ACP deshidratasa. Este escenario sugiere que la ruta de síntesis de lípido IV_A no se encuentra presente en el *Blattabacterium* Bge y que la asignación de la actividad EC 3.5.1.108 al gen BLBBGE_037 es posiblemente un error de anotación. En efecto, al comparar la secuencia de BLBBGE_037 con la de sus respectivos homólogos en las otras siete especies de *Blattabacterium* que hay secuenciadas, se encontró que la anotación del gen en las demás especies tiene asignado las actividades EC 4.2.1.58 y 4.2.1.59, actividades que se encontraban como huérfanas en el modelo *iCG238*. Como consecuencia, la anotación de BLBBGE_037 fue modificada del EC 3.5.1.108 a los EC 4.2.1.58 y 4.2.1.59. Las actividades de la ruta de síntesis de lípido IVA fueron eliminadas del modelo ya que no se identificó ningún gen que codifique para las actividades involucradas en dicha ruta.

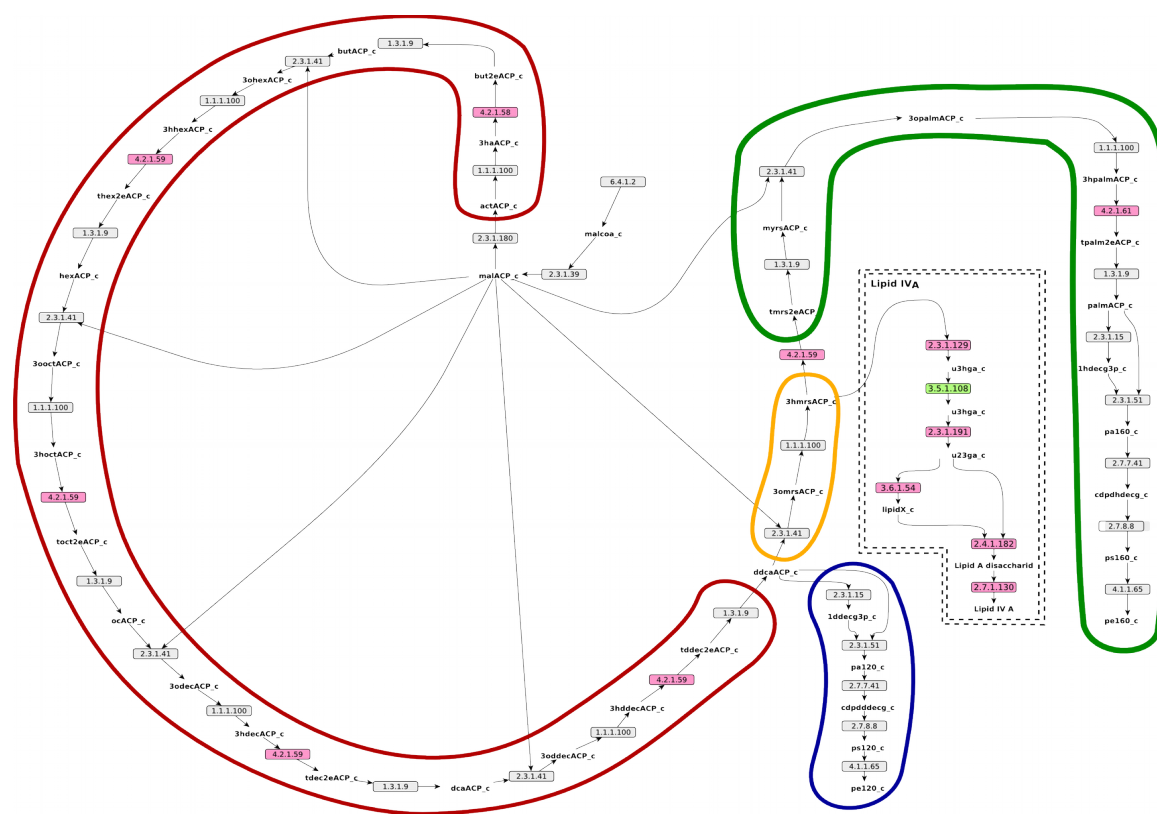


Figura 4.6. Representación del UM4 reformulado en el cual las reacciones C120SN y KAS16 se han descondensado. La representación también incluye tres bifurcaciones que corresponden a las rutas de síntesis de dos especies de fosfatidiletanolaminas y el lípido IVA. Los rectángulos de color representan las reacciones: los de color gris representan actividades con gen codificante asociado, los rosas se corresponden a reacciones huérfanas, i.e. sin gen codificante asociado y en verde se indican genes mal anotados. Las reacciones agrupadas como aparecen en el modelo *iCG238*: rojo C120SN; rojo + amarillo = KAS16; rojo + azul + amarillo + verde = PASYN_Bge DASYN_Bge + PSSA_Bge + PSD_Bge + C140SN + C141SN + C160SN + C161SN + C181SN.

UM5 – Ruta de síntesis del sirohemo

El UM5 está compuesto por una cadena lineal formada por dos reacciones y un transportador que importa porfobilinógeno al interior celular (ver figura 4.7a). Las dos actividades enzimáticas realizan la síntesis del uroporfirinógeno III a partir de porfobilinógeno, el cual es importado desde el exterior. Por su parte el uroporfirinógeno III, que en el modelo es identificado como *deadend* RNC, puede ser empleado como precursor para la síntesis de sirohemo, un grupo prostético de la familia de los grupos hemo, que ha sido identificado como cofactor de diversas enzimas, que incluyen a las nitrato y sulfito reductasas en las asimilación y desasimilación de los compuestos con sulfuro y nitrógeno (Crane et al., 1997).

La biosíntesis del sirohemo a partir del uroporfirinógeno III requiere de tres actividades enzimáticas, ninguna de las cuales ha sido incluida en el modelo *iCG238* (ver figura 4.7b). Sin embargo, una búsqueda específica en el genoma de *Blattabacterium* Bge permitió identificar los genes que codifican las actividades enzimáticas necesarias para completar la ruta de biosíntesis del sirohemo. En total se identificaron dos genes: BLBBGE_281 que codifica para la actividad 1.3.1.76 y el gen BLBBGE_278 que codifica una enzima que cataliza las actividades 2.1.1.107 y 4.99.1.4. Como se muestra en la figura 4.7b estas son las actividades que completan la ruta. El sirohemo es un cofactor esencial debido a su importante papel en el metabolismo del azufre y el nitrógeno. En consecuencia, al igual que los casos de cofactores, como la menaquinona o el piridoxal-5-fosfato, discutidos en las secciones previas, el sirohemo debe también formar parte de la ecuación de biomasa del modelo. De esta forma se representa la demanda del cofactor que es requerido por el organismo para poder crecer. Una vez incluidas las actividades ausentes y haber incluido al sirohemo a la ecuación de biomasa del modelo, las inconsistencias presentadas en el UM5 quedan resueltas.

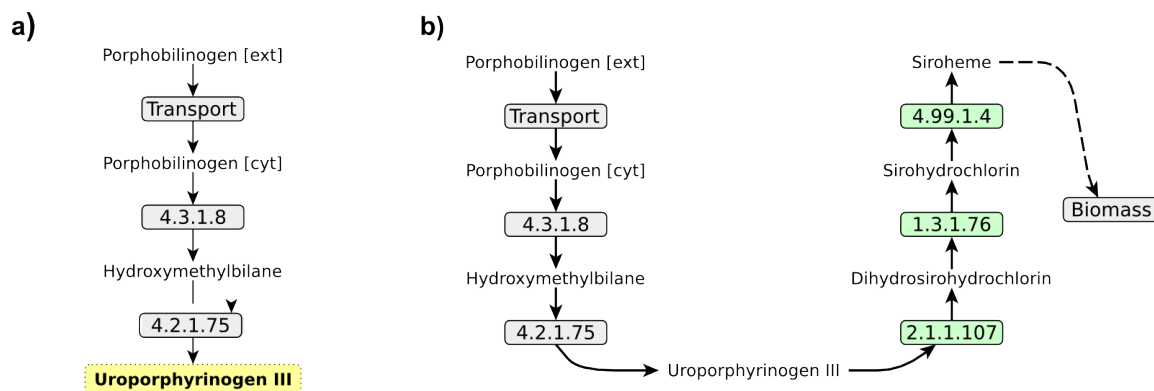


Figura 4.7. Representación del UM5 correspondiente a la ruta de síntesis del sirohemo a partir del porfobilinógeno. En a) se representa la estructura de la UM5 identificada en *iCG238* la cual incluye un transportador y dos actividades enzimáticas que desembocan en un metabolito RNC. En b) se representa la ruta de síntesis del sirohemo completa, después de haber encontrado los dos genes, BLBBGE_278 y BLBBGE_281, que codifican para las tres actividades no presentes en el modelo y que causaban la inconsistencia. En particular, se encontró que BLBBGE_278 codifica las actividades 2.1.1.107 y 4.99.1.4 mientras que BLBBGE_281 codifica la enzima que corresponde a la actividad 1.3.1.76. Finalmente, como se muestra en b) la resolución del UM también implica conectar el cofactor sirohemo a la ecuación de biomasa. Las rectángulos color verde corresponden a las actividades que fueron añadidas después de detectar los genes que codifican para las mismas en el genoma de *Blattabacterium* Bge.

UM6 – Ruta de biosíntesis de la prolina

En este caso el UM se encuentra compuesto por dos reacciones: la N-acetilornitina desacetilasa, codificada por el gen BLBBGE_320 (sin EC asignado) y la L-glutamato 5-semialdehído deshidratasa que corresponde a una reacción espontánea (ver figura 4.8). Dado que ambas reacciones se encuentran asociadas a la ruta de síntesis/degradación de la prolina, este escenario se podría sugerir que únicamente faltaría una actividad enzimática para completar la ruta lo cual se podría deber, o bien a un error de anotación o a una complementación metabólica entre el endosimbionte y su hospedador. Sin embargo, una inspección detallada del gen BLBBGE_320, el cual se encuentra asignado a la reacción NACODA, permite ver que dicho gen tiene asignadas las siguientes tres actividades:

- N-acetil-L-ornitina + H₂O → L-ornitina + acetato (EC 3.5.1.16)
- N-succinil-L,L-2,6-diaminopimelato + H₂O → L,L-diaminopimelato + succinato (EC 3.5.1.18)
- N-Acetil-L-glu 5-semialdehído + H₂O → Acetato + L-Glu 5-semialdehído (NACODA)

Tanto la primera como la segunda reacción, son actividades enzimáticas consistentes en el modelo que participan de la ruta de síntesis de la ornitina y la lisina. Por su parte la reacción (3) que es la reacción asociada al UM6, no fue encontrada en la base datos de actividades enzimáticas BRENDA (Schomburg et al., 2013). Sin embargo, se encontró que la reacción está definida en la base de datos BiGG (Schellenberger et al., 2010). Además, de acuerdo con MetaCyc, esta actividad se corresponde a EC 3.5.1.16 (N-acetil-ornitina desacetilasa, *argE* en *E. coli*) que también podría actuar sobre N-acetil-L-glu-semialdehído, aunque no se han encontrado referencias bibliográficas que la avalen. Por esa razón no está claro si la reacción ha sido caracterizada desde un punto de vista bioquímico, o por el contrario es una reacción ficticia o putativa formulada para completar algún *gap*. Por otra parte, algunos resultados experimentales indican que la prolina es el aminoácido más abundante en la hemolinfa de las cucarachas (Sowa and Keeley, 1996). Este dato apoya la hipótesis de que la prolina es un aminoácido proveído por el huésped y en consecuencia no sintetizado por el endosimbionte. Tomando en cuenta estos puntos, la hipótesis más plausible es que el UM6 sea un artefacto, producto de errores en la anotación y/o reconstrucción del modelo. Como consecuencia, las actividades incluidas en la UM6 fueron eliminadas del modelo.

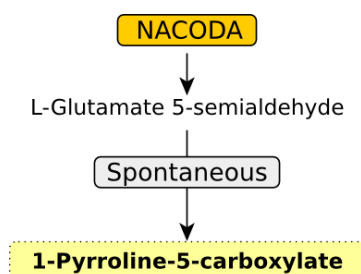


Figura 4.8. Representación del UM6 correspondiente un fragmento de la ruta de síntesis / degradación de la prolina. Los rectángulos de color representan las reacciones: los de color gris representan actividades con gen codificante asociado, mientras que los amarillos se corresponden a reacciones mal anotadas. Los rectángulos de borde punteado y color amarillo se corresponden a metabolitos RNC.

UM7-UM10 – El caso de las reacciones aisladas

El caso de los UM7 y UM8 cada uno incluye una única reacción que se corresponden con los transportadores de iones, Fe^{2+} y Mg^{2+} , respectivamente (figura 4.9a). Si bien ambos UMs incluyen también los flujos de intercambios asociados a los metabolitos extracelulares, no existe ningún proceso intracelular que consuma los iones y, por tanto, ambos metabolitos representan *deadends* de tipo RNC. Dado que la presencia de ciertos iones en el interior celular es esencial para el crecimiento de un organismo, una práctica común en la construcción de GEMMs es incluir dichos iones en la ecuación de biomasa (Thiele and Palsson, 2010). En el presente caso, como consecuencia de incluir los iones Fe^{2+} y Mg^{2+} en la ecuación de biomasa, ambas UMs se desbloquean. En el caso del UM9, la reacción aislada incluida se corresponde con la superóxido dimutasa (EC 1.15.1.1), la cual consume superóxido para producir peróxido (figura 4.9b). Esta actividad suele operar junto a la actividad catalasa (EC 1.11.1.6) en la ruta de detoxificación de los radicales libres, tales como superóxido y el peróxido de hidrógeno. Dado que ambas actividades tienen un gen asignado y que existe evidencia genómica así como simulaciones que sugieren el carácter aeróbico del metabolismo de *Blattabacterium* Bge (González-Domenech et al., 2012; Sabree et al., 2009), es altamente probable que la superóxido dimutasa desempeñe un papel importante en el metabolismo del endosimbionte.

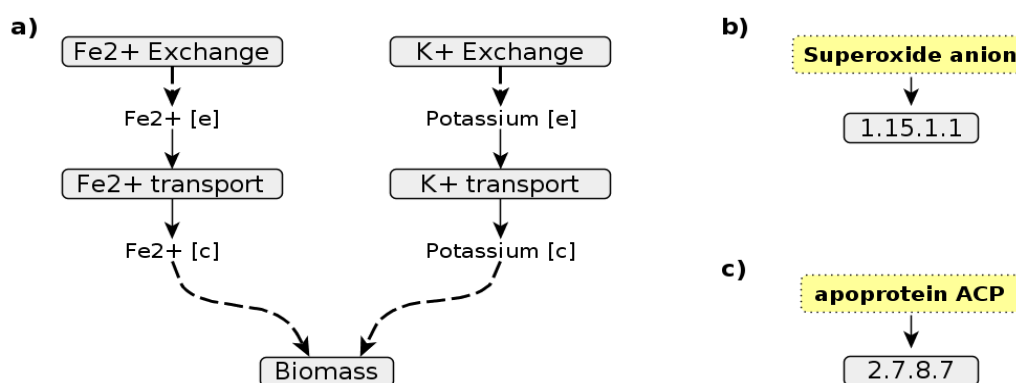


Figura 4.9. Representación de los UMs 7-10. En a) los UMs asociados con transporte de iones se resuelven al incluir el ion en la ecuación de biomasa. En b) y c) se representan los UM 9 y 10, respectivamente. En ambos casos los componentes incluidos en el UM fueron eliminados como parte del proceso de revisión.

La actividad catalasa no se encuentra bloqueada en el modelo debido a que el peróxido es un subproducto en el último paso de las síntesis del piridoxal-5-fosfato. Sin embargo, el sustrato de la superóxido dimutasa, el anión superóxido, se encuentra como un metabolito *deadend* de tipo RNP (figura 4.9b). Esto se debe a que la formación de tales iones es un subproducto de la respiración aeróbica, y su formación solo se da en ciertas condiciones que no se pueden representar de forma natural en un modelo estequiométrico y por tanto, los procesos de formación del superóxido se encuentran fuera del alcance del modelo. Como consecuencia de lo expresado anteriormente, la superóxido dimutasa puede ser eliminada del modelo sin que esto afecte la estructura de la red.

Finalmente, el análisis del UM10 indica también un caso de reacción aislada: la actividad EC 2.7.8.7 que cataliza la activación de una apoproteína en una proteína transportadora de acilos (ACP, acylt-carrier-protein), un transportador de grupos acilo, importante en las biosíntesis de ácidos grasos (figura 4.9c). Dado que el proceso de síntesis de proteínas no se encuentra explícitamente

incluido en el modelo *iCG238*, este no incluye una reacción que produzca la apoproteína, dando como resultado un metabolito deadend RNP que a su vez bloquea a la actividad EC 2.7.8.7. Como en el caso del UM9, previamente discutido, cuando un metabolito es producido por un proceso que se encuentra por fuera del alcance del modelo, este se manifiesta como un *deadend* causando el bloqueo de una o más reacciones. En el presente trabajo tales inconsistencias no tienen una resolución directa y, como consecuencia, los componentes involucrados pueden quitarse del modelo sin que esto afecte a la estructura del mismo.

4.3.2 Actualizando el modelo metabólico de *Blatabacterium Bge iCG238*.

El proceso de revisión mediante el análisis de módulos inconexos (UM), descrito en el apartado previo, dio como resultado la eliminación de seis genes asociados a reacciones del modelo *iCG238*, debido a que se detectaron errores en la anotación, o debido que las actividades catalizadas por los productos de dichos genes se encuentran fuera del alcance del modelo. A su vez, se identificaron ocho genes, previamente no incluidos en el modelo, que codifican para un total de nueve actividades. Además, la resignación e inclusión de nuevas actividades, así como la eliminación actividades huérfanas, condujo a la eliminación de 71 reacciones y 2 flujos de intercambios del *iCG238*, y la inclusión de 56 nuevas reacciones junto a 3 flujos de intercambio. En forma adicional, se introdujeron 13 nuevos metabolitos en la formulación de la ecuación de biomasa (ver [tabla 4.4](#)).

Tabla 4.4 Descripción de los nuevos metabolitos incluidos en la ecuación de biomasa de *iMP240*.

ID	Nombre	Fórmula
2dmmql8	2-Demetilmenaquinol 8	C ₅₀ H ₇₂ O ₂
mql8	Menaquinol 8	C ₅₁ H ₇₄ O ₂
5mthf	5-Metiltetrahidrofolato	C ₂₀ H ₂₄ N ₇ O ₆
amet	S-Adenosil-L-metionina	C ₁₅ H ₂₃ N ₆ O ₅ S
pydx5p	Piridoxal-5-fosfate	C ₈ H ₈ N ₀ O ₆ P
sheme	Sirohemo	C ₄₂ H ₃₆ FeN ₄ O ₁₆
udcpdp	Undecaprenil difosfato	C ₅₅ H ₈₉ O ₇ P ₂
murein5px4p	Disacarido pentapeptido de murein (2 unidades)	C ₇₇ H ₁₁₇ N ₁₅ O ₄₀
pe120	Fosfatidiletanolamina (didodecanoil, n-C12:0)	C ₂₉ H ₅₈ N ₁ O ₈ P ₁
pe160	Fosfatidiletanolamina (dihexadecanoil, n-C16:0)	C ₃₇ H ₇₄ N ₁ O ₈ P ₁
fe2	Fe ²⁺	Fe
k	Potasio	K
mg2	Magnesio	Mg

Dado que no hay información experimental disponible acerca de la composición bioquímica de la biomasa de *Blattabacterium Bge*, se utilizó como referencia la formulación de la biomasa del GEMM de *E. coli iJO1366* (Orth et al., 2011). En particular, los coeficientes estequiométricos de los nuevos cofactores incluidos en la nueva ecuación de biomasa del modelo, corresponden a los del modelo de *E. coli iJO1366*. Si bien, *Blattabacterium Bge* y *E. coli* son organismos filogenéticamente distantes, y por tanto su composición bioquímica puede llegar a diferir en forma

significativa, es muy posible que los coeficientes de los cofactores en *Blattabacterium* Bge deben ser, aproximadamente, del mismo orden de magnitud que en *E. coli* (Feist and Palsson, 2010; Henry et al., 2010). De esta forma es posible capturar las necesidades de crecimiento del organismo desde un punto de vista cualitativo. Las nuevas modificaciones introducidas en el modelo dieron lugar a una nueva versión del modelo de *Blattabacterium* Bge, denominado *iMP240* (Ponce-de-Leon et al., 2013). Los detalles sobre las modificaciones introducidas en el modelo se encuentran en la [tabla S.4.2](#).

4.3.3 Análisis de la reorganización de los subconjuntos de reacciones

Como se pone de manifiesto en las secciones previas, donde se detallan los pasos seguidos para curar cada UM, el proceso de revisión del modelo metabólico implica la incorporación de nuevas reacciones así como la eliminación de otras y en algunos casos la modificación de la estequiometría, o la reversibilidad/irreversibilidad. Estos cambios afectan a la estructura de la red y como consecuencia, las propiedades estructurales del sistema se verán modificadas. Por ejemplo, las modificaciones mencionadas tenderán a causar una reorganización en las relaciones de acoplamiento entre los flujos de la red. En particular, en este apartado se presentan los resultados del análisis de la reorganización de los conjuntos de reacciones parcial y completamente acopladas (FCS, *full coupling sets*) (Burgard et al., 2004) también denominadas subconjuntos enzimáticos (Pfeiffer et al., 1999) o subconjuntos de reacciones (Poolman et al., 2007). Como se discute en el [apartado A.2.3](#) del apéndice A, el conjunto de FCSs es una propiedad estructural e invariante de una red estequiométrica, y su análisis puede brindar información acerca de la regulación del sistema (Notebaart et al., 2008).

La [tabla 4.5](#) resume el número de FCS identificado en ambas versiones del modelo, así como el número de reacciones que contiene cada FCS. Para comparar la reorganización de FCS se empleó el índice de Jaccard como medida de similitud entre los FCS de ambos modelos. El índice fue calculado para cada par de FCS como el número de elementos (*i.e.* reacciones) en la intersección dividido el número de elementos en la unión de ambos FCS y, como consecuencia, el índice se encuentra acotado entre cero y uno. Cuanto mayor es el valor del índice, mayor es el número de reacciones compartidas entre dos FCS, siendo 1 en el caso que ambos FCS sean idénticos y 0 en el caso de que los FCS comparados no tengan ningún elemento en común. Al analizar la [tabla 4.5](#) la mayor diferencia en la reorganización de los FCS es la presencia de un FCS compuesto por 20 reacciones que está presente en *iMP240* pero en *iCG238*. Esta gran diferencia se explica debido a que en *iCG238* esas 20 reacciones no fueron incluidas de forma explícita sino que se encuentran condensadas en una única reacción (como se explica en la discusión del UM4). Esto mismo ocurre en el caso del FCS que contiene 13 reacciones y uno de los FCS de 7 reacciones identificados en *iMP240*. Además, otro de los FCS de 7 reacciones presentes en *iMP240*, que se corresponde con la ruta del síntesis del piridoxal-5-fosfato y tampoco se encuentra presente en *iCG238*; esto es debido a que las reacciones correspondientes a la ruta de síntesis del piridoxal-5-fosfato se encuentra bloqueadas en *iCG238* y, por definición, no pueden formar parte de ningún FCS. En forma adicional, se detectaron 3 FCS que son prácticamente idénticos entre ambos modelos, pero que difieren únicamente en una reacción; por ejemplo, el FCS de 18 reacciones en *iMP240* se corresponde al FCS presente *iCG238* que contiene 17 reacciones. Más allá de las diferencias antes descritas, no se detectaron cambios drásticos en la organización de FCS entre ambos modelos.

Tabla 4.5. Reorganización de los subconjuntos de reacciones completamente acopladas (FCS) en los modelos iCG238 e iMP240. Las columnas de izquierda a derecha indican: el número de reacciones que pertenecen al FCS; ii) y iii) cuantos FCS con un número dado de reacciones hay en los modelos iMP240 e iCG238, respectivamente; iv) cuantos pares de FCS de un tamaño dado son iguales entre ambos modelos; y v) cuantos FCS con un índice de Jaccard mayor a 0,75 han sido encontrados, tomando como referencia de la comparación del modelo iMP240. El número entre paréntesis indica la diferencia en reacciones para un par de FCS de la siguiente forma: un número positivo indica que el FCS presente en iCG238 tiene ese número de reacciones adicionales, y un valor negativo indica lo contrario.

# Reacciones en FCS	iMP240	iCG238	J _{1.0} *	J _{0.75}
2	27	28	21	--
3	8	6	3	1 (+1)
4	2	4	1	--
5	4	3	2	--
6	3	3	2	--
7	3	1	1	1 (+1)
8	--	1	--	--
9	4	3	3	--
10	1	1	--	1 (+1)
11	--	1	--	--
13	1	--	--	--
17	--	1	--	--
18	1	--	--	1 (-1)
20	1	--	--	--

4.3.4 Análisis comparativo en la predicción de medios mínimos

Dado un modelo metabólico con una ecuación de biomasa que represente el crecimiento del organismo es posible calcular o predecir cuál es el conjunto mínimo de nutrientes esenciales para el crecimiento de un cierto organismo. En este sentido, el modelado basado en restricciones permite generar hipótesis de medios mínimos de crecimiento a partir de un modelo. De esta forma es posible identificar nuevos componentes, previamente no considerados, que un organismo necesite para crecer. Estas predicciones pueden evaluarse en forma directa en un laboratorio, y en ciertos casos posibilitan la definición de un medio mínimo de crecimiento. Dada la imposibilidad de cultivar las bacterias endosimbiontes primarios y la compleja relación que existe entre el metabolismo del endosimbionte y su huésped, la predicción *in-silico* de medios mínimos puede ser de gran utilidad para entender la relación de dependencia que existe entre los dos miembros de la asociación.

El medio mínimo *in-silico* fue predicho para las dos versiones del modelo, *i.e.* iCG238 y iMP240, empleado el procedimiento descrito en el apartado A.2.4 del apéndice A. Los resultados obtenidos para el caso de iCG238 se presentan en la tabla 4.6 e incluyen: cuatro aminoácidos; oxígeno, iones y algunos cofactores, o precursores de cofactores como el nicotinato, que luego de ser importado se emplea para sintetizar NAD. Estos resultados reproducen los obtenidos previamente (González-Domenech et al., 2012). Sin embargo, si bien las predicciones realizadas

empleando la versión revisada del modelo *iMP240* incluyen todos los compuestos predichos por *iCG238* (excepto el S-dihidroorotato), estas también sugieren que el endosimbionte requeriría seis compuestos adicionales para poder crecer (ver [tabla 4.6](#)). El S-dihidroorotato es un precursor en la biosíntesis de las pirimidinas que en el caso de *iCG238* debe ser importado del exterior para que pueda haber síntesis de nucleótidos, ya que el modelo no incluye la actividad EC 3.5.2.3 que cataliza la conversión del N-carbamoyl-L-aspartato en S-dihidroorotato. Sin embargo, la revisión de la anotación del genoma de *Blattabacterium* Bge permitió encontrar que el gen BLBBGE_317 codifica para la actividad EC 3.5.2.3. Como consecuencia, al incluir en el modelo dicha actividad asociada a su respectivo gen, las predicciones de *iMP240* indican que el endosimbionte podría llevar a cabo la síntesis de bases pirimidínicas *de novo*, *i.e.* sin la necesidad de importar el S-dihidroorotato del exterior.

Tabla 4.6. Identificación de medios mínimos. Tabla comparativa entre los medios de cultivo mínimos predichos *in-silico* empleando la versión revisada del modelo *iMP240* y su versión previa *iCG238*.

Componentes de Medio	<i>iMP240</i>	<i>iCG238</i>
Tiamina	Requerido	Requerido
Nicotinato	Requerido	Requerido
Sodio	Requerido	Requerido
L-Glutamina	Requerido	Requerido
Sulfato	Requerido	Requerido
R-Pantotenato	Requerido	Requerido
Fosfato	Requerido	Requerido
L-Asparagina	Requerido	Requerido
L-Prolina	Requerido	Requerido
Glicina	Requerido	Requerido
O ₂	Requerido	Requerido
S-Dihidroorotato	--	Requerido
R-Mevalonato	Requerido	--
Glicerol	Requerido	--
Porfobilinógeno	Requerido	--
Fe ²⁺	Requerido	--
K ⁺	Requerido	--
Mg ²⁺	Requerido	--
Total	17	12

Los nuevos requerimientos predichos incluyen: 1) dos nuevos iones (K⁺, Mg²⁺) que fueron añadidos en forma directa a la ecuación de biomasa y el Fe²⁺ que es empleado para la síntesis de sirohemo, cofactor que también forma parte de la biomasa; 2) dos precursores que son necesarios para la síntesis de cofactores; y 3) el glicerol. Dentro de los precursores para la síntesis de cofactores se encuentra el mevalonato que, como ya fue discutido, es un compuesto necesario para la síntesis de las quinonas, MQ y DMK (ver subapartado sobre la discusión de la curación de la [UM1 – Ruta de biosíntesis de Menaquinol](#)). Por tratarse, muy posiblemente, de un organismo aeróbico estricto, las quinonas serían fundamentales para el funcionamiento de la cadena

respiratoria En consecuencia el mevalonato sería un compuesto esencial para el crecimiento del simbionte. Algo similar ocurre con el porfobilinógeno, que es precursor de la síntesis del sirohemo. Dada la importancia del sirohemo en el metabolismo del azufre y el nitrógeno se espera que la síntesis de este cofactor sea vital para endosimbionte y, por ende, para su huésped. Como consecuencia, el porfobilinógeno sería también un compuesto esencial para el endosimbionte, ya que luego de ser importado sería empleado para la síntesis de sirohemo.

Dado que los cofactores mencionados fueron incluidos en la biomasa, se espera que la versión revisada del modelo incluya los precursores respectivos como requerimientos nutricionales esenciales. Sin embargo, la presencia del glicerol en el medio mínimo fue algo no tan previsible. De acuerdo con las predicciones de *iMP240*, el glicerol importado sería fosforilado mediante la actividad EC 2.7.1.30 y el glicerol-3-fosfato obtenido es empleado para la biosíntesis de distintas especies de fosfatidilglicerol que forman parte de la de la ecuación biomasa en ambos modelos. En el caso de *iCG238* el glicerol-3-fosfato es sintetizado a partir de la dihidroxiacetona-fosfato, mediante la actividad EC 1.1.5.3 operando en el sentido reverso. Sin embargo, no fue posible encontrar en la bibliografía evidencias que indiquen o sugieran que la actividad EC 1.1.5.3 es reversible. Por lo tanto, si la reacción se considera irreversible en condiciones fisiológicas, el glicerol-3-fosfato no podría ser producido por el endosimbionte y como consecuencia, el glicerol debería ser importado del medio y fosforilado en el interior de la célula para poder sintetizar las especies de fosfatidilglicerol que se integrarán a la membrana de la bacteria.

4.3.5 Fragilidad, robustez y reducción genómica.

El conjunto de genes metabólicos esenciales de *Blattabacterium* Bge fue calculado *in-silico* empleando el modelo *iMP240* y los resultados fueron comparados con el conjunto de genes esenciales predichos previamente empleando el modelo *iCG238*. Para que la comparación fuese posible, la predicción de los genes esenciales se llevó a cabo de la misma forma que se realizó en la publicación de la primera versión (González-Domenech et al., 2012), es decir empleando FBA (ver apartado A.2.1 del apéndice A). La figura 4.10 resume los resultados de la comparación entre ambos modelos. El conjunto de 172 genes (~72%) predichos como esenciales en *iCG238* son predichos también como esenciales en *iMP240*. En forma adicional, con *iMP240* se predicen 30 nuevos genes esenciales, 25 de los cuales se encuentran en *iCG238* (predichos como no esenciales) y 5 fueron incluidos en *iMP240* como consecuencia de la revisión. Esto resulta en un total de 202 genes esenciales (~84%) predichos por *iMP240*. El aumento en la fragilidad de la red frente a la pérdida de genes que se observa luego de realizar la revisión del modelo se debe a la incorporación de nuevos compuestos en la ecuación de biomasa. Esto trae como resultado que los genes que codifican las enzimas cuyas actividades permiten realizar la síntesis de los compuestos incluidos la ecuación de biomasa, serán predichos como esenciales. Por ejemplo, los genes que codifican para la biosíntesis del menaquinol se predicen como no esenciales al emplear *iCG238* debido a que dicho cofactor no se encuentra en la ecuación de biomasa, lo que a su vez es una de las causas del bloqueo de la ruta (ver sección 2.5). Sin embargo, dado que el modelo predice que diversas actividades esenciales para el organismo requieren de la presencia de menaquinol (*e.g.* reacciones de la cadena respiratoria o la dihidroorotato deshidrogenasa) es de esperar que los genes de la ruta de síntesis del cofactor también los sean. De igual forma, los genes requeridos para la síntesis de los nuevos compuestos que se han añadido a la biomasa de *iMP240* son también predichos como

esenciales. Los 38 genes que son predichos como no esenciales por el modelo *iMP240* codifican para un total de 45 actividades enzimáticas distribuidas en diversas rutas metabólicas (ver [tabla 4.7](#)), principalmente la ruta de recuperación de nucleótidos, los procesos de transporte y las reacciones del ciclo de Krebs.

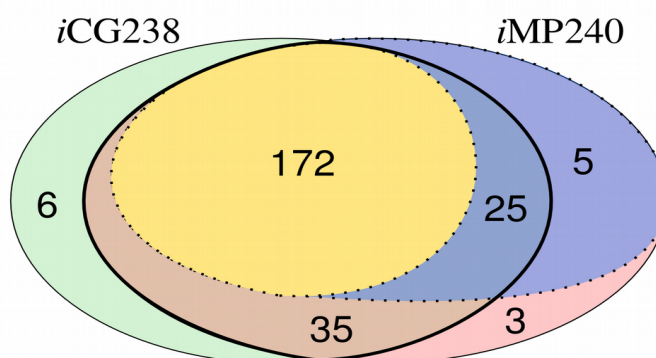


Figura 4.10. Comparación de genes esenciales en los modelos *iCG238* y *iMP240*. El diagrama de Venn representa las diferencias en los genes predichos como esenciales en los modelos *iCG238* y *iMP240*. Los conjuntos dibujados con una línea continua fina representan los genes incluidos en cada modelo. El conjunto delimitado por la línea continua gruesa representa el conjunto de genes presentes en ambos modelos, es decir, la intersección. Los conjuntos definidos por las líneas punteadas indican los genes predichos *in-silico* como esenciales.

La presencia de genes no esenciales indica que el sistema metabólico del endosimbionte aún conserva un cierto grado de robustez frente a la pérdida de genes. Dicho de otra forma, es posible que el proceso de reducción genómica aún no haya alcanzado su límite, lo que sugiere que el sistema podría reducirse aún más. Considerando este resultado se realizó un estudio *in-silico* de doble *knockout* para evaluar la posibilidad que dentro del conjunto de genes no esenciales, existan pares letales que indiquen la posibilidad de epistasis. El estudio se ha realizado evaluando todos los posibles pares entre los 38 genes no letales. La [figura 4.11](#) muestra los resultados del estudio de *knockout* dobles realizado sobre *iMP240*. Estos indican que existen 23 pares o combinaciones de genes cuya eliminación simultánea se traduce en la imposibilidad del metabolismo para exhibir crecimiento. Algunas de las combinaciones letales se deben a la eliminación simultánea de dos genes que codificaban para la misma función, *i.e.* que la letalidad se produce al perder una función y su copia redundante. Por ejemplo, existen dos genes que codifican para un transportador de hierro y dado que el hierro es un compuesto esencial para la síntesis de sirohemo, la pérdida de ambas copias sería letal para el endosimbionte. Algo similar ocurre con los genes *BLBBGE_290* y *BLBBGE_576*, que codifican para la misma proteína, la uroporfirinógeno III sintasa (EC 4.2.1.75), enzima involucrada en las síntesis de tetrapirroles y que en el caso de *Blattabacterium* Bge participa en la síntesis del sirohemo. Es llamativo el hecho de que el simbiote conserve dos copias de un mismo gen lo que podría indicar que codifica una función relevante. Por otra parte, si bien la similitud de secuencias entre ambas copias del gen es muy baja (< 40%), lo que podría sugerir que los genes desempeñan funciones diferentes, ambas presentan un dominio conservado de tipo HEM4 lo que indicaría que ambas conservan la actividad enzimática EC 4.2.1.75.

Tabla 4.7. Genes no esenciales agrupados por ruta metabólica. La tabla incluye los genes predichos como no esenciales en *iMP240* agrupados por la ruta metabólica en la que participa la enzima codificada.

Ruta metabólica	Nº de genes no esenciales
Ruta de recuperación de nucleótidos	18
Transporte, extracelular	10
Ciclo de los ácidos tricarboxílicos	6
Glicólisis/Gluconeogénesis	4
Biosíntesis de cofactores	4
Metabolismo de la glicina y serina	2
Otras funciones	2
Reacciones anapleróticas	2
Metabolismo de la arginina y la prolina	2
Fosforilación oxidativa	2
Ruta de las pentosas fosfato	1
Metabolismo de la fructosa y la manosa	1
Metabolismo del glutamato	1
Metabolismo de cofactores	1

Otros casos de *knockout* dobles letales no tienen una explicación tan directa, es decir que no son dos copias de un gen que codifica la misma función. Por ejemplo, cinco genes involucrados en la ruta glicólisis/gluconeogénesis que codifican las siguientes actividades: EC 4.2.1.11 (BLBBGE_264), EC 2.7.2.3 (BLBBGE_294), EC 5.4.2.12 (BLBBGE_231), EC 5.3.1.1 (BLBBGE_104) y EC 4.1.1.49 (BLBBGE_77), no son letales si son eliminados en forma individual. Sin embargo, algunas de las combinaciones de *knockout* dobles pueden ser letales. Por ejemplo si se inactiva la actividad EC 4.2.1.11, esto impediría la formación de fosfoenol-piruvato (necesario por ejemplo para la síntesis de corismato). Pero el metabolismo del organismo se puede ajustar redirigiendo flujo a través de la actividad EC 4.1.1.49 para poder generar el fosfoenol-piruvato. La pérdida de otras de las actividad mencionadas son también compensadas en forma similar a la descrita, aunque en algunos casos implican el reajuste de un número considerablemente mayor de flujos, aunque el valor de la tasa de crecimiento óptima no se ve afectado. Esto indica que pese a la reducción genómica, la red metabólica del endosimbionte aún conserva cierto grado de robustez distribuidas, es decir que el sistema puede compensar la falla o pérdida de funciones redirigiendo el flujo a través de rutas alternativas (Wagner, 2005). Además, las actividades mencionadas forman parte del metabolismo central del carbono cuyo diseño, altamente flexible, permite la interconversión de los distintos precursores por rutas alternativas (Noor et al., 2010). Algo similar ocurre con ciertas enzimas del ciclo de Krebs del endosimbionte, en donde las actividades citrato sintasa (EC 2.3.3.1), aconitasa (EC 4.2.1.3) e isocitrato deshidrogenasa son redundantes (EC 1.1.1.42) respecto a la oxoglutarato deshidrogenasa (EC 1.2.4.2), como ya ha sido predicho con *iCG238* (González-Domenech et al., 2012).

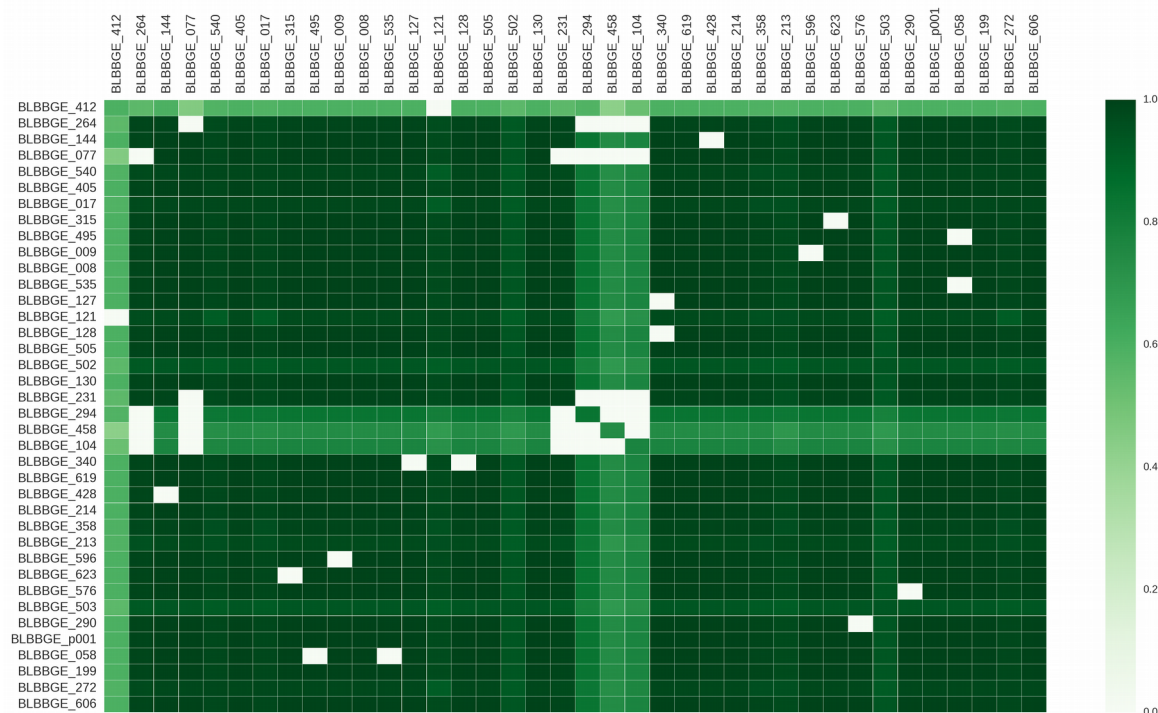


Figura 4.11. Predicciones in-silico de la viabilidad de knockout dobles en el modelo iMP240. La figura representa la matriz de todos los pares posibles de knockout dobles realizados sobre el conjunto de 38 genes no esenciales. La escala de colores indica el valor de la tasa de crecimiento relativo al valor predicho por el sistema sin perturbar.

4.4 Discusión

El cálculo de los módulos inconexos permitió identificar y resolver todas las inconsistencias presentes en el modelo de *Blattabacterium* Bge iCG238 y, al mismo tiempo, mejorar la anotación del genoma del organismo. Como se demuestra a lo largo de los distintos apartados, la representación e inspección visual de los diferentes módulos inconexos (UM) es una fuente de información muy útil que puede ser empleada para la resolución de las inconsistencias durante la reconstrucción y revisión de un modelo metabólico. En este sentido, cada reacción bloqueada detectada en el modelo ha sido convenientemente desbloqueada, o quitada del modelo en aquellos casos donde no se encontró información suficiente que respaldará la presencia de esta en el metabolismo de la bacteria. Por ejemplo, en el caso del UM2 correspondiente a la ruta de recuperación de nucleótidos, todas las actividades enzimáticas identificadas como bloqueadas y que no tenían asociado un gen codificante, o aquellas reacciones bloqueadas introducidas al modelo como consecuencia de la baja especificidad de ciertas enzimas, fueron eliminadas del modelo. El proceso de revisión del modelo sirvió como guía para realizar una minuciosa reanotación del genoma de la cual resultó la detección de ciertas actividades enzimáticas codificadas por el genoma que no habían sido consideradas en la primera versión del modelo y que eran la causa de ciertas inconsistencias. Un ejemplo claro puede verse en el caso de la revisión del UM5 correspondiente a la ruta de biosíntesis del sirohemo en el que la incorporación al modelo de de tres nuevas actividades enzimáticas, resolvió las inconsistencias presente en la formulación de dicha ruta. Además, el análisis de consistencia permitió la identificación de nuevos componentes de la

biomasa, principalmente cofactores, no considerados previamente, como por ejemplo la menaquinona, el sirohemo y el piridoxal fosfato.

Como consecuencia del proceso de revisión y de los cambios introducidos en el modelo, se propone una nueva versión de la reconstrucción a escala genómica de *Blattabacterium* Bge, identificada como *iMP240*. Se realizaron diversos análisis *in-silico* con el objetivo de identificar el impacto que las modificaciones del modelo sobre las predicciones previamente realizadas, así como con el objetivo de caracterizar otras propiedades del sistema no analizadas previamente. Las modificaciones introducidas en el modelo no reflejaron grandes cambios en la reorganización del conjunto de subconjuntos de reacciones fuertemente acopladas o FCSs. Si bien, en la versión más reciente aparecen nuevos FCSs como consecuencia del desbloqueo y adición de nuevas reacciones, los FCSs compartidos por ambas versiones del modelo no presentaron diferencias notorias. Sin embargo, la modificación en la formulación de la ecuación de biomasa sí introdujo cambios importantes, que se reflejan en las diferencias de la predicción de medios mínimos y fragilidad de la red. Por un lado, *iMP240* predice la esencialidad de dos nuevos compuestos (sin incluir iones) como son el caso del glicerol y el mevalonato, mientras que tales compuesto no se predicen como esenciales al emplear *iCG238*. El caso del mevalonato es de especial interés en el contexto de los endosimbiontes debido a que sugiere un nuevo posible evento de complementación metabólica entre ambos organismos. Por otra parte, las diferencias en la formulación de la ecuación de biomasa de ambos modelos se reflejan en las estimaciones de la fragilidad de la red. Mientras que *iCG238* predice que un ~72% de los genes metabólicos son esenciales, las nuevas predicciones elaboradas con *iMP240* sugieren que el ~84% de los genes metabólicos son esenciales. La nueva predicción sugiere que el metabolismo del endosimbiontes es aún más frágil que lo predicho previamente. Esto es debido, en buena medida, al hecho de que los endosimbiontes se encuentran sometidos a un extremo proceso de reducción genómica en el que solo se retienen los genes esenciales para el crecimiento del sistema insecto-bacteria. En el presente caso, la mayoría de los nuevos genes predichos como esenciales están involucrados en la síntesis de cofactores que al ser haber sido incluidos en la ecuación de biomasa son predichos como esenciales por el modelo.

De acuerdo con los resultados del estudio, las inconsistencias detectadas en el modelo se deben a una de las siguientes causas: 1) errores en la anotación del genoma que resultan en la asignación incorrecta u omisión de actividades enzimáticas; 2) la ambigüedad por el sustrato de las enzimas, *i.e.* la baja especificidad por el sustrato que exhiben diversas enzimas, puede conducir a la inclusión errónea de reacciones que carezcan de sentido en un modelo; 3) la presencia de eventos de complementación metabólica entre el endosimbionte y su hospedador; y 4) la definición incompleta de la ecuación de biomasa del organismo. La identificación de este tipo de inconsistencias ha resultado ser muy sencilla mediante la inspección visual de los módulos inconexos. El caso de estudio expuesto en el presente capítulo ha tenido como objetivo por un lado, para poner a prueba el método de revisión de modelos metabólicos a través de la resolución de módulos inconexos, propuestos en la el [sección 3.6](#). Por otra parte el estudio también ha permitido poner de manifiesto cuáles son las principales fuentes de errores, ya sea en la anotación del genoma o en la reconstrucción de la red, que luego se manifiestan como inconsistencias en el modelo, con especial hincapié en el caso de la reconstrucción de modelos metabólicos de bacterias endosimbiontes.

Capítulo 5

5 Complementación metabólica en consorcios endosimbiontes

5.1 Introducción

Las especies que coexisten en un cierto espacio y tiempo conforman un red de interacciones que es moldeada tanto por factores bióticos como abióticos (Faust and Raes, 2012; Seth and Taga, 2014). Desde un punto de vista molecular, dichas redes pueden considerarse como un entramado de diversos metabolismos que se interconectan mediante el intercambio de compuestos. Por tanto, una aproximación para el estudio de las distintas formas de interacción entre organismos puede basarse en el análisis de las interacciones nutricionales. En este contexto, los posibles escenarios de interacciones nutricionales entre distintos organismos pueden agruparse de la siguiente manera: 1) competición por un recurso limitante; 2) sintrofia, en la cual un organismo consume un subproducto producido por otro organismo; y 3) *cross-feeding* o intercambio de nutrientes, en la cual un organismo que sobre-produce un determinado compuesto permite que otros organismos auxótrofos puedan sobrevivir. Mientras que en el caso de la competición entre dos especies por un recurso una de las especies excluirá, de forma inevitable, a la otra (Hardin, 1960; Seth and Taga, 2014), en los casos donde hay sintrofia o intercambio de nutrientes, la interacción tenderá a estabilizar la coexistencia de ambas especies, que de otra forma se volvería competitiva. En particular, cuando el intercambio de nutrientes o precursores es bidireccional y, por lo tanto, beneficioso para ambos miembros, la interdependencia nutricional conducirá, en la mayoría de los casos, a un proceso de co-evolución. Además, el proceso co-evolutivo en especies que establecen una dependencia nutricional puede dar lugar a asociaciones simbióticas, como es el caso de los insecto, como pulgones, cícadas y cucarachas, que alojan bacterias endosimbiontes en tejidos especializado.

La [figura 5.1](#) representa las distintas etapas del proceso por el cual se establece la simbiosis obligada entre un hospedador y un endosimbionte. Un aspecto clave de este proceso es que los endosimbiontes se encuentran sujetos a un proceso de reducción genómica extrema, como consecuencia del confinamiento a la vida intracelular, (Moran, 1996; Moran and Bennett, 2014; Moya et al., 2008). El proceso de reducción genómica comienza cuando la bacteria infecta al

hospedador y puede conducir: i) a la extinción del simbiote; ii) al reemplazo por una nueva bacteria con un genoma más completo; y iii) al establecimiento de una relación de dependencia obligada. Una vez que se ha establecido una simbiosis, es posible que una segunda infección no reemplace al primer simbiote, sino que se forme un consorcio de endosimbiontes dentro del hospedador (Martínez-Cano et al., 2015). Debido a la reducción genómica, los genomas de los endosimbiontes suelen perder rutas metabólicas completas, y de esta forma se vuelven auxótrofos para diversos compuestos, como nucleótidos o aminoácidos. En otros casos las pérdidas de genes son parciales, de manera que los genomas suelen codificar rutas metabólicas fragmentadas o que contienen *gaps*. En algunos casos las rutas fragmentadas siguen siendo funcionales pero el endosimbionte requiere de un cierto intermediario para completar el proceso de biosíntesis, lo que aumenta el nivel de dependencia entre el simbiote y su hospedador (e.g. el caso de la ruta de biosíntesis de menaquinol en *Blattabacterium* discutido en el apartado 4.3.1).

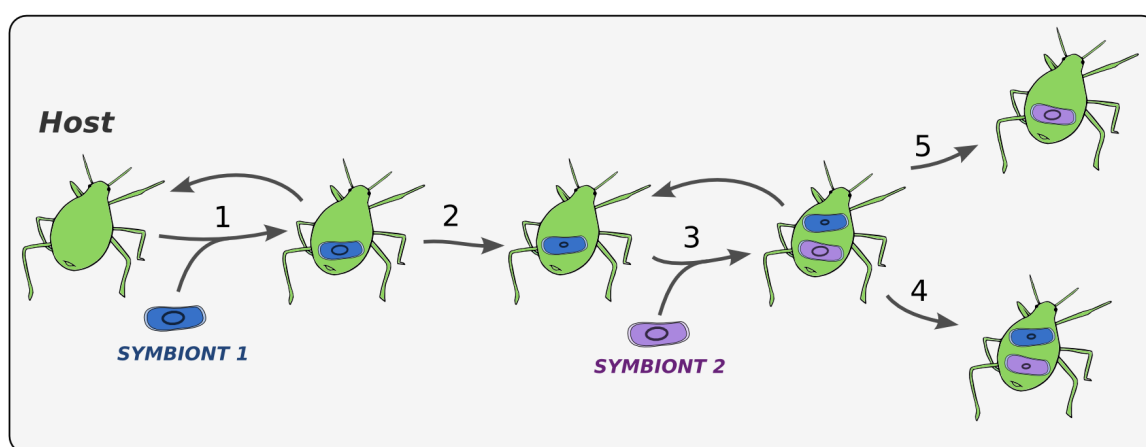


Figura 5.1. Evolución de consorcios endosimbiontes. La figura representa las posibles trayectorias evolutivas de la interacción entre insectos y bacterias endosimbiontes. Las flechas numeradas indican los diferentes estadios: 1) infección primaria; 2) reducción genómica y consolidación de la simbiosis; 3) infección secundaria (pérdida del endosimbionte secundario); 4) complementación metabólica y establecimiento de un consorcio; 5) reemplazo del endosimbionte primario por el secundario.

En el contexto de la evolución de endosimbiontes, el concepto de complementación metabólica o genómica es ampliamente empleado para describir las interacciones nutricionales, que como consecuencia de la reducción genómica, emergen entre los miembros de las asociaciones simbióticas (Baumann et al., 2006; Rao et al., 2015; Zientz et al., 2004). El concepto de complementación ha sido empleado en al menos dos sentidos diferentes. Por una lado, el término es empleado para referir al intercambio nutricional entre bacterias endosimbiontes y sus hospedadores, donde uno de los miembros produce y comparte uno o más compuestos esenciales (e.g. aminoácidos, vitaminas, etc) y vice-versa (López-Sánchez et al., 2009; Macdonald et al., 2012; Russell et al., 2014; Wu et al., 2006). En este sentido, el uso del término solapa con la definición de *cross-feeding*. Por otra parte, el término es también empleado para referir a escenarios más complejos, en donde rutas fragmentadas continúan siendo funcionales ya que se encuentran complementadas por genes del hospedador, o bien por genes de otro endosimbionte (Price and Wilson, 2014; Van Leuven et al., 2014). Es decir, que la pérdida y retención de genes da como resultado patrones genómicos en los que los genes que codifican para una ruta metabólica biosintética, se encuentran distribuidos en los genomas de un endosimbionte y su hospedador, o

entre diversos endosimbiontes. A lo largo de este capítulo se empleará el término “complementación metabólica” en el sentido de la segunda definición propuesta.

Se han descrito eventos de complementación metabólica entre bacterias endosimbiontes y sus hospedadores en diversos sistemas, en particular en insectos (Baumann et al., 2006). Por ejemplo, en diversos endosimbiontes responsables de proveer de aminoácidos esenciales a sus respectivos hospedadores (*e.g.* *Buchnera aphidicola*, *Tremblaya princeps*, *Candidatus Portiera aleyrodidarum*), se ha detectado la pérdida de ciertas actividades transaminasas, críticas en la síntesis de diversos aminoácidos (Baumann et al., 2006; Jiang et al., 2012; McCutcheon and von Dohlen, 2011; Rao et al., 2015; Sloan and Moran, 2012a; Zientz et al., 2004). En casos, como el de *B. aphidicola* BAp, se encontró que las transaminasas que perdió el endosimbionte podrían estar codificadas en el genoma del hospedador *Acyrtosiphon pisum*. Este hallazgo indica que las síntesis de los aminoácidos que requieren de dicha actividades transaminasas son sintetizados mediante la combinación de las capacidades metabólicas de ambos organismos (Hansen and Moran, 2011; Russell et al., 2013). Este mismo patrón también ha sido descrito en otros sistemas y se ha propuesto que el hospedador podría regular la producción de los aminoácidos sintetizados por el simbiote combinando la regulación de la actividad transaminasa con el suministro de precursores clave como la glutamina (González-Domenech et al., 2012; Poliakov et al., 2011; Price et al., 2014; Russell et al., 2014). Otros ejemplo de eventos de complementación metabólica también se han sido descrito en *Blattabacterium cuenoti*, endosimbionte de la cucaracha (González-Domenech et al., 2012; Ponce-de-Leon et al., 2013).

Escenarios más complejos de complementación metabólica pueden emerger cuando más de un endosimbionte coexisten en un mismo hospedador. Como muestra la [figura 5.1](#), después de que un endosimbionte consolida su relación del hospedador y pasa a ser un endosimbionte primario, su genoma se erosiona de forma irreversible lo que suele resultar en una reducción de las capacidades metabólicas del organismo. En tales casos es posible que se produzca una nueva infección por parte de un endosimbionte secundario (o facultativo) y que este establezca complementación metabólica con el simbiote primario, dando origen a un consorcio endosimbionte (Wu et al., 2006). Un ejemplo paradigmático de consorcio de endosimbiontes, es el que se ha descrito en el áfido del cedro *Cinara cedri* (Burke and Moran, 2011; Lamelas et al., 2011a; Pérez-Brocal et al., 2006). En este sistema coexisten dos poblaciones de bacterias endosimbiontes de especies diferentes. Por un lado, al igual que en la mayoría de los áfido, el endosimbionte primario es *B. aphidicola* BCc, y el segundo endosimbionte co-primario *Serratia symbiotica* SCc (Burke and Moran, 2011). En este sistema, muchas rutas biosintéticas, *e.g.* la síntesis de diversos aminoácidos esenciales, se encuentran codificadas únicamente en el genoma de uno de los simbiotes, lo que da lugar a que ocurra *cross-feeding*. Sin embargo, el caso del triptófano representa una excepción ya que los genes que codifican las actividades de la ruta se encuentran distribuidos entre ambos endosimbiontes: uno de los simbiotes (*B. aphidicola*) codifica las actividades del primer tramo de la ruta y el otro (*S. symbiotica*) codifica para el resto (Gosalbes et al., 2008). Como consecuencia, la biosíntesis del triptófano requiere de la presencia y coordinación de ambos endosimbiontes para poder ser realizada. Sorprendentemente, un caso convergente ha sido descrito en el psílido *Heteropsylla cubana* (Martínez-Cano et al., 2015), donde el simbiote primario, *Ca. Carsonella ruddii*, codifica el primer paso de la ruta, mientras que el endosimbionte secundario (que ha perdido casi todos lo

genes de las rutas de síntesis de aminoácidos) codifica para las actividades finales (Sloan and Moran, 2012b), y además el intermediario compartido es el mismo que para el caso de *C. cedri*. Otros ejemplos incluyen el consorcio endosimbionte de la cigarrilla de alas cristalinas *Homalodisca vitripennis*, que alberga a los endosimbiontes *Sulcia muelleri* y *Baumannia cicadellinicola* (McCutcheon and Moran, 2007). Escenarios más complejos han sido también descritos en el caso de una cícada del género *Tettigades* donde el endosimbionte primario *Candidatus. Hodgkinia. cicadicola* ha especiado de forma simpátrida en dos linajes citológicamente distintos pero genética y metabólicamente interdependientes (Campbell et al., 2015; Van Leuven et al., 2014).

La mayor parte de las complementaciones descritas se han identificado mediante análisis genómicos y se han limitado a describir tales eventos. Sin embargo, ninguno de estos estudios ha profundizado en el estudio de las posibles consecuencias que supone la división de una ruta metabólica en dos organismos. La presencia de la complementación metabólica entre los endosimbiontes y su hospedador, y en particular entre los miembros de un consorcio plantea diversas preguntas evolutivas: ¿la complementación supone algún tipo de ventaja adaptativa que sea beneficiosa para todo el sistema? Si la respuesta fuese no, se podría pensar que los patrones de complementación son producto de un proceso neutral de pérdida de genes en los que la ruta metabólica sigue siendo funcional (de lo contrario el sistema colapsaría). Por otra parte, si la respuesta es que sí, entonces surge la siguiente pregunta ¿en qué consiste dicha ventaja? Algunos trabajos teóricos han intentado dilucidar si organismos que exhiben algún tipo de cooperación, como *cross-feeding*, presentan algún incremento en su tasa de crecimiento, respecto a las cepas silvestres (Germerodt et al., 2016; Großkopf and Soyer, 2016; Pfeiffer and Bonhoeffer, 2004). Por otra parte, la complementación de una ruta metabólica entre dos endosimbiontes plantea ciertos problemas biofísicos; por ejemplo, el intercambio de intermediarios entre los endosimbiontes implica un problema de transporte. Por un lado está el hecho de que los endosimbiontes suelen disponer de un repertorio muy limitado de transportadores (Charles et al., 2011), y por otro está el hecho de que en general la mayor parte de los intermediarios de una ruta biosintética no suelen tener transportadores asociados. Esto implicaría que los intercambios tendrían que ser vía difusión y esto impone restricciones ya que no todos los compuestos pueden difundir a través de una bicapa lipídica.

Otro problema asociado a la complementación entre los miembros de un consorcio y su hospedador (y también al *cross-feeding*) es conocer cómo ajustan su metabolismo los endosimbiontes para sobre-producir ciertos compuestos de forma que satisfaga las demandas del hospedador. El flujo a través de una ruta metabólica puede ser incrementado mediante la modificación de distintas propiedades de las enzimas (Kacser and Burns, 1973). Mutaciones que afecten las constantes catalíticas, así como la afinidad por el sustrato de las enzimas, y que permiten aumentar el flujo, se pueden seleccionar (M. Ullah, H. Schmidt, 2006). En forma complementaria, el aumento del flujo puede realizarse mediante el incremento de las concentraciones de las enzimas, lo que implica modificar la expresión de los genes. Pero también hay que tener en cuenta los mecanismos de regulación que debieron estar presentes en los ancestros de los endosimbiontes. Si bien, los endosimbiontes suelen perder los mecanismos de regulación génica, el devenir de otros mecanismos como la regulación alostérica (muy frecuente en ruta de biosíntesis en la forma de inhibición por producto) no es tan claro. En efecto, en algunos estudios

se ha visto la pérdida de conservación en el sitio alostérico de una enzima (Jiménez et al., 2000), mientras que en otros casos se ha encontrado que los residuos que conforman el sitio de regulación se encuentran bien conservados (Lai et al., 1994; Rouhbakhsh et al., 1997).

Con el objetivo de profundizar en el estudio del fenómeno de la complementación metabólica, a lo largo de este capítulo se adoptaran dos estrategias complementarias. Por un lado, se llevará adelante la reconstrucción a escala genómica y el análisis de los metabolismos de *B. aphidicola* BCc y *S. symbiotica* SCc, endosimbiontes del cedro. Cada modelo será caracterizado de forma individual. Posteriormente, ambos modelos serán empleados en la construcción de un modelo compartimentalizado del consorcio endosimbionte del áfido del cedro. Dicho modelo de consorcio será empleado para estudiar las propiedades del diseño estequiométrico de la complementación metabólica en el caso de la ruta de biosíntesis del triptófano antes descrito. El objetivo es intentar identificar si el diseño en la complementación exhibido por el consorcio presenta ventajas frente a otros diseños alternativos que incluyan al caso de un consorcio ancestral hipotético en que la ruta estuviese codificada por ambos simbiontes. En forma complementaria, se plantea estudiar el fenómeno de la complementación desde un punto de vista cinético. Para esto se construirá un modelo cinético que represente el escenario en el que dos tipos de endosimbionte codifican una ruta metabólica necesaria para la síntesis de un producto esencial. Empleando dicho modelo se buscarán identificar condiciones necesarias para la emergencia de complementación metabólica, haciendo énfasis en tres aspectos: i) la presencia de inhibición por producto; ii) el papel de la permeabilidad de la membrana a los intermediarios y el producto final de la ruta; y iii) el coste asociado a la síntesis de las enzimas.

5.2 Materiales y métodos

5.2.1 Genomas anotados

Para realizar la reconstrucción y revisión de los modelos metabólicos a escala genómica de *Ca. Serratia symbiotica* SCc y *B. aphidicola* BCc se obtuvieron los genomas correspondientes, así como las bases de datos modelo-organismo reconstruidas en forma semiautomática y que se encuentran disponibles en la colección de PGDBs de BioCyc (Caspi et al., 2014). Las versiones de las bases PGDBs empleadas fueron: [SSYM568817](#) 19.0, para el caso de *Ca. Serratia symbiotica* SCc y [BAPH372461](#) 19.0 para el caso de *B. aphidicola* BCc (ambas disponibles en el Tier 3 de BioCyc 19.0). En el caso de la base de datos [BAPH372461](#), la versión pública disponible no incluye la información referente al plásmido pTpr-BCc con número de acceso [EU660486.1](#) (Gosalbes et al., 2008). Dicho plásmido, que solo incluye dos genes (*trpE* y *trpG*) fue añadido en forma manual a una versión local de la PGDB de *B. aphidicola* BCc, creada a través de la herramienta, *Pathway Tools* versión 19.0 (Karp et al., 2015). Una vez añadida la secuencia correspondiente del plásmido, se volvió a ejecutar el algoritmo *PathoLogic* (Green and Karp, 2004) para actualizar la predicción de rutas metabólicas del organismo, de manera que esta incluya los dos nuevos genes presentes en el plásmido. La información referente a diversas características de los genomas de ambas bacterias se encuentra resumida en la [tabla 5.1](#).

Tabla 5.1 Descripción de la anotación genómica de *Ca. S. symbiotica* SCc y *B. aphidicola* BCc

Organismo	Replicón	Genes	Proteínas	RNAs	Pseudogenes	Longitud (bp)
<i>S. symbiotica</i> SCc	NC_016632 (cr)	712	672	40	59	1.762.765
	Total:	712	672	40	59	1.762.765
<i>B. aphidicola</i> BCc	NC_008513 (cr)	430	358	72	3	416.380
	NC_011878 (pl)	5	5	0	0	6.054
	*EU660486.1 (pl)	2	2	0	0	2.795
	Total:	438	365	72	3	425.229

* El plásmido EU660486.1 no está incluido en la versión pública de la base de datos BAPH372461 y fue incluida de forma manual en una versión local de la misma. Abreviaciones: cr. (cromosoma); pl (plásmido)

5.2.2 Reconstrucción y refinado de modelos metabólicos

La reconstrucción y refinado de los modelos metabólicos se realizó siguiendo los pasos descritos en el protocolo propuesto por Thiele and Palsson (2010). Para llevar a cabo la reconstrucción metabólica de *Ca. S. symbiotica* SCc se tomó como referencia el modelo metabólico de *E. coli* K12 MG1655, debido a que es el organismo de vida libre filogenéticamente más cercano, que ha sido estudiado de forma extensiva y para el cual existe un modelo metabólico altamente refinado y validado. En concreto, el modelo metabólico empleado como referencia para la reconstrucción fue el denominado *E. coli* K12 iAF1260 (Feist et al., 2007b). El primer paso para la reconstrucción fue la identificación de genes ortólogos entre el genoma de *E. coli* K12 MG1655 disponible en EcoCyc (Weaver et al., 2014) y *Ca. S. symbiotica* SCc, empleando la versión del genoma disponible en MetaCyc (ver apartado 5.2.1). Las secuencias de los genes de cada organismo fueron extraídas y traducidas a aminoácidos; las secuencias de aminoácidos fueron comparadas contra las secuencias de *E. coli* K12 MG1655, empleando *Blastp* (Altschul et al., 1990) con un filtro de *e-value* de 10^{-15} y un mínimo de identidad de secuencia del 75%. Los ortólogos fueron identificados empleando la herramienta *OrthoMCL* (Li et al., 2003), combinada con el resultado obtenido del análisis con *Blastp*. En forma adicional, empleando la herramienta *PathwayTools* se extrajo la información de todas las reacciones y rutas metabólicas presentes en la anotación de la PGDB de *Ca. S. symbiotica* SCc, SSYM568817 19.0. Combinando esta información se realizó la construcción de la tabla de asociaciones gen-proteína-reacción (GPR). Empleando la tabla GPR y en combinación con el modelo iAF1260 junto con la base de datos BiGG (Schellenberger et al., 2010) se reconstruyó una primera versión de la red *Ca. S. symbiotica* SCc. La ecuación de biomasa fue construida empleando como referencia la de *E. coli* K12 iAF1260 y ajustando los coeficientes estequiométricos de acuerdo al procedimientos que emplea el *pipeline* del SEED (Henry et al., 2010). El modelo metabólico obtenido fue revisado empleando la aproximación de los módulos inconexos (ver sección 3.6 y el apartado 4.3.1) y la información disponible en la bibliografía referente al metabolismo del organismo (Burke and Moran, 2011; Lamelas et al., 2011a). Para el caso de *B. aphidicola* BCc, se empleó como referencia el modelo metabólico denominado BCc, previamente publicado (Belda et al., 2012). Dado que el modelo contenía inconsistencias, éste fue revisado en forma manual empleando la aproximación de los módulos inconexos, combinada con la

información disponible en la bases de datos modelo-organismo (BAPH372461), así como el modelo metabólico de *B. aphidicola* BAp previamente publicado (MacDonald et al., 2011) el cual se encuentra disponible en el repositorio público BioModels (MODEL1012300000).

5.2.3 Construcción de una ecuación de biomasa de *C. cedri*

En forma similar a la definición de la biomasa de los simbiontes, la ecuación de biomasa del áfido se define incluyendo los distintos componentes en la proporciones estequiométricas correspondientes. Dado que el modelo del consorcio no incluye el metabolismo del hospedador, la ecuación de biomasa del áfido solo incluye los compuestos que le son provistos por el consorcio de endosimbiontes, *i.e.* los aminoácidos esenciales, vitaminas y cofactores. Los coeficientes estequiométricos de los aminoácidos fueron estimados a partir de los del análisis de la composición del proteoma de los áfidos *Aphis fabae* y de *A. pisum* (Douglas et al., 2001; Russell et al., 2014). Dado que las mediciones de la composición fueron realizadas en dos especies de áfidos diferentes, esto permite suponer que los valores son extrapolables al caso de *C. cedri*. En el presente caso, hemos empleado los valores de las medidas experimentales de la composición en aminoácidos de la biomasa de *A. pisum*, realizados por Russell et al., (2014), debido a que la medidas fueron realizadas recientemente en base datos de proteómica. Estos valores fueron normalizados para que representen la composición en 1gDW de biomasa del áfido. En forma adicional se añadieron los cofactores que el consorcio de endosimbiontes provee al hospedador. En este caso, dado que no existe una estimación de su proporción en la biomasa del organismo, se introdujeron valores con órdenes de magnitud muy por debajo de la demanda de los aminoácidos, pero que reflejen el hecho de que tales compuestos son esenciales para el hospedador (ver [tabla S.5.5](#)).

5.2.4 Definición de una restricción de contorno

La definición de condición de borde del sistema, en un modelo metabólico reconstruido a escala genómica, puede llevarse a cabo estableciendo límites en los flujos de intercambio (ver [apartado A.1.3](#) del apéndice A). Por ejemplo, estableciendo una cota sobre el flujo de intercambio de glucosa se puede representar el límite que existe en la tasa de importación, para una cierta concentración del azúcar. Sin embargo, al modelar condiciones de crecimiento en medios complejos, es más difícil la definición de los límites sobre los flujos de intercambio, además de que cada cota implica un parámetro adicional. Por otra parte existe un límite físico en el total de moléculas que pueden atravesar la membrana en un cierto instante de tiempo (que se debe a que la superficie de la membrana es, a fin de cuentas, finita) (Zhuang et al., 2011). Dado que los endosimbiontes deben importar un gran número de compuestos del exterior y también exportar nutrientes, el número de parámetros, *i.e.* de cotas en los flujos de intercambio que deben establecerse es elevado, y no hay información experimental disponible. Una manera de reducir el número total de parámetros es introducir una restricción que limite el número total de átomos de carbono que atraviesan la membrana en cada instante de tiempo (Burgard et al., 2001). La restricción se puede expresar como la siguiente desigualdad lineal:

$$\sum_{i \in T_\alpha} v_i \cdot n_i \leq L \quad \forall \alpha \in \{BCc, SCc\} \quad (\text{Ec 5.1})$$

Donde $T_\alpha \subset J$ es el conjunto de índices de los transportadores presentes en el endosimbionte α , v_i es el flujo a través del transportador i , n_i es el número de átomos de carbono de la molécula

transportada, y L es la cota, *i.e.* el parámetro que establece el máximo número de átomos de carbono que se pueden intercambiar entre el compartimento del endosimbionte α y el compartimento extracelular. Es importante hacer notar que para garantizar que los valores de los flujos sean mayores o iguales a cero, los transportadores reversibles deben descomponerse en dos reacciones irreversibles con direcciones opuestas. Esta transformación no afecta la estructura del espacio de flujos (Wagner and Urbanczik, 2005), y garantiza que la contribución de los flujos de transporte en la ecuación 5.1 sea positiva, o nula.

5.2.5 Estimación de permeabilidades

Se han empleado diversos parámetros para estimar la capacidad de los distintos metabolitos de difundir a través de membranas. Para cada compuesto presente en los modelos se consideraron: (i) masa molecular; (ii) el número de donadores y aceptores de enlaces hidrógeno; (iii) el coeficiente de partición octanol-agua; y (iv) el área topológica de superficie polar (TPSA, Topological Polar Surface Area). La información estructural de los metabolitos fue extraída de la base de datos PubChem (Kim et al., 2015). A su vez, todos los parámetros fisicoquímicos fueron calculados empleando OpenBabel (a través de la interfaz, PyBel) con excepción del logaritmo del coeficiente de partición octanol-agua (LogP) que fue predicho empleando dos estimadores alternativos: i) el estimador AlogP (Tetko and Poda, 2004) y ii) el estimador XLOG3 (Cheng et al., 2007). De forma complementaria, los parámetros y estimadores antes descritos fueron empleados para calcular la regla del 5 de Lipinski (Lipinski et al., 2001) así como su versión extendida (Veber et al., 2002). La regla del 5 de Lipinski son cuatro criterios empíricos que los compuesto con buena permeabilidad suelen cumplir, e incluyen: i) que el compuesto no tenga más de cinco donadores de puentes de hidrógeno; ii) que compuesto tenga menos de 10 aceptores de puentes de hidrógeno; iii) que la masa molecular sea inferior a 500 Da; y iv) que el coeficiente de partición octanol-agua del compuesto sea menor a cinco. La regla del 5 enuncia que los compuestos que exhiben una buena permeabilidad suelen violar como máximo uno de los cuatro criterios previamente enunciados. A su vez, la extensión de la regla propuesta por Veber et al. (2002), incluyen los siguientes dos criterios adicionales: i) que la superficie polar del compuesto sea menor a 140 \AA^2 y ii) que el número total de donadores y aceptores de puentes de hidrógeno sea menor que 12. Por último, para predecir la permeabilidad de los compuesto, se emplearon dos de los clasificadores QSAR, basado en reglas, propuestos por Pham-The et al. (2013), que permiten clasificar la permeabilidad de los compuestos en: Alta (A), Media Alta (MA), Media, Media Baja (MB) y Baja (B). De los distintos clasificadores propuestos, se han empleado los clasificadores denominados 1PRule y 3PRule; el primero clasifica la permeabilidad de un compuesto empleando únicamente la PSA y el segundo combina el peso molecular, el PSA y LogD. Dado que no se dispone de un estimador de LogD, se empleó el estimador AlogP. en su lugar Los resultados se encuentran en la [tabla S.5.6](#) del apéndice B.

5.2.6 Alineamiento múltiple de secuencias

Las secuencias de aminoácidos del componente I de la antranilato sintasa (*trpE*) proveniente de doce cepas de *Buchnera* diferentes (incluyendo la cepa BCc), junto con el homólogo correspondiente en *Escherichia coli* y *Serratia marcescens*, fueron descargados de las base de datos UniProtKB (UniProt-Consortium, 2014). Los identificadores de las catorce secuencias se encuentran en la [tabla 5.2](#). El alineamiento múltiple fue realizado empleando el paquete de software

T-coffee (Notredame et al., 2000), ejecutado con los parámetros por defecto. La representación del alineamiento se llevó a cabo utilizando la herramienta *Jalview* (Waterhouse et al., 2009).

Tabla 5.2 Descripción de las secuencias de *trpE*. Descripción de las catorce secuencias empleadas en el alineamiento múltiple realizado para evaluar el grado de conservación del sitio alostérico del componente I de la antranilato sintasa de *B. aphidicola* BCc.

ID (UniProtKB)	Organismo	Largo (aa)
B4YQR1	<i>B. aphidicola</i> (<i>Cinara cedri</i>)	519
Q5WQ06	<i>B. aphidicola</i> (<i>Cinara tujafilina</i>)	514
G2LQ71	<i>B. aphidicola</i> (<i>Uroleucon ambrosiae</i>)	525
Q44695	<i>B. aphidicola</i> (<i>Acyrtosiphon pisum</i>)	521
Q89A29	<i>B. aphidicola</i> (<i>Baizongia pistaciae</i>)	519
Q44697	<i>B. aphidicola</i> (<i>Diuraphis noxia</i>)	519
Q9ZER9	<i>B. aphidicola</i> (<i>Pemphigus spyrothecae</i>)	520
Q44689	<i>B. aphidicola</i> (<i>Rhopalosiphum maidis</i>)	514
Q44691	<i>B. aphidicola</i> (<i>Rhopalosiphum padi</i>)	514
P42387	<i>B. aphidicola</i> (<i>Schizaphis graminum</i>)	515
Q44598	<i>B. aphidicola</i> (<i>Schlechtendalia chinensis</i>)	530
Q9ZES2	<i>B. aphidicola</i> (<i>Tetraneura caerulescens</i>)	526
P00895	<i>E. coli</i> K12	520
P00897	<i>S. marcescens</i>	519

5.3 Resultados

5.3.1 Reconstrucción metabólica de los endosimbiontes del áfido del cedro

Descripción de los modelos

La reconstrucción de un modelo metabólico de *S. symbiotica* SCc comenzó con la búsqueda de genes ortólogos en *E. coli*, como se detalla en el apartado 5.2.2. De las 712 secuencias anotadas como genes en el genoma de *S. symbiotica* SCc (ver tabla 5.1) un total de 665 fueron identificadas como genes ortólogos con respecto al genoma de *E. coli* K12 (MG1655). Dichas secuencias fueron filtradas para eliminar todos aquellos genes que no codificasen actividades enzimáticas, obteniéndose un total de 236 genes. En forma adicional, los 47 genes no identificados como ortólogos fueron analizados en forma manual y se identificaron 2 genes codificantes de actividades enzimáticas: el gen SCc_569 que codifica para NAD⁺ sintasa dependiente de NH₃ (EC 6.3.5.1 y EC 6.3.1.5), y el gen SCc_469 que codifica la subunidad beta de la riboflavina sintasa (EC 2.5.1.9).

La construcción de la tabla de asociación GPR (Gen-Proteína-Reacción) se realizó en base al listado de 236 genes metabólicos, empleando como referencia los complejos predichos por *Pathwaytools* así como el modelo metabólico de *E. coli* iJO1366. A partir de dicha tabla se infirieron las reacciones presentes en el metabolismo de *S. symbiotica* SCc, las cuales fueron extraídas de la base de datos BiGG (Schellenberger et al., 2010). De esta forma se obtuvo una primera versión de la red, la cual incluía 208 reacciones, 17 transportadores con gen asociado y 5

reacciones de transporte vía difusión. La ecuación de biomasa introducida en el modelo es una versión modificada de la presente en *iAF1260*, de la cual se eliminaron componentes de membrana y cofactores no presentes en *S. symbiotica* SCc. Los coeficientes fueron corregidos siguiendo la metodología propuesta por Henry et al. (2010), que implica corregir la estequiometría de los desoxinucleótidos y los aminoácidos empleando la frecuencia de las bases en el genoma y de los aminoácidos en el proteoma (virtual) del organismo. Finalmente se añadieron los flujos de intercambio correspondientes y una reacción de consumo de ATP de mantenimiento.

Dado que esta primera versión del modelo no predijo producción de biomasa se realizó un análisis de consistencia empleando la aproximación de los módulos inconexos (ver [sección 3.6](#)). Como resultado se identificaron 8 *gaps* en distintas rutas metabólicas, reacciones que fueron completados en forma manual, lo que implicó añadir las siguientes reacciones huérfanas al modelo: i) dos actividades de la ruta de síntesis de coenzima A (EC 4.1.1.36 y EC 6.3.2.5); ii) dos actividades de las biosíntesis del tetrahidrofolato (EC 4.1.2.25 y EC 3.6.1.-) involucradas en la síntesis del dihidropterin-CH₂OH-difosfato; iii) una actividad de la ruta de síntesis del piridoxal fosfato (EC 1.4.3.5); y iv) tres actividades involucradas de la ruta de síntesis de distintos componentes de membrana. Combinando el análisis de módulos inconexos con la predicción de medios mínimos (ver [apartado A.2.4](#) en el apéndice A) se identificó un conjunto de 19 compuestos que deben ser necesariamente intercambiados con el huésped; 11 de dichos compuestos deben ser importados e incluyen aminoácidos y precursores, y 8 son productos secundarios que deben ser excretados (los detalles sobre las condiciones de crecimiento serán abordados en el siguiente apartado). Como consecuencia de estos resultados se añadieron al modelo un total de 19 reacciones de transporte así como los correspondientes flujos de intercambio. Las reacciones de transporte incluidas carecen de un gen codificante y por tanto fueron designadas como reacciones huérfanas. Finalmente, debido a que diversos estudios proponen que el papel de *S. symbiotica* SCc en la simbiosis es la síntesis de diversas vitaminas y cofactores (Lamelas et al., 2011a, 2011b), se añadieron 24 reacciones de transporte que permiten exportar dichos compuestos. El resultado de la reconstrucción metabólica de *S. symbiotica* SCc es un modelo a escala genómica del metabolismo del endosimbionte denominado *iSCc236*, que incluye 236 genes los cuales codifican para un total de 239 actividades enzimática y 35 transportadores. Además, *iSCc236* incluye 8 reacciones huérfanas necesarias para la producción de ciertos componentes de la biomasa que no tienen gen asociado y 19 transportadores huérfanos también predichos como esenciales (para detalles del modelo ver [tabla S.5.1](#) en el apéndice B).

En el caso de *B. aphidicola* BCc, el modelo publicado por Belda et al., (2012) fue revisado combinando la aproximación de los módulos inconexos con la anotación funcional de genoma. Como consecuencia del análisis, 8 genes fueron eliminados del modelo por alguna de las siguientes tres razones: i) la actividad enzimática asociada se encuentra fuera del alcance del modelo, *i.e.* la reacción tiene lugar en un proceso que no se encuentra representado en el modelo; ii) la actividad enzimática se encontraba aislada y por lo tanto bloqueada; iii) el gen se encontraba mal anotado y la actividad enzimática no tiene lugar en el metabolismo del organismo. En forma complementaria se identificaron dos genes, previamente no incluidos en la versión del modelo publicada por Belda et al., (2012) que codifican actividades involucradas en la síntesis del lipoato (ver [tabla S.5.2](#) en el apéndice B). Dado que la versión previamente publicada del modelo no incluía compartimento

extracelular ni reacciones de transporte, se incluyeron los 62 metabolitos extracelulares así como las 62 reacciones de transporte correspondientes. En este caso, dado que en *B. aphidicola* BCc no se conocen transportadores con genes asociados (Charles et al., 2011), todos los transportes son considerados reacciones huérfanas, excepto los 5 casos de transporte que se sabe pueden ocurrir por difusión.

Tabla 5.3. Descripción de los modelos metabólicos de *S. symbiotica* SCc iSCc236 y *B. aphidicola* BCc iBCc98. La tabla describe la estructura de la reconstrucciones metabólica a escala genómica de los endosimbiontes primario y co-primario del áfido del cedro.

Elemento	iSCc236	iBCc98
Genes metabólicos	236	98
† Reacciones con gen asociado	209	95
Reacciones huérfanas	11	8
† Transportadores con gen asociado	30	5
Transportadores huérfanos	49	58
Metabolitos (intracelulares / extracelulares)	267 / 78	155 / 65
Flujos de intercambio	78	65

† El recuento de las reacciones y transportadores con gen asociado incluyen los procesos espontáneos y no incluyen la ecuación de biomasa.

Finalmente, se revisó la formulación de la ecuación de biomasa del modelo previamente publicado (Belda et al., 2012), a la cual se añadieron: nueve aminoácidos que no estaban presente, los cuatro desoxirribonucleótidos, y dieciséis cofactores y coenzimas que no estaban incluidas en la versión original (ver [tabla S.5.2](#) en el apéndice B). El resultado de la revisión dio lugar a una nueva versión del modelo metabólico de *B. aphidicola* BCc que ha sido denominado iBCc98. Como se muestra en la [tabla 5.3](#), el modelo iBCc98 incluye 98 genes que codifican para un total de 95 actividades enzimática y 8 reacciones huérfanas de las cuales 6 son catalizadas por las siguientes 3 actividades enzimáticas que corresponden a dos actividades transaminasa (EC 2.6.1.42 y EC 2.6.1.58) y una desaminasa (EC 4.3.1.19). En el caso de las transaminasas, algunos experimentos sugieren que dichas reacciones son catalizadas por el huésped (Hansen and Moran, 2011; Wilson et al., 2010). Por su parte, en el caso de la actividad EC 4.3.1.19 se han propuesto distintas explicaciones alternativas (Douglas et al., 2002): i) que otra enzima del endosimbionte cataliza dicha reacción, pero con una menor especificidad por el sustrato; ii) que una enzima accesoria podría realizar dicha actividad; y iii) que, como en el caso de las transaminasas, dicha actividad sea catalizada por el huésped. Por simplicidad en la formulación del modelo, las tres actividades han sido añadidas a la red metabólica del endosimbionte como reacciones huérfanas. En forma similar, otras dos reacciones huérfanas correspondientes a las actividades 2.1.1.197 (Malonil-CoA metiltransferasa) y la 2.3.1.47 (8-amino-7-oxononanoato sintasa) ambas involucradas en la ruta de síntesis del 8-amino-7-oxononanoato, también fueron añadidas al modelo.

Análisis *in-silico* de condiciones de crecimiento

Con el objetivo de establecer las condiciones de contorno o interfaces entre los endosimbiontes y su huésped (el áfido del cedro), se realizó un estudio *in-silico* para predecir las requerimientos

nutricionales de *B. aphidicola* BCc y *S. symbiotica* SCc. El cálculo de las condiciones mínimas de crecimiento de cada bacteria se realizó aplicando la aproximación descrita en el apartado A.2.4 del apéndice A sobre los modelos metabólicos descritos en el apartado previo. Los resultados del análisis de medios mínimos se encuentran resumidos en la tabla 5.4. Por un lado *iSCc236* predice que *S. symbiotica* SCc, requiere un total de 22 compuestos orgánicos. Estos incluyen a la hipoxantina como precursor para la síntesis de nucleótidos, un total de quince aminoácidos, y algunos cofactores o precursores para la síntesis de los mismos¹. Pese a su reducida capacidad para sintetizar aminoácidos, *S. symbiotica* SCc es capaz de sintetizar la cisteína *de novo*, mientras que el aspartato, el glutamato y la glicina son sintetizados a partir de la asparagina, la glutamina y la serina, respectivamente.

Tabla 5.4. Componentes esenciales del medio mínimo predichos in-silico por *iSCc236* y *iBCC98*. La primera columna indica las clases de compuesto, mientras que la segunda y tercera corresponden a los requerimientos predichos para *S. symbiotica* SCc y *B. aphidicola* BCc, respectivamente. Los aminoácidos resaltados son los provistos por el hospedador.

Clase	<i>S. symbiotica</i> SCc <i>iSCc236</i>	<i>B. aphidicola</i> BCc <i>iBCC98</i>
Nucleótidos	Hipoxantina	dATP, dCTP, dGTP, dTTP
Aminoácidos	Ala, Arg, Asn, Gln, His, Ile, Leu, Lys, Met, Phe, Pro, Thr, Ser, Tyr, Val, Anth*	Ala, Asn, Asp, Cys, Glu, Pro, Ser, Tyr, Trp* Ornitina, H-cyc
Coenzimas	Porfobilinógeno, Tiamina, 8AONN, (R)-Pantotenato, Shiquimato	NADP, NAD, FAD AMP, GTP, CTP, UTP Riboflavina, UQ8, Piridoxal-5P Tiamina-2P, Malonil-CoA, THF, SAM
Fuente de carbono	D-Glucosa / D-Fructosa	D-Glucosa-6P
Compuestos inorgánicos	Zn ²⁺ , SO ₄ ²⁻ , Fe ²⁺ , K ⁺ , O ₂ , HO ₄ P ²⁻	H ₂ S, Fe ²⁺ , K ⁺ , HO ₄ P ²⁻

Las abreviaturas de los metabolitos se encuentra en las tablas suplementarias S.5.1 y S.5.2.

* BCc es capaz de sintetizar el antranilato, pero ha perdido los genes que codifican las demás actividades de la ruta, mientras que SCc emplea el antranilato producido por BCc para poder sintetizar triptófano.

Estas predicciones son congruentes con los resultados previamente obtenidos en bases a estudios de genómica comparativa (Lamelas et al., 2011a). Sin embargo, si bien Lamelas y colaboradores proponen que *S. symbiotica* SCc sería capaz de sintetizar asparagina, las predicciones de *iSCc236* indican que el organismo no es capaz de sintetizar dicho aminoácido y como consecuencia este debe ser provisto por el hospedador. Además, *S. symbiotica* SCc es capaz de sintetizar triptófano pero a partir del antranilato que, como ha sido mostrado, es provisto por *B. aphidicola* BCc, en lo que constituye un caso paradigmático de complementación metabólica en endosimbiontes

1 Los detalles sobre la relación entre las capacidades biosintéticas del organismo y los requerimientos nutricionales, serán tratados en el subapartado: Capacidades biosintéticas y energéticas predichas por *iSCc236* y *iBCC98*.

(Gosalbes et al., 2008; Lamelas et al., 2011a; Manzano-Marín et al., 2016; Martínez-Cano et al., 2015). Otro caso de complementación metabólica entre ambos simbioses ocurre en la ruta de síntesis de la biotina. En efecto el modelo predice que *Serratia* debe importar 8-amino-7-oxononanoato (8AONN), precursor para la síntesis de biotina que como se verá en el apartado [Capacidades biosintéticas y energéticas predichas por iBCc98](#), este compuesto es producido por *Buchnera*, como ya ha sido propuesto (Manzano-Marín et al., 2016). Finalmente, *iSCc236* predice que el endosimbionte también requiere diversos compuestos inorgánicos y cinco precursores que son empleados en la síntesis de distintos cofactores. Dentro del conjunto de compuestos inorgánicos se encuentra el oxígeno, lo que sugiere que el organismo podría ser aeróbico estricto. En el caso de *iSCc236*, la dependencia del oxígeno no es consecuencia de una necesidad energética, sino que, de acuerdo con el modelo, el oxígeno es necesario para oxidar el ubiquinol que se genera en la síntesis *de novo* de pirimidinas. Además, el oxígeno también es empleado como aceptor de electrones durante la síntesis del sirohemo, cofactor esencial para la asimilación del azufre durante la síntesis de la cisteína. Finalmente, las predicciones indican que la red metabólica de *S. symbiotica* SCc puede emplear como fuentes de carbono, tanto la glucosa como la fructosa de forma indistinta.

Por su parte, *iBCc98* predice que *B. aphidicola* BCc debe importar todos los desoxinucleótidos, así como los nucleósidos trifosfato. Debido al proceso de reducción genómica extrema que ha sufrido el simbiote, este requiere también la incorporación de todos los cofactores necesarios por las diferentes enzimas presentes en el metabolismo, como por ejemplo la ubiquinona, la riboflavina y el tetrahidrofolato. Si bien es ampliamente aceptado que la principal función de *Buchnera* en la relación de simbiosis es la de sintetizar aminoácidos esenciales, en el caso de *B. aphidicola* BCc, ésta requiere la incorporación de un total ocho aminoácidos. Además, en el caso de la síntesis de triptófano esta se realiza de forma conjunta entre ambos endosimbiontes (Gosalbes et al., 2008). En el caso de *B. aphidicola* BCc, el organismo carece de sistemas de transporte PTS y de la capacidad de fosforilar la glucosa. De acuerdo con las predicciones del modelo, la única forma que el organismo tiene de emplear una fuente de carbono es a través de azúcares fosforilados, *e.g.* glucosa-6-fosfato, ribulosa-5-fosfato, y fructosa-6-fosfato. Sin embargo, dado que los demás azúcares fosfato son convertidos a glucosa-6-fosfato antes de ser metabolizados, su inclusión en el modelo no aporta información adicional. Como consecuencia, *iBCc98* solo incluye el transporte de la glucosa-6-fosfato como fuente de carbono.

Capacidades biosintéticas y energéticas predichas por *iSCc236*

Empleando el modelo metabólico de *S. symbiotica* *iSCc236*, descrito en el apartado previo, se realizó el análisis de las capacidades biosintéticas y energéticas de dicho organismo. Las capacidades biosintéticas fueron evaluadas mediante el cálculo de los requerimientos mínimos necesario para la síntesis individual de cada componente de la biomasa. Estos patrones se corresponden a un subconjunto de las conversiones elementales y permiten inferir qué compuestos del medio mínimo son necesario para sintetizar cada componente de la biomasa del organismo (Urbanczik and Wagner, 2005). Por su parte, las capacidades energéticas son evaluadas a través del análisis de la distribución de flujos que optimiza la producción de ATP. La [figura 5.2](#) representa la relación entre los compuestos sintetizados por *S. symbiotica* SCc (columnas) y los compuestos que requiere para realizar dichas síntesis (filas). Como se puede ver en la figura, de los cinco

aminoácidos que *S. symbiotica* SCc es capaz de sintetizar, el glutamato y el aspartato son sintetizados en forma trivial mediante la desaminación de la glutamina y la asparagina, respectivamente. De forma similar, la glicina sería sintetizada a partir de la serina mediante la actividad hidroximetil-transferasa (EC 2.1.2.1). En el caso de la cisteína, *S. symbiotica* SCc incluye todos los genes de la ruta de biosíntesis y, como consecuencia las predicciones indican que el organismo es capaz de sintetizar dicho aminoácido *de novo*, a partir de serina y sulfato. El caso del triptófano, como ya fue mencionado en el apartado previo, es quizás el ejemplo más paradigmático de complementación metabólica entre miembros de un consorcio de dos endosimbionte. Los genes que codifican las actividades de dicha ruta metabólica se encuentran distribuidos entre los genomas de ambos organismos (Gosalbes et al., 2008); concretamente *S. symbiotica* SCc ha perdido los primeros genes de la ruta (*trpGE*) pero ha conservado las actividades finales que permiten la síntesis del aminoácido a partir del antranilato (*trpDCBA*). *B. aphidicola* BCc por su parte, ha retenido las actividades que permiten realizar la síntesis del antranilato a partir de corismato (*trpGE*), este último sintetizado a partir de la eritrosa4-fosfato (*aroH, B, Q, y E*).



Figura 5.2. Mapa de conversiones elementales predicho por iSCc236. Las filas representan los compuestos importados y las columnas los componentes de la biomasa sintetizados por la bacteria. Los cuadros de color indican que compuestos importados son necesarios para la biosíntesis de cada componente de la biomasa. La glucosa y el oxígeno se excluyeron de la figura, ya que son requeridos en la síntesis de todos los componentes.

Un caso particular es el de la lisina ya que el genoma de *S. symbiotica* SCc codifica para todas las actividades de la ruta, con excepción del último paso que convierte el *meso*-diaminopimelato en lisina. Esto podría sugerir que la lisina es sintetizada como consecuencia de una complementación metabólica. Sin embargo, el *meso*-diaminopimelato es también necesario para la síntesis de petidoglicano, y como se verá el siguiente apartado *B. aphidicola* BCc conserva la ruta completa de síntesis de lisina. Esta evidencia pone en duda la complementación metabólica entre *S. symbiotica* SCc y el hospedador para la síntesis del aminoácido, previamente propuesta (Lamelas et al., 2011a). De acuerdo con lo anterior, *S. symbiotica* SCc exhibe una capacidad limitada para la

síntesis de aminoácidos. Sin embargo, las predicciones indican que el endosimbionte aún sería capaz de sintetizar los componentes de los ácidos nucleicos así como diversos cofactores. En particular, *iSCc236* predice que el organismo es capaz de sintetizar purinas a partir de la glutamina y la hipoxantina. Este resultado se encuentra en discordancia con los resultados previamente publicados que proponen que el organismo no sería capaz de sintetizar IMP (Lamelas et al., 2011a). Sin embargo, el genoma de *S. symbiotica* SCc incluye el gen (SCc_272) que codifica para la actividad hipoxantina fosforribo-transferasa (EC 2.4.2.8) y que permite la síntesis de IMP a partir de la hipoxantina y el PRPP (el cual puede ser sintetizado *de novo*, ya que el organismo conserva la ruta de las pentosas fosfato).

Además de los nucleótidos, el organismo codifica para diversas rutas metabólicas de síntesis de vitaminas y cofactores. En efecto, la capacidad de sintetizar riboflavina (vitamina B2) ha sido propuesta como el papel fundamental en el establecimiento de *S. symbiotica* SCc como un endosimbionte co-obligado en los áfidos de la familia *Lachninae* (Manzano-Marín and Latorre, 2014). Por otra parte, las predicciones de los requerimientos nutricionales sugieren que como consecuencia del proceso de reducción genómica, *S. symbiotica* SCc también ha establecido complementación metabólica con el hospedador en la ruta de síntesis del NAD⁺. El genoma de *S. symbiotica* SCc solo codifica las dos últimas actividades, por lo que el β-nicotinato D-ribonucleótido debe ser provisto por el hospedador para que el simbiote pueda completar la síntesis del NAD⁺. Este patrón es similar al caso identificado en *Blattabacterium* (ver apartado 4.3.4), sin embargo, *Blattabacterium* incluye una actividad adicional de la ruta por lo que es capaz de sintetizar el cofactor a partir del nicotinato en lugar de a partir de β-nicotinato D-ribonucleótido. Al comparar los requerimientos que *S. symbiotica* SCc presenta para la síntesis de los demás cofactores, respecto a las predicciones obtenidas para *Blattabacterium*, es notorio el hecho que ambos organismos tienen los mismos requerimientos tanto para la síntesis de la coenzima A y el sirohemo, los cuales son sintetizados a partir de (R)-pantotenato y del porfobilinógeno, respectivamente. Este resultado sugiere cierto grado de convergencia evolutiva en la complementación metabólica que existe entre los endosimbiontes y sus respectivos hospedadores.

Por su parte, la síntesis del tetrahidrofolato (THF) requiere del corismato como precursor y como consecuencia, éste había sido sugerido como una necesidad de *S. symbiotica* SCc (Lamelas et al., 2011a). Sin embargo, la bacteria conserva los dos últimos pasos de la ruta de síntesis del corismato (EC 2.5.1.19 y 4.2.3.5). Esto sugiere que el shiquimato, y no el corismato, debe ser importado para la síntesis del THF y sus derivados. Si se tiene en cuenta que los genomas disponibles de diversas *Serratia* de vida libre, codifican las seis actividades necesarias para la síntesis de corismato a partir del 3-dehidrokinolato (datos no mostrados), el hecho de que el genoma de *S. symbiotica* SCc sólo codifique para las últimas tres actividades y que *B. aphidicola* BCc sea capaz de sintetizar shiquimato, representa otro posible evento de complementación metabólica entre ambos endosimbionte, no descrito previamente. Por último, el modelo predice que *Serratia* es capaz de sintetizar biotina a partir del 8AONN, que como se ha propuesto, este podría estar siendo complementado por el metabolismo de *Buchnera* (Manzano-Marín et al., 2016). Desde el punto de vista de la energética *S. symbiotica* SCc presenta los genes codificantes para todas las actividades enzimáticas de la ruta glicolítica de manera que es capaz de degradar la glucosa o la fructosa para dar piruvato, 2 ATP y 2 NADH. Sin embargo, a diferencia con sus parientes más cercanos, ya sean

facultativos o de vida libre, *S. symbiotica* SCc ha perdido los genes que codifican las actividades del ciclo de Krebs. Como consecuencia, los dos moles de piruvato (por mol de glucosa) que se obtienen como subproducto de la glicólisis sólo pueden ser fermentados. La fermentación realizada a través la síntesis de acetyl-CoA y acetyl-fosfato, permite la fosforilación de dos moléculas de ATP adicionales y genera como producto dos moléculas de acetatos, que son excretadas. De esta forma, el rendimiento en la formación de ATP por fosforilación a nivel de sustrato es de 4 moléculas de ATP por glucosa consumida. Pero además de la fosforilación a nivel de sustrato, la fermentación de la glucosa produce equivalentes de reducción en forma de cuatro moles de NADH por mol de glucosa. Dado que el genoma de *S. symbiotica* SCc tiene los genes codificantes para la cadena respiratoria y la ATP sintasa, el organismo es capaz de emplear el NADH generado para sintetizar dos moléculas de ATP adicionales². De forma que al considerar la red completa, *iSCc236* predice que el rendimiento energético neto es de 6 moles de ATP por mol de glucosa. Este valor es considerablemente menor al rendimiento predicho por el modelo de *E. coli* K12 (*MG1655*) *iAF1260*, el cual tienen un valor de 23.5 moles de ATP por mol de glucosa (datos no mostrados). La gran diferencia se debe al hecho de que el metabolismo de *E. coli* incluye las actividades que forman ciclo de los ácidos tricarbónicos, y este le permiten oxidar el piruvato hasta CO₂ y emplear el NADH y el FADH₂ obtenidos para sintetizar ATP adicional, combinando la cadena de transporte de electrones con la ATP sintasa translocadora de protones.

Capacidades biosintéticas y energéticas predichas por *iBCc98*

En forma equivalente al análisis realizado en el apartado previo, en este se describen las capacidades metabólicas predichas por el modelo metabólico revisado de *B. aphidicola* *iBCc98*. La [figura 5.3](#) representa la relación entre los compuestos sintetizados por *B. aphidicola* BCc (columnas) y los compuestos que requiere para realizar dichas síntesis (filas), que corresponden a un subconjunto de las conversiones elementales de la red (ver [apéndice A](#)). Como se puede ver en la figura, *B. aphidicola* BCc es capaz de sintetizar (de forma no trivial) ocho aminoácidos. En el caso de la arginina, el endosimbionte requiere la ornitina que debe ser provista por el hospedador (u otra bacteria), a diferencia con la *Buchnera* de *Acyrtosiphon pisum*, que conserva dicha ruta íntegra. Además, *B. aphidicola* BCc también ha perdido la capacidad de sintetizar cisteína y alanina, posiblemente por ser ambos aminoácidos no-esenciales. El modelo también predice que el endosimbionte es capaz de sintetizar fenilalanina a partir del corismato mediante las actividades EC 4.2.1.51 y EC 5.4.99.5, codificadas por el gen BCc_244. Sin embargo, al igual que con los aminoácidos ramificados (*i.e.* valina, leucina e isoleucina) el genoma de *B. aphidicola* BCc ha perdido el gen que codifica la actividad transaminasa necesaria para completar dicha síntesis. Esta actividad es plausiblemente complementada por el hospedador o catalizada por alguna otra enzima que exhiba baja especificidad por el sustrato. Las predicciones también incluyen como requerimientos al antranilato y al 8AONN, intermediarios de la ruta de síntesis del triptófano y la biotina, respectivamente. Estos compuestos serían provistos por *Serratia* de manera que las rutas se encuentran complementadas entre ambos endosimbiontes, como ya fue comentado. Además, el

2 Las 4 moléculas de NADH generadas en la glicólisis son oxidadas para reducir cuatro quinonas. La oxidación de las últimas permite el bombeo de diez protones. Este gradiente de protones es balanceado de la siguiente forma: seis protones son bombeados desde el exterior, mediante la bomba ATP sintasa, dando lugar a la síntesis de dos moléculas de ATP. Los restante cuatro protones se compensan con los que son liberados en otras reacciones, *e.g.* la hidrólisis del ATP (algo similar a lo que ocurre en *B. aphidicola* BCc, como se verá en el siguiente apartado)

organismo carece de rutas de síntesis de cofactores y también de nucleótidos, de forma que dichos compuestos deben ser provistos o bien por el hospedador o bien por el otro miembro del consorcio, *i.e.* *S. symbiotica* SCc. Finalmente, *iBCc98* predice que *B. aphidicola* BCc solo sería capaz de sintetizar unos pocos lípidos de membrana, resultado ya descrito en estudios previos (Lamelas et al., 2011a).

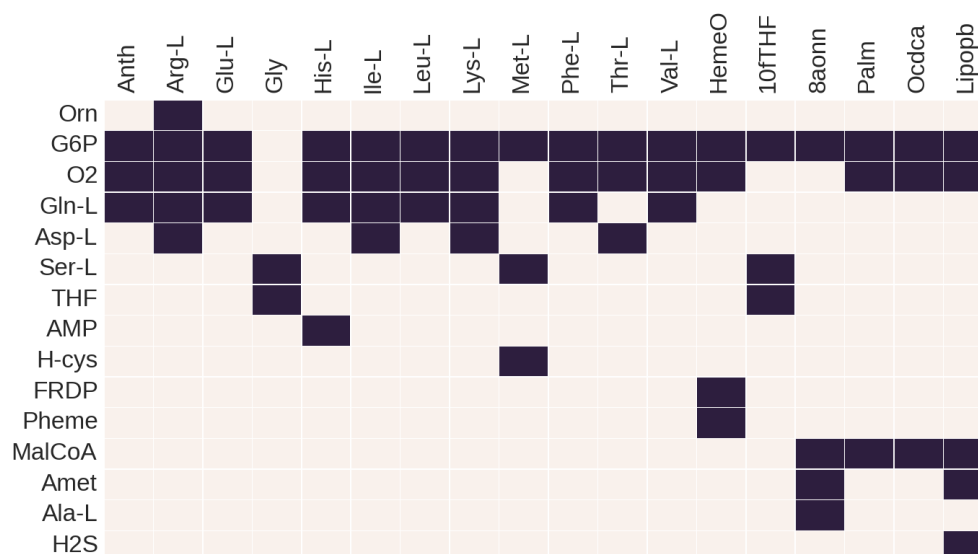
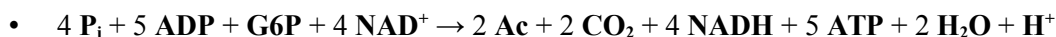


Figura 5.3. Mapa de conversiones elementales predicho por *iBCc98*. Las filas representan los compuestos importados y las columnas los componentes de la biomasa sintetizados por la bacteria. Los cuadros de color indican qué componentes importados son necesarios para la biosíntesis de cada componente de la biomasa.

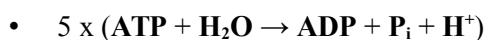
Desde el punto de vista del catabolismo, el genoma de *B. aphidicola* BCc conserva los genes de la ruta glicolítica, y al igual que *S. symbiotica* SCc, es capaz de fermentar la glucosa produciendo acetato, pero con un rendimiento de 5 moléculas de ATP en lugar de 4, ya que, de acuerdo con el modelo, debe importar la glucosa fosforilada³. No obstante, a diferencia con *S. symbiotica* SCc, este simbiote no ha conservado los genes que codifican para la ATP sintasa, de forma que el organismo solo es capaz de sintetizar ATP a través de la fosforilación a nivel de sustrato. Sin embargo, el genoma conserva todos los genes que codifican para los complejos NADH-deshidrogenasa y citocromo *bo*₃ (ubiquinol oxidasa). La presencia de dichos complejos en ausencia de un sistema ATP-sintasa es notable y las posibles explicaciones a este patrón no son ni intuitivas ni triviales. De acuerdo con las predicciones de *iBCc98*, dichos sistemas podrían ser empleados para “disipar” el poder reductor que se forma en la glicólisis u otros procesos. Sin embargo, esto plantea el siguiente problema de balances: por cada molécula de NADH que se oxida a NAD⁺ se expulsan 4 protones fuera de célula, *i.e.* que en el citosol de la bacteria se genera un consumo neto de protones. Este desequilibrio o gradiente es en general utilizado por la bomba ATP sintasa que permite acoplar el flujo de reentrada de protones a la síntesis de ATP. Pero dado que en el endosimbionte no hay ATP sintasa, los protones tienen que ser balanceados mediante algún mecanismo alternativo. De acuerdo con el modelo, la degradación de una molécula de glucosa-6-P

3 Dado que el genoma de *B. aphidicola* BCc no codifica ni para el sistema de transporte PTS ni para la actividad hexoquinasa, el modelo predice que el endosimbionte solo es capaz de degradar la glucosa-6-fosfato u otros azúcares fosforilados. De esta forma el ATP que se perdería al fosforilar la glucosa por una u otra vía, no se está contando en este caso y por este motivo el rendimiento es mayor.

(G6P) hasta acetato da como resultado neto la siguiente ecuación:



De forma que se genera un incremento de un protón, que debe ser balanceado, y cuatro moléculas de NADH que deben ser oxidadas. Al considerar la ecuación de la hidrólisis de ATP, que representa el consumo de energía de los distintos procesos celulares no incluidos en el modelo (*e.g.* replicación, transcripción, traducción):



El balance total, respecto al NADH y los protones, es de cuatro y seis, respectivamente, por molécula de G6P. Nótese que la ecuación está ponderada por cinco para que el balance global de ATP/ADP (al considerar la glicólisis y el gasto) sea cero. Si las cuatro moléculas de NADH son oxidadas a través de los complejos NADH-deshidrogenasa y citocromo *bo*₃:



El balance global de protones pasa a ser negativo, es decir que de alguna forma, se deben generar protones para que el balance sea cero. Al maximizar la síntesis de ATP empleando *iBCc98* y analizar la distribución de flujos óptima, las predicciones indican que solo una molécula de NADH sería oxidada a través de la cadena respiratoria (lo que deja un balance positivo de tres protones), mientras que las otras tres moléculas del cofactor serían oxidadas en el ciclo del tetrahidrofolato (EC 2.1.2.1, EC 1.5.1.20 y EC 2.1.1.13):



Además de oxidar el NADH balancea los tres protones restantes. Sin embargo, el ciclo de los folatos consume en forma neta homocisteína y serina para dar como productos glicina y metionina. Es decir, que de acuerdo con el modelo, la oxidación del NADH y el balance de los protones intracelulares, producidos como consecuencia de la fermentación de una molécula de G6P, implicaría el uso del ciclo de los folatos que conduciría al consumo de homocisteína y serina y la producción y excreción de glicina y metionina. Una forma sencilla de verificar este resultado es introducir en el modelo un flujo de reentrada de protones que permita compensar el desequilibrio que genera la actividad de la cadena respiratoria. Si bien la introducción de este flujo auxiliar carece de una justificación fisiológica apropiada, este permite verificar que tanto la reoxidación del NADH como el balance de los protones son la causa del alto flujo predicho para el ciclo de los folatos en la distribución de flujos que optimiza la producción de ATP (o biomasa).

La [figura 5.4](#) resume de forma esquemática el sistema con el que se regeneraría el NAD⁺ y los protones serían balanceados en el metabolismo de *B. aphidicola* BCc, de acuerdo con las predicciones de *iBCc98*. En dicho esquema, por cada dos moléculas de NADH que se oxidan, se consumen tres protones intracelulares y se convierten dos moléculas de homocisteína y dos serina en dos moléculas glicina y dos de metionina. De acuerdo con las predicciones de *iBCc98*, la producción de energía en *B. aphidicola* BCc se encontraría acoplada a la interconversión de homocisteína y serina en glicina y metionina. Como ya se ha mencionado, estas predicciones son contra-intuitivas, ya que plantean una forma poco ortodoxa de interpretar la función de la cadena respiratoria, que usualmente está asociada a la generación de una fuerza protón-motriz, que en

última instancia conduce a la síntesis de ATP. En cualquier caso, es posible que exista un mecanismo alternativo para la regeneración del NAD^+ y/o para balancear el gradiente de protones generado por la cadena respiratoria. Para explorar si el metabolismo de *Buchnera* dispone de mecanismos alternativos para regenerar el NAD^+ , a la vez que se mantiene el balance en los protones, se realizó el siguiente experimento: se limitaron a cero las entradas homocisteína y serina y con FBA se obtuvo la distribución de flujos que optimiza la producción de ATP. Bajo la nueva restricción, la solución produce ATP con un rendimiento de ATP por glucosa-6-fosfato importada de $\sim 1:1$. Curiosamente, la distribución de flujos produce como subproducto, un subconjunto de los nutrientes que el endosimbionte aporta al sistema. Estos incluyen a la valina, el 8AONN, y el antranilato. Este resultado puede sorprender a primera vista, ya que permite suponer que las capacidades energéticas del organismo se encuentran acopladas a la sobre-producción de compuestos de interés para todo el sistema. Sin embargo, al igual que con el ciclo de los folatos, si se relaja el balance de los protones, las sobre-producciones desaparecen y se recupera la solución en la que la glucosa es fermentada a acetato.

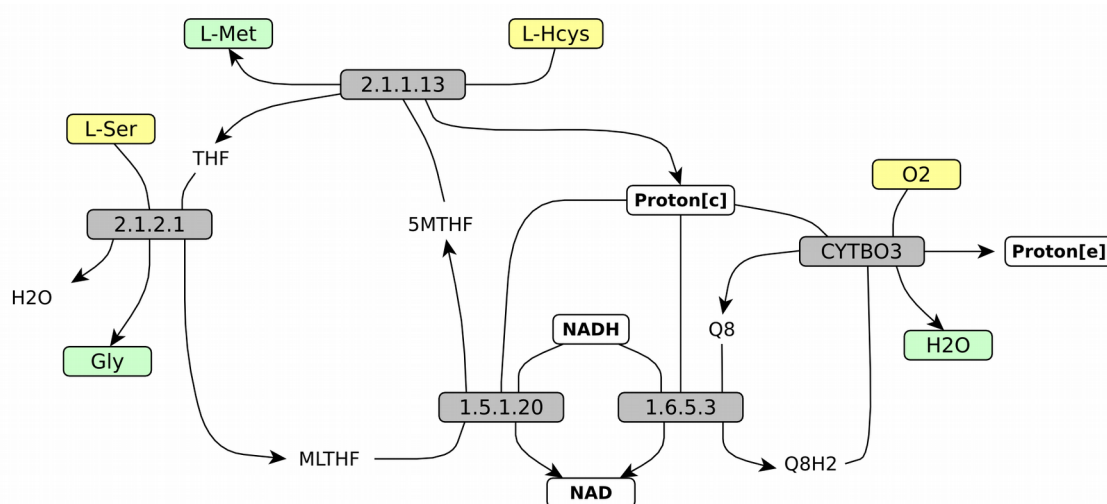


Figura 5.4. Esquema metabólico de regeneración de NADH y balance de protones predicho por *iBCc98*. El esquema incluye las actividades enzimáticas del ciclo de los folatos, el complejo NADH deshidrogenasa y la citocromo oxidasa. Los rectángulos amarillos y los verdes, representan los sustratos y productos, respectivamente. Los símbolos [c] y [e] en los protones denotan a los protones citosólicos y extracelulares, respectivamente. Los metabolitos sin indicación corresponden a metabolitos citosólicos.

Los resultados presentados plantean la posibilidad de que en el metabolismo de *B. aphidicola* BCc, el catabolismo y el anabolismo se encuentren fuertemente acoplado. Esto implica que la producción de energía en forma de moléculas de ATP conlleva la sobre-producción de nutrientes como aminoácidos, cofactores e incluso intermediarios de la complementación metabólica. Estos resultados deben ser considerados con cautela ya que la predicción podría representar un artefacto, consecuencia de la falta de información en la construcción del modelo. De manera que si existiese algún mecanismo alternativo para compensar el desequilibrio de protones generado por el uso de la cadena respiratoria, las sobre-producciones predichas por *iBCc98*, no necesariamente estarían acopladas a la síntesis de ATP. Por ejemplo, el gradiente de protones generado por la cadena respiratoria podría acoplarse para transportar compuestos en contra de gradiente, mediante sistemas de tipo simporte o antiporte. Sin embargo, la ausencia de transportadores codificados en el genoma

iBCc98 no apoyan esta conjetura. Además, si el mecanismo empleado por *Buchnera* para regenerar el NAD⁺ es a través del ciclo de los folatos, o mediante otras sobre-producciones, tanto la NADH-deshidrogenasa como el citocromo *bo₃* pasan a ser actividades enzimáticas residuales, e incluso prescindibles. No obstante, si un organismo con el grado de reducción genómica que presenta *B. aphidicola* BCc, conserva intactos los genes de la cadena respiratoria, estos deben tener un papel importante en la fisiología del organismo.

Por otra parte, los sistemas metabólicos de los endosimbiontes se encuentra muy reducidos y como consecuencia, es posible que su funcionamiento sea particularmente diferente al que se espera para un organismo de vida libre. Por ejemplo, la reconstrucción y el análisis metabólico del endosimbionte de la mosca blanca del tabaco, *Ca. Portiera aleyrodidarum*, sugieren que las rutas de generación de energía de este endosimbionte se encuentran acopladas a la síntesis de aminoácidos y carotenoides (Calle-Espinosa et al., 2016). En el caso de *B. aphidicola* BCc, ocurre algo equivalente, ya que como muestran las predicciones, ciertos procesos de biosíntesis se encuentren acoplados a la producción de energía. En ambos casos, el fenómeno es una consecuencia del tamaño reducido de los metabolismos de ambos endosimbiontes, consecuencia del proceso de reducción genómica al que se encuentran sujetos. En cualquier caso, en ausencia de resultados experimentales o información adicional que permita establecer la relación entre el desequilibrio de protones, los procesos de biosíntesis y producción de energía, es fundamental considerar ambos escenarios y comparar las predicciones obtenidas al considerar el balance estricto de los protones y cuando se relaja esta restricción, permitiendo la re-entrada de protones.

Fragilidad y robustez genética en los endosimbiontes del áfido del cedro

La fragilidad/robustez de las redes metabólicas de los endosimbiontes del áfido del cedro fue evaluada realizando un análisis de *knockouts in-silico* empleando en forma alternativa FBA y MOMA (ver apartado A.2.5 del apéndice A). Al analizar el metabolismo reconstruido de *B. aphidicola* BCc, el modelo *iBCc98* predice que un total de 73 de los 98 genes metabólicos (~75%) presentes en el modelo serían esenciales para el endosimbionte. Los 23 genes “dispensables”, codifican para un total de 7 actividades enzimáticas, todas involucradas en el catabolismo: cadena respiratoria (2), rutas de las pentosas fosfato (2), fermentación del piruvato (2) y glicólisis (1). Sin embargo, al realizar el análisis empleando MOMA el resultado indica que 93 de los genes (~95%) son esenciales para el organismo. En este nuevo escenario, los únicos cinco genes prescindibles codifican enzimas de la glicólisis (1), ruta de las pentosas fosfato (1), síntesis del 8AONN (1) y síntesis del triptófano (2)⁴. En este caso, los resultados no se ven afectados por el balance de protones, ya que al relajar la restricción se obtienen los mismos resultados.

En el caso de *S. symbiotica* SCc, de los 236 genes metabólicos incluidos en *iSCc236*, un total de 202 (~85%) son predichos como esenciales por el modelo al realizar la predicción empleando FBA. Sorprendentemente, el mismo resultado es obtenido si el cálculo se realiza empleando MOMA. En este organismo, el conjunto de 34 genes dispensables codifican para un total de 36 actividades enzimáticas involucradas en su mayoría, en rutas biosintéticas (e.g. biosíntesis de nucleótidos,

4 Si se tiene en cuenta que el triptófano que consume *Buchnera*, los sintetiza *Serratia*, que para sintetizar el aminoácido requiere del antranilato producido por *Buchnera*, se puede concluir que los dos genes que codifican para la antranilato sintasa son esenciales. El mismo argumento puede emplearse con la síntesis del 8AONN el cual es empleado por *Serratia* para sintetizar biotina.

cofactores), aunque también incluye genes involucrados en el catabolismo, como la ruta de las pentosas fosfato y la glicólisis. Si se comparan ambos organismos, el porcentaje de genes esenciales predichos por FBA, el hecho de que el porcentaje de genes esenciales sea menor en el caso de *B. aphidicola* BCc puede parecer, en un principio, paradójico ya que la red metabólica de *B. aphidicola* BCc es considerablemente más pequeña que la de *S. symbiotica* SCc. No obstante, si se considera el número de compuestos que deben ser importados del exterior por cada uno de los endosimbiontes, se ve que mientras que *S. symbiotica* SCc requiere unos 22 metabolitos (sin incluir los compuestos inorgánicos), *B. aphidicola* BCc requiere 29 metabolitos, entre en los que se encuentran los nucleótidos, y cofactores fundamentales como el NAD⁺ y la coenzima-A. Además, si se comparan las predicciones obtenidos con MOMA la situación se invierte. Por último, si no se consideran los genes de la ruta de síntesis de la pared celular presentes en *S. symbiotica* SCc (ya que los endosimbiontes los suelen perder), el número de genes esenciales se reduce a un total de 182 (~77%).

En forma adicional se realizó un estudio de la robustez distribuida mediante el análisis de letales sintéticos (o dobles letales), *i.e.* los pares de genes no esenciales que al ser inactivados en forma simultánea son predichos letales (Wagner, 2005). En el caso de *B. aphidicola* BCc, *iBCc98* predice que ~22% de las combinaciones de pares entre los 24 genes no esenciales, son letales sintéticos (o dobles letales). Por su parte, *iSCc236* predice que solo un ~5% de las combinaciones de pares posibles entre los 34 genes no esenciales, son letales en *S. symbiotica* SCc. Además, si el mismo cálculo se realiza sin considerar los genes de la pared celular, el porcentaje de letales sintéticos desciende al ~2%. El porcentaje de letales sintéticos en *S. symbiotica* SCc es sensiblemente menor al predicho para *B. aphidicola* BCc, lo que apoya la idea intuitiva de que el metabolismo del primero es más robusto que el del segundo debido a su mayor tamaños.

5.3.2 Definición de un modelo metabólico del consorcio endosimbionte

Existen diversos estudios con modelos metabólicos de un único endosimbionte (Belda et al., 2012; Calle-Espinosa et al., 2016; González-Domenech et al., 2012; Pál et al., 2006; Ponce-de-Leon et al., 2013; Russell et al., 2014; Thomas et al., 2009). En dichos estudios, el metabolismo del endosimbionte se ha modelado empleando un compartimento para representar el espacio intracelular del endosimbionte, y un compartimento extracelular que representa el citosol del bacteriocito, *i.e.* las células especializadas del hospedador en que se encuentran alojadas las bacterias. En este apartado se busca modelar el consorcio compuesto por los endosimbiontes primario y co-primario del áfido del cedro, *B. aphidicola* BCc y *S. symbiotica* SCc. Existen diversas aproximaciones para formular un modelo metabólico que incluya distintos organismos, siendo la diferencia principal, la forma en que se definen los distintos compartimentos (Taffs et al., 2009), Un aproximación bastante extendida es modelar cada organismo como un compartimento individual en forma análoga a la forma de representar los distintos orgánulos en modelos metabólicos de células eucariotas (Klitgord and Segrè, 2010a). Además, es necesario embeberlos en un compartimento que represente el medio extracelular compartido, a través del cual pueden intercambiarse distintos compuestos. En este punto, es importante hacer notar que un mismo metabolito que esté presente en distintos compartimentos es representado como metabolitos diferentes, de forma que el añadir compartimentos se aumenta la complejidad del modelo.

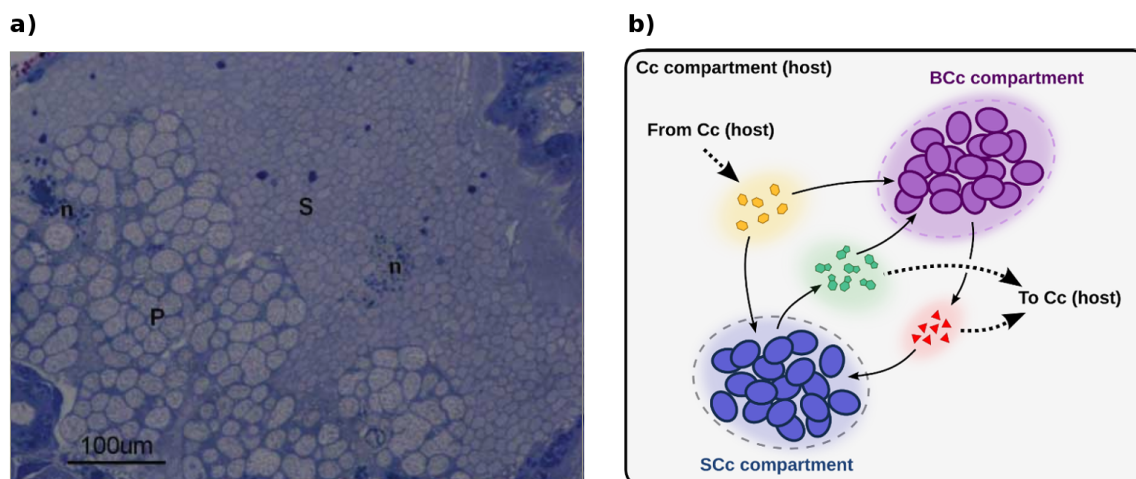


Figura 5.5. Consorcio endosimbionte del áfido del cedro. En a) se muestra una micrografía de los bacteriocitos del áfido *C. cedri*. P, simbionte primario (*B. aphidicola*); S, simbionte co-primario (*S. symbiotica*); n, núcleo del bacteriocito. Imagen extraída de (Pérez-Brocal et al., 2006). En b) se representa de forma esquemática, la definición de compartimentos adoptados en el modelo metabólico del consorcio endosimbionte del áfido del cedro. Las flechas sólidas representan los procesos de transporte que tienen lugar como consecuencia de la complementación metabólica o el cross-feeding. Las flechas punteadas representan flujos de intercambio así como los metabolitos que son consumidos por el hospedador.

En caso del áfido del cedro, *Buchnera* y *Serratia* no coexisten en los mismos bacteriocitos, sino que se encuentran segregadas en distintas regiones del órgano (ver figura 5.5a). Sin embargo, el tipo de célula en el que se encuentran alojadas ambas especies bacterianas, *i.e.* los bacteriocitos, forman parte de un mismo tejido u órgano (*i.e.* el bacterioma). De esta manera se puede suponer que el ambiente celular de los distintos bacteriocitos es equivalente, y que por su proximidad física estos pueden intercambiar distintos compuestos. Estas suposiciones permiten simplificar la formulación del modelo. De acuerdo con lo anterior, la reconstrucción del modelo metabólico del consorcio se llevó a cabo combinando los modelos metabólicos de los endosimbiontes primario y co-primario (*iBCc98* y *iSCc236*), en un modelo denominado *iBSCc*. El modelo incluye los siguientes tres compartimentos: i) un compartimento que representa a la población de *Buchnera*; ii) un compartimento que representa la población de *Serratia*; y iii) un único compartimento “extracelular” que representa al hospedador (*C. cedri*), y en el que se encuentran contenidos los dos anteriores (figura 5.5b). Como consecuencia de esta formulación, la frontera o borde del sistema queda definida por el compartimento que representa al hospedador. Por su parte, los flujos de intercambio que atraviesan dicha frontera representan los metabolitos que son provistos por el áfido, así como aquellos que este consume o excreta.

La descripción del modelo *iBSCc*, en término de compartimentos se describe en la tabla 5.5. El conjunto de metabolitos presentes en el compartimento extracelular es la unión de los metabolitos presentes en los compartimentos extracelulares de los modelos de los simbiontes individuales *iBCc98* (63) y *iSCc236* (79) y son un total de 86. Los metabolitos compartidos por ambos modelos son un total de 52, e incluye a los metabolitos que ambos organismos obtienen del hospedador, los productos de excreción comunes y los metabolitos que son intercambiados entre los endosimbiontes, *i.e.* los puntos de intercambio de las complementaciones. Los metabolitos del compartimento extracelular que son únicamente consumido/producido por uno de los

endosimbiontes se corresponden con requerimientos o productos de excreción específicos de cada simbiote.

Tabla 5.5. Descripción del modelo iBSCc en término de los metabolitos presentes en cada compartimento y los intercambios posibles. La tabla describe el modelo en relación con sus compartimentos (ver figura 5.5b) y los metabolitos presentes en cada uno y que pueden ser intercambiados entre estos.

Elemento / Compartimento	SCc (<i>Serratia</i>)	BCc (<i>Buchnera</i>)	Cc (extracelular)
Metabolitos importados	34	36	26
Metabolitos exportados	34	20	57
Metabolitos libres	12	7	3
Metabolitos total	264	142	86

En forma adicional, se añadieron cinco reacciones al compartimento extracelular. Estas reacciones fueron incluidas porque se ha propuesto que son realizadas por el hospedador y juegan un papel relevante en la complementación metabólica entre los endosimbiontes y su hospedador (Hansen and Moran, 2011; Poliakov et al., 2011). Las reacciones añadidas incluyen: 1) la conversión de fenilalanina en tirosina; 2-4) la producción de homocisteína y adenosil-metionina a partir de la cisteína; y 5) una reacción auxiliar que presenta la asimilación del ácido sulfhídrico, producido por *Serratia*, para la formación de lipoato. Todos los detalles sobre el modelo metabólico de consorcio iBSCc se encuentra en la tabla S.5.3 del apéndice B

Definiendo una función objetivo para el consorcio de endosimbiontes

Una vez que se ha definido el modelo metabólico de consorcio es posible analizar los patrones de flujos que surgen como consecuencia de la complementación metabólica. Dado que la complementación metabólica es una consecuencia de un largo proceso co-evolutivo entre diversos organismos (~200ma), una pregunta que surge es la siguiente: ¿los patrones de complementación metabólica observados presentan algún tipo de “diseño” característico que sea consecuencia de la optimización de alguna función? Empleando herramientas de optimización es posible analizar las capacidades metabólicas de la red observadas. A su vez, se pueden construir redes con diseños alternativos y comparar sus capacidades, respecto a la red original. No obstante, para poder llevar adelante esta clase de experimentos *in-silico*, es preciso definir una función objetivo (o de coste) que brinde un criterio para poder comparar distintas redes o diseños. Un problema que surge al abordar el análisis del consorcio es cómo definir una función objetivo que integre a los tres miembros de la asociación simbiótica. Si bien se han desarrollado estrategias para el análisis de comunidades microbianas que permiten predecir distintos tipos de interacciones (Stolyar et al., 2007; Taffs et al., 2009; Vallino, 2003; Zomorodi and Maranas, 2012), el caso de un consorcio endosimbionte es sumamente particular. En primer lugar, el proceso co-evolutivo al que han estado sometido los miembros de un consorcio endosimbionte se ha extendido en un tiempo que supera los 100 millones de años (Pérez-Brocal et al., 2006). Este proceso ha dado lugar a la emergencia de complementación metabólica que a su vez conduce a una interdependencia, en principio, obligada entre los miembros de la asociación (*i.e.* ninguno de los miembros puede prescindir del resto)⁵. Estas características suponen que las tasas de crecimientos de los miembros de un consorcio se

5 A escala evolutiva existe la posibilidad de que un endosimbionte sea remplazado por otro.

encuentren estrechamente relacionadas. Concretamente, en el caso del sistema conformado por *B. aphidicola* BSg y su hospedador el áfido *Schizaphis graminum*, sistema similar al consorcio de *C. cedri*, existen resultados experimentales que muestran como las dinámicas de crecimiento del insecto y el endosimbionte se encuentra estrechamente relacionadas (Baumann and Baumann, 1994; Baumann et al., 1996).

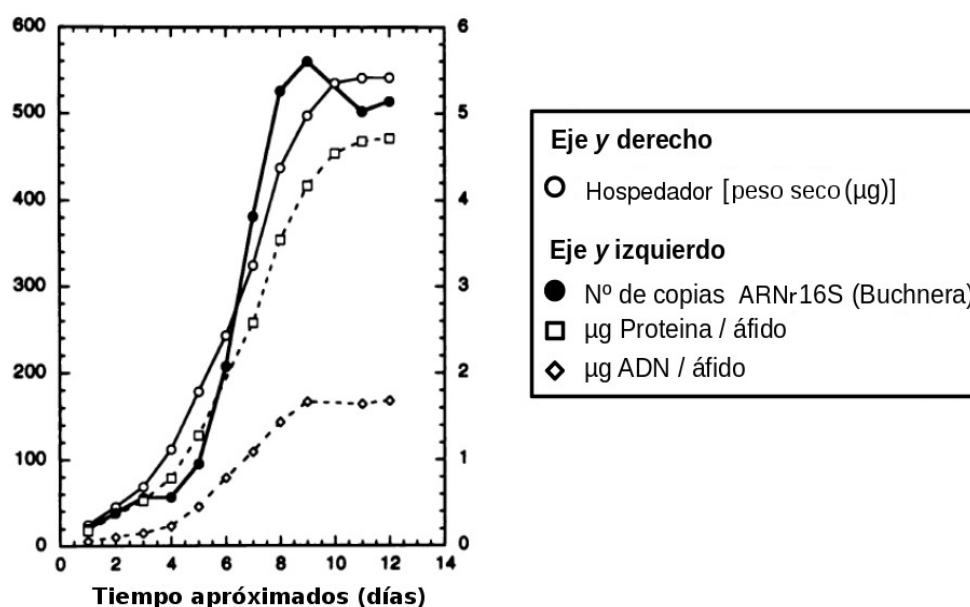


Figura 5.6. Cinética del crecimiento de *S. graminum* (áfido) y *B. aphidicola* BSg (endosimbionte). El eje de las *x* representa el tiempo aproximado en días. La leyenda de la figura se encuentra en el cuadro de la derecha. Modificada de Baumann y Baumann (1994).

La figura 5.6 representa distintas curvas obtenidas por Baumann y Baumann (1994) en mediciones de variables como la masa en peso seco de los áfidos, el número de individuos, así como del número total de copias de RNA ribosomal 16S del simbionte *B. aphidicola* BSg. Los resultados experimentales representados en la figura 5.6 muestran una relación estrecha entre el crecimiento de la masa del áfido y el incremento en el número de copias del gen rRNA 16S. De manera que si se supone que el número de copias del gen rRNA 16S permite estimar el número de simbiontes, esta medida se puede comparar con la masa del áfido durante la fase de crecimiento. En efecto, el test de correlación de Pearson indica que la correlación entre ambas variables es de ~ 0.97 con un *p*-valor de 10^{-7} (ver tabla S.5.4 en el apéndice B). Si bien el número de copias del gen rRNA 16S puede presentar problemas para la estimación de abundancias en comunidades microbianas, debido a que el número de copias del gen puede variar entre los genomas de distintas especies (Větrovský and Baldrian, 2013), en el estudio descrito sólo se analiza la abundancia de un especie bacteriana, cuyo genoma sólo codifica para un única copia del gen rRNA 16S. Estudios de la estimación del número de endosimbiontes realizado por recuento manual de individuos en distintas secciones de muestras del insecto, observadas mediante microscopía electrónica, han permitido estimar la población de bacterias de forma directa (Mira and Moran, 2002). Los tamaños poblacionales medidos por Mira y Moran concuerdan con los valores previamente estimados por métodos moleculares (Baumann and Baumann, 1994), así como con los de otros estudios similares (Humphreys and Douglas, 1997). Esto indica, que el número de copias del gen rRNA 16S puede

emplearse como estimador del número de individuos. Además, un estudio reciente realizado empleando citometría de flujo ha permitido estimar los tamaños poblacionales de *Buchnera* en distintos estadios de crecimiento del áfido *A. pisum*. Estos resultados son también consistentes con los estudios antes citados (Simonet et al., 2016). Simonet y colaboradores muestran que el endosimbionte crece de forma exponencial durante el desarrollo del insecto.

De acuerdo con los resultados de los estudios expuestos, parece razonable suponer que el número de copias de rRNA ribosomal 16S se encuentra relacionado con el número de individuos, o la biomasa total de endosimbiontes. Combinando estos resultados con los datos sobre la dinámica de crecimiento del áfido *S. graminum* y su endosimbionte *B. aphidicola* BSG, previamente obtenidos (Baumann and Baumann, 1994; Baumann et al., 1996), se puede deducir que la relación de masas entre hospedador y simbiote se mantiene relativamente constante durante los primeros estadios de crecimiento del áfido. Por otra parte, Kerner y colaboradores construyeron un sistema experimental de *crossfeeders* sintéticos, empleando dos cepas *E. coli* auxótrofas para distintos aminoácidos (Kerner et al., 2012). Los resultados obtenidos en dicho trabajo indican que luego de un cierto tiempo, las tasas de crecimiento de ambas cepas, adoptan el mismo valor. En base a estos resultados, los autores desarrollaron un modelo dinámico (basado en ecuaciones diferenciales) que muestra cómo la interdependencia nutricional mutua y obligada entre organismos *cross-feeders* conduce a que las tasas de crecimiento de ambos organismos adopten el mismo valor. Además, a partir de las ecuaciones se pueden deducir dos expresiones analíticas que relacionan por un lado la tasa de crecimiento y por otro la relación entre el número de individuos de cada cepa, con las tasas de producción de los compuestos intercambiados (*i.e.* aminoácidos) y la estequiometría del compuesto requerido, en la biomasa de cada cepa (Kerner et al., 2012). Si bien el sistema de *cross-feeders* sintéticos es muy diferente a un consorcio endosimbionte, la dependencia nutricional obligada es una característica común a ambos.

De acuerdo con los datos experimentales disponibles y los argumentos antes expuestos, en lo que sigue se supondrá que las tasas de crecimiento de los endosimbiontes y el hospedador se encuentran acopladas y que la relación de biomasa entre los tres organismos permanece constante durante la fase de desarrollo del áfido en estadio de ninfa hasta alcanzar el estado adulto (días 2 al día 8). En base a lo anterior, la definición de una función objetivo que permita calcular las capacidades metabólicas del consorcio durante la fase de crecimiento exponencial de los simbiotes, puede realizarse empleando la misma idea con la que se define la ecuación de biomasa en el modelo metabólico de una bacteria de vida libre (ver apartado A.1.4 del apéndice A). Si se considera que la biomasa del insecto está formada por tres componentes que son *Buchnera*, *Serratia* y el áfido y se conoce la estequiometría, o la relación de masas entre los tres organismos, se puede introducir una reacción artificial o ecuación de meta-biomasa que represente el crecimiento de todo el sistema, en forma similar al modelo propuesto por Stolyar y colaboradores (2007). Esto requiere la introducción de una ecuación de biomasa que represente el crecimiento del huésped. La formulación de dicha ecuación fue realizada en forma similar a como ha sido propuesto por Calle-Espinosa et al. (2016). De esta forma, dicha ecuación solo incluye el consumo de los aminoácidos provistos por los endosimbiontes, en proporciones basadas en dos conjuntos de medidas experimentales, previamente realizadas (Douglas et al., 2001; Russell et al., 2014). Si bien los valores experimentales empleados fueron medidos para dos especies de áfido diferentes a *C.*

cedri, las medidas obtenidas en ambos trabajos presentan una correlación de Pearson $r = 0,85$ con un p -valor de 10^{-6} , lo que indica que la composición en aminoácidos entre distintas especies de áfidos no exhibe gran variación (ver apartado 5.2.3 para más detalles).

Un vez que cada miembro del sistema tiene definida su ecuación de crecimiento, es necesario incluir un metabolito auxiliar en cada una de las ecuaciones que representen la biomasa que se forma durante el crecimiento. Si bien, la relación de masas entre insecto y simbiontes cambia una vez que el insecto alcanza la adultez (Simonet et al., 2016), la relación parece mantenerse relativamente constante, al menos durante los primeros días de la fase de crecimiento (Baumann and Baumann, 1994; Baumann et al., 1996). Dado que una vez que el áfido alcanza la madurez, la población de endosimbiontes tiene a reducirse, e incluso, en algunos casos como en el caso de los gorgojos, estos son reciclados mediante apoptosis y autofagia de los bacteriocitos (Vigneron et al., 2014), los efectos de la complementación metabólica entre los endosimbiontes y el hospedador deben ser más relevantes durante el crecimiento del sistema que cuando éste estaciona. Por este motivo, la introducción de una ecuación de meta-biomasa parece justificada para representar y analizar el crecimiento del sistema en las primeras fases de desarrollo del áfido. Entonces, teniendo las ecuaciones de crecimiento de cada uno de los organismos y empleando los metabolitos auxiliares que representan la biomasa de cada miembro, estos “metabolitos biomasa” son introducidos como sustratos de la ecuación de meta-biomasa.

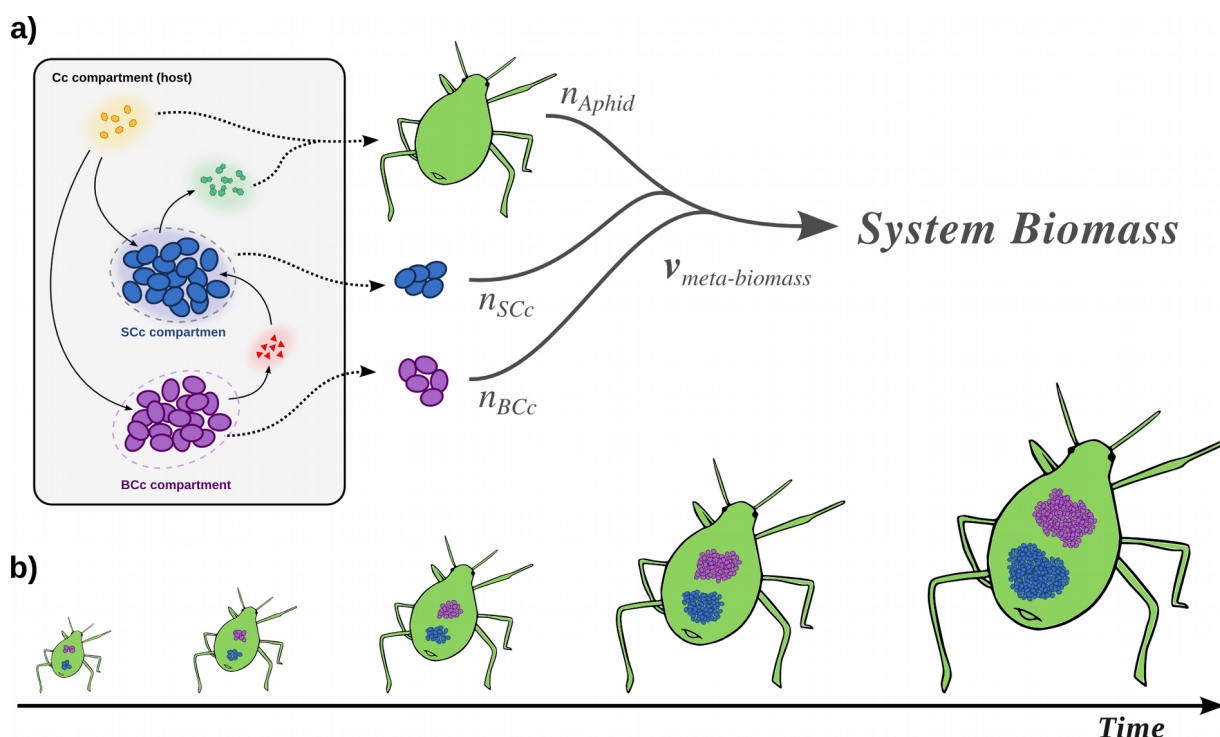


Figura 5.7. Definición de una ecuación de meta-biomasa para modelar el crecimiento de un consorcio endosimbionte y su hospedador. En a) se representa en forma esquemática el modelo con tres compartimentos (miembros) y sus ecuaciones de biomasa respectivas (flechas punteadas). La flecha que conecta las biomásas de los tres miembros representa la meta-biomasa o crecimiento del sistema, y las anotaciones $n_{\%alpha} \in \{BCc, SCc, Aphid\}$ representan la estequiometría, o relación de masas, de cada miembro en relación al sistema completo. En b) se representa el crecimiento acoplado del sistema y muestra como se mantendría la relación de masas entre los tres miembros como consecuencia del acoplamiento.

Como muestra la [figura 5.7](#), la ecuación de meta-biomasa es una reacción ficticia que consume los metabolitos auxiliares que representan la biomasa de cada miembro del consorcio en proporciones estequiométricas fijas. Como consecuencia de esta estequiometría, las tasas de crecimiento de cada miembro individual quedan acopladas. Si bien no se dispone de datos sobre la relación entre la biomasa de los simbiontes y su hospedador el áfido del cedro, existen datos para el sistema de *S. graminum* y *A. pisum* que indican que los endosimbiontes presentan entre el 5-15% de la biomasa total del sistema (Baumann et al., 2006; Whitehead and Douglas, 1993). En el presente estudio, dichos valores fueron empleados como referencia para establecer la estequiometría o relación de masas entre el insecto y sus endosimbiontes. Concretamente, y por simplicidad, se supondrá una relación entre la masa de los endosimbiontes y la del insecto normalizada a 100, de forma que la masa de endosimbiontes y el insecto representan el 10% y el 90% de la masa total, respectivamente. Dado que no se ha encontrado información que indique la relación de tamaños entre las poblaciones de ambos endosimbiontes, se supondrá que éstas se encuentran en la misma proporción y como consecuencia, se supondrá que la masa de cada población individual de endosimbiontes representa un 5% de la masa total del sistema. Esta relación es empleada para definir los coeficientes estequiométricos en la ecuación de meta-biomasa como sigue: $n_{BCc}=n_{SCc}=0,05$ y $n_{Aphid}=0,9$ (ver [figura 5.7](#)). Una vez introducida la ecuación de meta-biomasa, se verificó mediante FBA que el modelo predice un flujo $v_{\text{meta-biomasa}} > 0$.

5.3.3 Definición de una condición de contorno

Antes de poder realizar simulaciones con el modelo de consorcio es preciso definir una condición de borde del sistema. Esto implica definir las cotas de los flujos de intercambios que representan las tasas máximas y mínimas de consumo y producción de nutrientes (ver [apartado A.1.3](#) del apéndice A). Como se vio en el [apartado 5.3.1](#) los modelos individuales de los endosimbiontes predicen que cada organismo debe importar un número elevado de compuestos, lo que representa un gran número de parámetros⁶. Además, es de interés poder controlar tales parámetros de forma independiente para cada endosimbionte. Si bien para esta clase de sistemas no se dispone de medidas experimentales que puedan ser empleadas como referencia, es posible definir una condición de borde que permita reducir el número de parámetros. En concreto se empleó una aproximación similar a la propuesta Burgard *et al.* (2001) para realizar identificar redes metabólicas mínimas viables, en ambientes ricos en nutrientes. La aproximación consiste en introducir una restricción adicional (por cada endosimbionte) que limite el flujo total de átomos de carbono intercambiados por cada endosimbionte (ver el [apartado 5.2.4](#) de la sección Materiales y Métodos). De esta forma, el número de parámetros que representan las cotas de intercambio se reduce únicamente a dos, *i.e.* el transporte máximo entre cada endosimbionte y el espacio o compartimento extracelular. Además, con la introducción de estas restricciones se normalizan las distribuciones de flujos en patrones cualitativos relativos a los parámetros de control (*i.e.* las cotas en los flujos de átomos de carbono que cada endosimbionte pueden intercambiar con el exterior). Desde un punto de vista biofísico se ha visto que en determinadas condiciones de crecimiento la superficie total de la célula pueda imponer un restricción en la carga máxima de moléculas que se

6 Los modelos predicen que *Serratia* y *Buchnera* deben importar un total de 22 y 30 compuestos orgánicos, respectivamente. Esto es un número muy elevado en comparación, por ejemplo, con una bacteria libre como *E. coli*, que puede crecer empleado la glucosa como única fuente de carbono y energía, además de fuentes inorgánicas de nitrógeno, azufre y fósforo.

pueden intercambiar por unidad de tiempo (Zhuang et al., 2011). En el caso de los endosimbiontes, dado que estos deben importar y exportar un gran número de compuestos es posible que las restricciones de transporte introducidas representen, en cierta medida, un límite real del sistema (además de reducir el número de parámetros). En lo que sigue, las cotas de transporte de cada endosimbionte se fijarán en 100 unidades de flujo arbitrarias (UFA), en caso contrario se mencionará el valor de forma explícita. Por último, realizando un FVA se ajustaron las direcciones de los de intercambio. Además se eliminaron aquellos intercambios que al considerar el modelo del consorcio ya no son necesarios (e.g. antranilato, 8AONN).

5.3.4 Análisis de patrones de crecimiento óptimo

Para validar la reconstrucción del modelo de consorcio, así como la definición de las restricciones de transporte descritas en el apartado previo, se calculó la distribución de flujos que maximiza la producción de meta-biomasa del sistema, aplicando FBA. El valor del óptimo obtenido de esa manera fue mayor a cero ($v_{\text{Meta-Biomasa}}^* = 0.37 \text{ UFA}$), lo que indica que la definición del modelo de consorcio combinado con la ecuación de meta-biomasa y las restricciones de transporte, permiten simular el crecimiento del sistema. Empleando la distribución de flujos predicha con FBA, se analizaron los flujos de transporte entre los miembros del consorcio, para identificar los eventos de *cross-feeding* y complementación metabólica, predichos por el modelo. De acuerdo con lo anterior, las predicciones obtenidas con FBA indican que *B. aphidicola* BCc sintetiza y provee al hospedador y a *S. symbiotica* SCc diez aminoácidos; *B. aphidicola* BCc también sintetiza antranilato, shiquimato y 8AONN, que son exportados a *S. symbiotica* SCc, que los emplea en la biosíntesis de triptófano, THF y biotina, respectivamente (productos que son a su vez compartidos con los demás miembros del consorcio). De esta manera, el modelo predice que habría tres eventos de complementación metabólica, teniendo lugar entre ambos endosimbiontes. Tanto el caso de la complementación en la ruta del triptófano, así como en la de la biotina, fueron inferidos en estudios previos, basados en análisis de los genomas (Gosalbes et al., 2008; Manzano-Marín et al., 2016). Sin embargo, el caso del shiquimato representa un evento de complementación metabólica plausible, no descrito previamente como tal; aunque sí se ha descrito *cross-feeding* de folato entre *Serratia grimesii* y *Treponema primitia* AS, ambos miembros de la microbiota intestinal de la termita *Zootermopsis angusticollis* (Graber and Breznak, 2005). En el caso de *C. cedri*, dado que los insectos y en general lo animales no son capaces de sintetizar *de novo* el folato y este no esta presente en la dieta de los pulgones, se espera que *Serratia* provea al áfido y a *Buchnera* de dicho cofactor.

Con respecto al papel de *S. symbiotica* SCc en el consorcio, las predicciones indican que este sintetiza *de novo*, i.e. sin la necesidad de complementación: cisteína, los cuatro desoxinucleótidos; los cuatro nucleósidos trifosfato, y trece cofactores y coenzimas. Muchos de estos compuestos son compartidos con *Buchnera* y/o con el hospedador. No obstante, el modelo de consorcio no incluye el metabolismo del áfido que posiblemente sea capaz de sintetizar algunos de estos compuestos, e.g. los desoxinucleótidos y los nucleósidos trifosfato. Si bien no se dispone de evidencia experimental que permita definir quién provee a *B. aphidicola* BCc de nucleótidos y de cofactores como el NAD⁺, parece más probable que dicha función sea realizada por los bacteriocitos que contienen a *Buchnera*. En efecto, si bien el genoma del áfido del cedro no ha sido secuenciado, hay disponible un genoma secuenciado y anotado para el caso de *A. pisum* (Consortium, 2010). Este, al

igual que la mayoría los eucariotas los áfidos incluyen las rutas de síntesis de purinas y pirimidinas (Vellozo et al., 2011), indicando que *C. cedri* también debería de tenerlas. Esto, vuelve notorio el hecho de que *S. symbiotica* SCc conserve intactas las rutas de síntesis de purinas y pirimidinas, mientras que *B. aphidicola* BCc las ha perdido por completo. A pesar de lo antes expuesto, en el presente estudio se supondrá que es *S. symbiotica* SCc quién provee a *Buchnera* de nucleótidos y cofactores y al áfido de cofactores como la biotina, la riboflavina y el TFH (Manzano-Marín et al., 2016). Como ya fue discutido al realizar el análisis de los modelos individuales de cada endosimbionte, ambos requieren oxígeno y excretan diversos compuestos, que incluyen productos de fermentación como el acetato, el formato y el succinato, así como intermediarios de rutas truncadas, como el caso de AICAR. Este último es muy posiblemente reciclado por el bacteriocito en la síntesis de purinas, como ya ha sido propuesto (Calle-Espinosa et al., 2016; Lamelas et al., 2011a; Zientz et al., 2004).

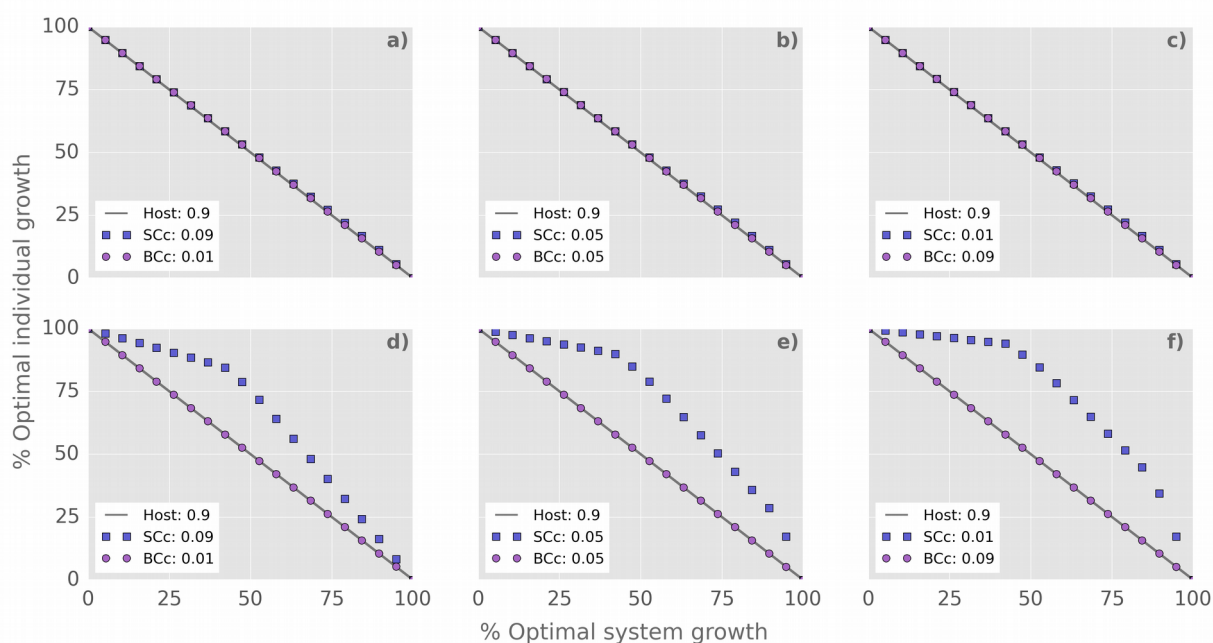


Figura 5.8. Perfiles de crecimiento de los endosimbiontes para distintas estequiometrías de la meta-biomasa. Las curvas representan la variabilidad en la tasa de crecimiento individual de cada miembro del consorcio al exigir distintos porcentajes de crecimiento óptimo del sistema. En cada panel se emplean distintos valores de estequiometría de la meta-biomasa (ver leyenda). De a-c y en d-f se representan los experimentos exigiendo el balance estricto de protones en BCc y relajando dicha restricción, respectivamente.

Si bien la relación de masas entre simbioses y hospedadores parece mantenerse relativamente constante, al menos durante las primeras fases de crecimiento del áfido, la utilización de la ecuación de meta-biomasa impone una restricción muy fuerte al sistema ya que implica acoplar las tasas de crecimiento de los miembros de consorcio con una estequiometría fija. Sin embargo, la relación de masas entre los miembros podría diferir de los valores empleados en este estudio. Para evaluar si los patrones de complementación antes descritos se ven afectados por cambios en la estequiometría de la meta-biomasa se realizó un análisis de sensibilidad, variando los coeficientes de la meta-biomasa. Para dichos experimentos se plantearon tres esquemas de estequiometría diferentes. Además los experimentos fueron realizados teniendo en cuenta el balance estricto de los

protones en el compartimento correspondiente a BCc, y relajando dicha restricción, ya que como fue descrito en el [apartado 5.3.1](#), el metabolismo energético predicho por *iBCc98* se encuentra extremadamente limitado debido a la falta de una ATPasa translocadora de protones que permita la reentrada de protones a la célula. Para cada uno de los seis escenarios se analizó la variabilidad en las tasas de crecimiento óptimas que cada organismo individual podría alcanzar, a la vez que se exige que el crecimiento total del sistema completo (*i.e.* la meta-biomasa) alcance un mínimo porcentaje del valor óptimo predicho con FBA. Por ejemplo, si se exige que la meta-biomasa o el crecimiento del sistema alcance el 50% de su máximo, ¿sería posible que alguno de los endosimbionte pueda aumentar su tasa de crecimiento? Con este experimento se busca evaluar hasta qué punto el crecimiento acoplado limita los crecimientos de los endosimbiontes individuales y qué efecto tiene la estequiometría del acoplamiento en este fenómeno. Los resultados del experimento se muestran en la [figura 5.8](#).

En primer lugar los resultados muestran que si se considera la restricción que impone el balance de protones en BCc, la tasa de crecimiento individual de cada miembro del sistema decrece en forma lineal con respecto a la tasa de crecimiento del sistema completo, es decir que al exigir un mayor valor en el flujo del meta-biomasa. Esto implica que en la medida que el sistema hospedador-consorcio se acerca a su tasa de crecimiento óptimo, los miembros individuales van perdiendo flexibilidad en su capacidad de crecer por encima de dicho valor. Este fenómeno es independiente de la estequiometría de la meta-biomasa. Sin embargo, si se relaja la restricción sobre el balance de protones en BCc, los resultados muestran de forma clara, que si el sistema hospedador-consorcio crece por debajo de su óptimo, *S. symbiotica* SCc, podría crecer más rápido que los demás miembros del sistema. Como consecuencia, si por restricciones cinéticas o regulatorias el sistema hospedador-consorcio no puede alcanzar la tasa de crecimiento máxima predicha por FBA, *S. symbiotica* SCc es el único miembro del consorcio que sería capaz de incrementar su propia tasa de crecimiento por el límite que le impone la estequiometría. Este resultado podría deberse al hecho de que *S. symbiotica* SCc se encuentra más cerca de un organismo de vida libre, pese a haberse convertido en un endosimbionte obligado. Por otra parte, si se consideran los patrones de crecimiento óptimo y se fija la atención en qué nutrientes y en qué proporciones son aportados por cada endosimbionte, se ve de forma clara que *B. aphidicola* BCc tiene un papel más importante en el consorcio que su co-simbionte, lo que refleja el mayor tiempo que lleva co-evolucionando con el hospedador. Si bien, los compuestos aportados por *S. symbiotica* SCc son también esenciales para el sistema y por eso es considerado un endosimbionte co-primario, el coste que esto le representa es considerablemente menor con respecto a lo aportado por *B. aphidicola* BCc, al menos en términos estequiométricos. Sin embargo, el hecho de que ciertas rutas metabólicas esenciales se encuentren complementadas entre ambos endosimbiontes sugiere que *S. symbiotica* SCc se ha ido integrando en la simbiosis, aunque en un grado considerablemente menor que su compañera *B. aphidicola* BCc.

5.3.5 Experimentos *in-silico* de reducción: el caso del triptófano

Los insectos, al igual que el resto de los animales carecen de ruta de síntesis del triptófano, y como consecuencia este es un aminoácido esencial para los mismos. En diversos insectos muchos de los aminoácidos esenciales son provistos por bacterias endosimbiontes. Por ejemplo, en las cucarachas el triptófano es producido por el endosimbionte *Blattabacterium* (ver [apartado 4.3.4](#)).

En el caso de los áfidos en general esta función la desempeña *B. aphidicola* (ver figura 5.9a) (Martínez-Cano et al., 2015; Rouhbakhsh et al., 1996). Sin embargo, como se ha discutido a lo largo de este capítulo, en el caso del áfido del cedro la ruta de síntesis del triptófano se encuentra complementada entre los endosimbiontes *B. aphidicola* BCc y *S. symbiotica* SCc (ver figura 5.9b).

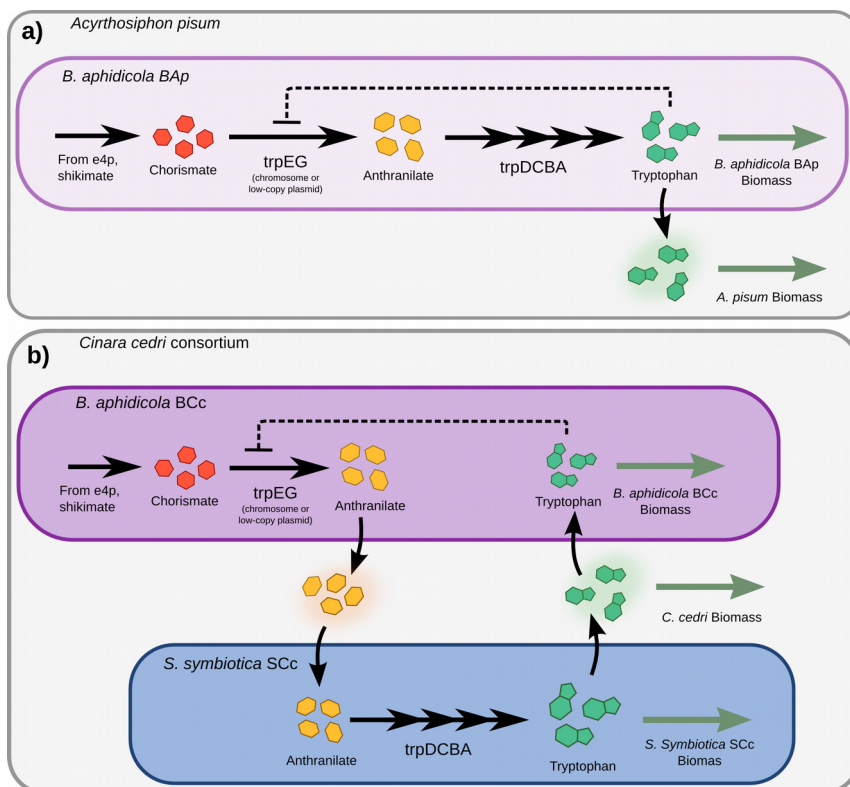


Figura 5.9 Estructura de la ruta de biosíntesis del triptófano en dos sistemas endosimbionte-hospedador diferentes. En a) se representa la estructura de la ruta presente en la cepa de *Buchnera* presente en el áfido *A. pisum*. En b) se representa la complementación metabólica de la ruta del triptófano entre *B. aphidicola* BCc y *S. symbiotica* SCc en el áfido *C. cedri*: los genes *trpEG* solo se encuentran presentes en *Buchnera*, mientras que los genes *trpDCBA* solo son codificados por *Serratia*.

La ruta biosintética del triptófano tiene como sustrato inicial el corismato e involucra 5 actividades enzimáticas que se encuentran altamente conservadas en diversos genomas microbianos. Sin embargo, los genes que codifican dichas actividades exhiben variaciones en el número y en la organización a nivel del genoma, debido en buena medida a los numerosos eventos de fusión y división que han ocurrido a lo largo de su historia evolutiva (Priya et al., 2014). Los genes de la ruta, cuyo número total varía de 5 a 7 (*trpA-G*) suelen encontrarse agrupados en un operón cuya transcripción se encuentra, por lo general, regulada negativamente por el triptófano (Crawford, 1989). Además de los mecanismos de regulación y atenuación de la transcripción, también se han descrito mecanismos de regulación traduccionales y post-traduccionales. La presencia de múltiples niveles de regulación a los que se encuentra sometida la síntesis del triptófano posiblemente está asociada al alto costo que representa la síntesis de dicho aminoácido. En efecto, estimaciones experimentales indican que el coste por mol de triptófano producido es de ~78 moles de ATP (Yanofsky, 1981). En efecto, en un estudio más reciente en el que se comparan diversas estimaciones del coste energético de la síntesis de triptófano se encuentra un consenso claro en que este es el aminoácido más costoso, posiblemente debido a la compleja estructura de

doble anillo y al elevado número de moléculas costosas que involucra su síntesis (Barton et al., 2010). Además, simulaciones realizadas empleando una reconstrucción a escala genómica de *E. coli* predicen que la síntesis de triptófano tiene el rendimiento más bajo de todos los aminoácidos (Kaleta et al., 2013).

La mayoría de los genomas de *Buchnera* disponibles incluyen todos los genes que codifican las actividades necesarias para sintetizar triptófano a partir del corismato, este último también producido por la bacteria. El primer paso de la ruta es catalizado por la antranilato sintasa, codificada por los genes *trpEG*; en los pasos siguientes, el antranilato es convertido en triptófano mediante la acción de cuatro actividades enzimáticas, codificadas por los genes *trpABCD* (figura 5.9a). Sin embargo, como ya se ha discutido en el apartado previo, en el áfido del cedro la ruta del triptófano se encuentra distribuida entre los endosimbiontes *B. aphidicola* BCc y *S. symbiotica* SCc (Gosalbes et al., 2008; Sloan and Moran, 2012b). La distribución de los genes es la siguiente: el genoma de *B. aphidicola* BCc codifica *trpGE* y el de *S. symbiotica* SCc codifica *trpABCD* (ver figura 5.9b). Dado que los endosimbiontes están sujetos a un proceso de reducción genómica extrema, el evento de complementación es en buena medida una consecuencia de dicho proceso. Por otra parte, es claro que esta forma de síntesis de triptófano funciona y permite satisfacer las necesidades nutricionales del áfido del cedro y los simbiontes. De otra manera no habría sido posible que el evento se fijase en una población ancestral. Sorprendentemente, la misma complementación también ha sido descrita en el consorcio endosimbionte albergado por el psílido *Heteropsylla cubana*, que incluye a *Ca. Carsonella ruddii* y otro endosimbionte, clasificado tentativamente como secundario (Sloan and Moran, 2012b). Es importante destacar que si bien los áfidos y los psílicos pertenecen al mismo suborden (Sternorrhyncha), sus endosimbiontes respectivos pertenecen a taxones diferentes, lo que indica eventos de simbiosis diferentes (Baumann et al., 2006). Dicha independencia sugiere que la complementación metabólica en la ruta del triptófano que exhiben ambos consorcios es un evento de convergencia evolutiva. Además, recientemente el mismo evento ha sido también descrito en el consorcio del pulgón gigante del sauce (*Tuberolachnus salignus*), un pariente del áfido del cedro que también alberga a una *Buchnera* y una *Serratia* (Manzano-Marín et al., 2016). Si bien es posible que el origen del evento de complementación haya tenido lugar en un ancestro común de ambos pulgones, el análisis filogenético de ambos áfidos y sus endosimbiontes, apuntan a que la complementación en la ruta del triptófano no estaba presente en el ancestro común de ambos sistemas (Manzano-Marín et al., 2016), lo que señala otro posible caso de convergencia evolutiva. Por otra parte, en todos los sistemas mencionados que presentan el evento de complementación en la ruta del triptófano, la distribución de las actividades es la misma: *trpGE* en un organismo y *trpDCBA* en el otro (ver figura 5.9). Este hecho conduce a la pregunta de si la organización de la ruta del triptófano distribuida en dos organismos, presenta alguna ventaja respecto a otros posibles diseños. De esta manera, empleando la reconstrucción del consorcio *iBSCc* se realizaron una serie de experimentos *in-silico*, para investigar si la complementación metabólica en la ruta del triptófano exhibida por el consorcio de endosimbiontes del áfido del cedro, presenta alguna ventaja frente a diseños alternativos en la organización de la ruta. Para realizar los experimentos primero se construyó una versión del modelo de consorcio *iBSCc* que representa un hipotético escenario ancestral en el cual tanto *Buchnera* como *Serratia* son capaces de sintetizar triptófano.

Tabla 5.6. Actividades enzimáticas y genes involucrados en la síntesis del triptófano a partir de la eritrosa-4-fosfato. La tabla incluye las reacciones de la ruta de síntesis del shiquimato, el corismato y el triptófano. La columna FCS indica los cuatro conjuntos de acoplamiento completo en que se agrupan las trece reacciones. Los identificadores de las reacciones corresponden a la base de datos BiGG y la columna Acc. indica el nombre del locus para cada gen en los genomas de *B. aphidicola* BCc y *S. symbiotica* SCc.

FCS	Numero EC	R. ID	Formula	Gen	Acc
FCS1	2.5.1.54	DDPA	e4p + h ₂ o + pep → 2dda7p + pi	<i>aroH</i>	BCc_077
	4.2.3.4	DHQS	2dda7p → 3dhq + pi	<i>aroB</i>	BCc_352
	4.2.1.10	DHQD	3dhq → 3dhsK + h ₂ o	<i>aroQ</i>	BCc_251
	1.1.1.25	SHK3Dr	3dhsK + h + nadph → nadp + skm	<i>aroE</i>	BCc_312
FCS2	2.7.1.71	SHKK	atp + skm → adp + h + skm5p	<i>aroK</i>	SCc_638
	2.5.1.19	PSCVT	pep + skm5p → 3psme + pi	<i>aroA</i>	SCc_515
	4.2.3.5	CHORS	3psme → chor + pi	<i>aroC</i>	SCc_617
FCS3	4.1.3.27	ANS	chor + gln-L → anth + glu-L + h + pyr	<i>trpEG</i>	<i>Cc_pT01</i> & <i>BCc_pT02</i>
FCS4	2.4.2.18	ANPRT	anth + prpp → ppi + pran	<i>trpD</i>	SCc_377
	5.3.1.24	PRAIi	pran → 2cpr5p	<i>trpC</i>	SCc_378
	4.1.1.48	IGPS	2cpr5p + h → 3ig3p + co ₂ + h ₂ o	<i>trpC</i>	SCc_378
	4.1.2.8	TRPS3	3ig3p → g ₃ p + indole	<i>trpB</i>	SCc_379
	4.2.1.122	TRPS2	indole + ser-L → h ₂ o + trp-L	<i>trpA</i>	SCc_380

Como se muestra en la [tabla 5.6](#), la síntesis de triptófano requiere como precursor el corismato que se sintetiza a partir del shiquimato, el cual a su vez es producido a partir de la eritrosa-4-fosfato (E4P), a través de cuatro reacciones enzimáticas. Como se muestra en la tabla, el genoma de *B. aphidicola* BCc ha retenido todos los genes que codifican las actividades necesarias para sintetizar antranilato a partir de la E4P, lo que implica que también es capaz de sintetizar el corismato a partir de este mismo precursor. Por su parte, el genoma de *S. symbiotica* SCc conserva, por un lado, los genes para la síntesis de corismato a partir de shiquimato y por otro, los genes para la síntesis del triptófano a partir del antranilato. De acuerdo con lo anterior, para la construcción del consorcio ancestral se añadieron las actividades enzimáticas codificadas por los genes ausentes en cada modelo. Es decir que las actividades codificadas por *trpDCBA* fueron añadidas en el compartimento correspondiente a *Buchnera*, mientras que en el compartimento correspondiente a *Serratia*, se introdujeron las actividades codificadas por los genes *trpGE* (i.e. la antranilato sintasa I). Con respecto a la síntesis del shiquimato a partir de la E4P, la información disponible en BioCyc 19.0, indica que de las diecisiete cepas de *Serratia* (correspondientes a nueve especies diferentes, para las cuales hay disponibles un genoma y una PGDB reconstruida), dieciséis presentan los genes para la síntesis de shiquimato a partir de E4P, siendo *S. symbiotica* SCc la única especie en que es están ausentes. De acuerdo con esta información, es razonable suponer en el genoma del ancestro de *S. symbiotica* estuviesen presentes dichos genes (así como los del triptófano). Como consecuencia, las actividades codificadas por los genes *aroH*, *B*, *Q*, y *E* (necesarias para sintetizar shiquimato a partir de eritrosa-4-fosfato) fueron añadidas al compartimento SCc. Finalmente, las cotas de los flujos de transportes del shiquimato, corismato, antranilato y triptófano, fueron

establecidas de manera que estos compuestos pueden ser intercambiados libremente entre ambos endosimbiontes. Esta versión del modelo de consorcio será referida como $iBSCc_{Ances}$, ya que representa un plausible consorcio ancestral.

Los experimentos *in-silico*, diseñados para analizar la eficiencia de posibles diseños alternativos de la complementación consisten en simular de forma exhaustiva todos los posibles esquemas de complementación metabólica entre los endosimbiontes que permiten la síntesis de triptófano. A continuación, para cada esquema se evalúa la eficiencia del sistema mediante la optimización de dos objetivos alternativos: la producción de meta-biomasa $v_{meta-biomasa}$ y la producción de triptófano v_{Trp} . Los esquemas analizados incluyen el caso del ancestro hipotético $iBSCc_{Ances}$ (donde la ruta está presente en ambos simbiontes), así como todos los demás diseños (escenarios de pérdida de genes) que son funcionales, *i.e.* que permiten sintetizar triptófano. Dado que por su estructura, la ruta incluye conjuntos de reacciones estequiométricamente acopladas, no es necesario realizar las combinaciones con el conjunto completo de veintiséis genes (trece en cada endosimbionte) sino que es suficiente con elegir una actividad (o gen) por cada FCS. Esto es posible ya que los genes que codifican actividades que pertenecen a un mismo conjunto de acoplamiento completo (FCS) representan una clase de equivalencia, y bajo la suposición de estado estacionario, estas se comportan como si fueran una única reacción (ver apartado A.2.3 en el apéndice A).

Dentro de cada endosimbionte, las trece reacciones analizadas se encuentran agrupadas en un total de cuatro FCS simétricos (en el modelo del ancestro $iBSCc_{Ances}$ la ruta está presente en cada compartimento). Dado que hay cuatro FCS en cada organismos y que para que se pueda completar la síntesis de triptófano al menos una de las copia de cada FCS ha de ser funcional (si se pierden las dos copias un FCS en forma simultánea el sistema no sería viable), el número máximo de FCS que se pueden perder es cuatro. Dicho de otra forma, los diseños mínimos que permiten la síntesis de triptófano son las combinaciones de cuatro FCS que no incluyen FCS redundantes. Por cada FCS se eligió como representante la última reacción del subconjunto (*e.g.* en el caso del FCS2 se seleccionó la CHORS). Cada uno de los diseño se representa con uno o más símbolo de tipo Δ_{ANS}^{BCc} , donde el superíndice indica el endosimbionte o compartimento y el subíndice la actividad que se supone pérdida. Por ejemplo, Δ_{ANS}^{BCc} corresponde al caso en *Buchnera* ha perdido la ANS, mientras que *Serratia* tiene la ruta completa. Teniendo en cuenta lo anterior se plantearon y analizaron los siguientes diseños: i) el caso del sistema redundante (*i.e.* el consorcio ancestral) en el que se encuentran activas las dos copias de cada enzima; ii) los ocho casos posibles de *knockout* simples (*i.e.* los cuatro FCS en cada uno de los organismos, *e.g.* Δ_{ANS}^{BCc}); iii) todas las combinaciones de pares posibles de FCS, *e.g.* Δ_{ANS}^{BCc} & Δ_{CHORS}^{SCc} ; iv) las combinaciones posibles de *knockout* triples, *e.g.* Δ_{ANS}^{BCc} & Δ_{CHORS}^{SCc} & Δ_{TRPS}^{SCc} ; y v) todas las combinaciones de *knockout* cuádruples, *e.g.* Δ_{SHK3Dr}^{BCc} & Δ_{ANS}^{BCc} & Δ_{CHORS}^{SCc} & Δ_{TRPS2}^{SCc} , que corresponden a los diseños mínimos. Si bien, el número total de diseños es 163, hay un total de 82 casos que corresponden a situaciones donde hay al menos un doble letal (*e.g.* Δ_{ANS}^{BCc} & Δ_{ANS}^{SCc}) y por lo tanto carecen de interés ya que representan versiones no funcionales del sistema. Como consecuencia, el análisis se centrará en los 81 diseños viables. Empleando FBA, cada uno de los diseños del consorcio fue caracterizado empleando las dos funciones objetivo, antes mencionadas: la maximización del flujo de producción de meta-biomasa y la maximización del flujo de producción de triptófano. En cada caso, el valor óptimo obtenido es normalizado con

respecto al valor óptimo predicho con el modelo del consorcio ancestral $iBSCc_{Ances}$, para poder evaluar cómo se ve afectado *fitness* o eficiencia del sistema, en la medida que se van perdiendo genes. Los óptimos normalizados para el caso de la meta-biomasa y el triptófano se denotan como $v_{meta-bio}^*$ y v_{Trp}^* , respectivamente. Finalmente, como se vio en el apartado 5.3.1 el balance de protones en el compartimento de *Buchnera* implica una restricción que de ser relajada afecta en forma sensible las predicciones del modelo. Para controlar el efecto de dicha restricción en los resultados, los experimentos se realizaron por un lado, considerando la restricción sobre el balance estricto de los protones y, por otro, relajando la misma.

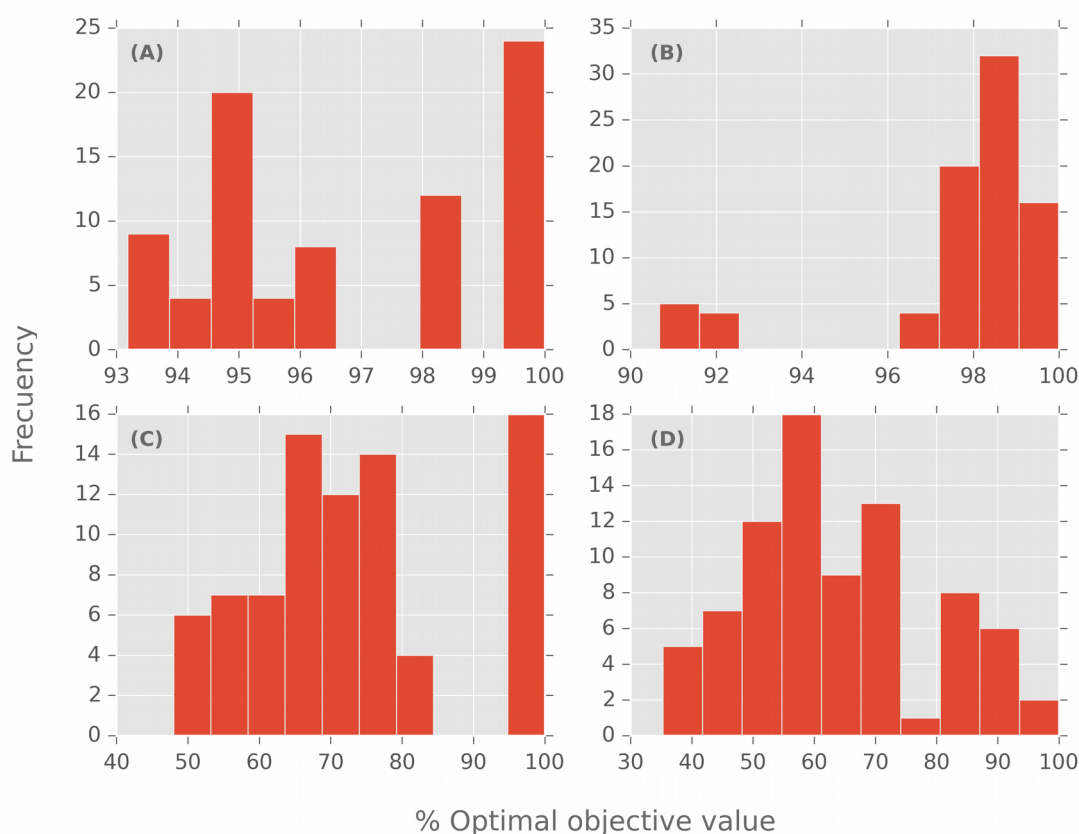


Figura 5.10. Distribuciones de los óptimos de *fitness* de los distintos diseños de la complementación en la ruta del triptófano. En los cuatro casos analizados, los valores de *fitness* óptimo obtenido para cada uno de los 81 experimentos, fueron normalizados contra el valor máximo identificado en la solución predicha por $iBSCc_{Ances}$. En (A) y (B) se representan los resultados para la maximización de la meta-biomasa, con y sin restricción en el balance de protones de *Buchnera*, respectivamente. En (C) y (D) se representan los resultados de la maximización de la producción de triptófano, con y sin restricción en el balance de protones, respectivamente.

La figura 5.10 resume los resultados de los experimentos realizados para evaluar los distintos diseños de los consorcios⁷. Los resultados indican que si el *fitness* del sistema es evaluado en términos de la producción de meta-biomasa, los 81 diseños viables permiten alcanzar un flujo de producción de meta-biomasa $v_{meta-biomasa}^* > 90\%$ con respecto al óptimo identificado para $iBSCc_{Ances}$. Además, el mismo comportamiento cualitativo se obtiene al considerar la restricción sobre el balance de protones como al relajar la misma (figuras 5.10AB). Esto indica que cualquier escenario de pérdida de genes conduce a diseños en los cuales el crecimiento del sistema disminuye como

⁷ Los resultados detallados de las simulaciones se encuentran en la tabla suplementaria S.5.6.

máximo, un 10% respecto al ancestro hipotético. En el caso de la maximización de la meta-biomasa con balance estricto de protones, los resultados muestran que en la distribución de flujo óptima que exhibe $iBSCc_{Ances}$, todas las reacciones de los cuatro FCS de *B. aphidicola* BCc tienen un flujo de cero y sería *Serratia* quién produciría todo el triptófano, y el corismato necesarios para satisfacer las demandas de *Buchnera* y al hospedador. Como consecuencia todos los escenarios que involucran pérdida de genes en *Buchnera* pero no en *Serratia*, alcanzan el mismo valor de *fitness* que $iBSCc_{Ances}$. Si bien, este resultado es contrario al diseño de la complementación que exhibe el consorcio real del áfido del cedro, si se exploran soluciones quasi-óptimas ($v_{meta-bio}^* \geq 99,5\%$) se encuentran diseños donde *Buchnera* carece las actividades del FCS4 (síntesis del triptófano a partir de antranilato) y *Serratia* las del FCS3 (*i.e.* la antranilato sintasa), los cuales corresponden al caso del consorcio real. Los resultados obtenidos al relajar la restricción del balance protones son cualitativamente muy similares a los descritos para el caso del balance de protones.

Como se comentó previamente, los valores máximos de producción de meta-biomasa exhibidos por los distintos diseños no exhiben gran variabilidad. De hecho, si se consideran todos los diseños cuyo valor óptimo normalizado sea $v_{meta-bio}^* \geq 95\%$, se encuentran diseños de consorcio totalmente complementarios. De esta manera no es posible identificar patrones cuyo *fitness* sea sensiblemente superior o inferior al resto. Es posible que los cambios en la eficiencia asociados a cada diseño no se reflejen en el valor del óptimo de la función objetivo, debido a que el triptófano sólo representa un pequeña proporción de la meta-biomasa ya que su proporción en cada biomasa individual es también muy pequeña. Por otra parte, si el *fitness* de los diseños se evalúa empleando el flujo máximo de producción de triptófano los resultados cambian notoriamente. En primer lugar, los distintos diseños exhiben un gran variabilidad en el valor óptimo normalizado, el cual varía entre un 50 y el 100%, y entre 35 y 100%, para los experimentos con y sin restricción en el balance de protones de *Buchnera*, respectivamente (ver figuras 5.10CD).

Al analizar los experimentos donde se exige un balance estricto de protones, lo primero que llama la atención es que el diseño Δ_{TRPS2}^{BCc} exhibe el mismo valor óptimo en la producción de triptófano que el modelo del ancestro, es decir que en este escenario, *Buchnera* podría haber perdido todas las actividades correspondientes al FCS4 sin que esto tuviera efecto alguno sobre la eficiencia de la producción del triptófano. Además, si se analizan diseños cuyo valor objetivo se encuentra muy próximo al óptimo ($v_{Trp}^* \geq 99,5\%$) aparece el caso Δ_{TRPS2}^{BCc} & Δ_{ANS}^{SCc} que es muy similar al diseño del consorcio real. Sin embargo, a diferencia con el consorcio de *C. cedri*, en este diseño *Serratia* produciría una pequeña cantidad de corismato (a partir de la E4P), que sería empleado por *Buchnera* para producir más antranilato. Sin embargo, si se exploran soluciones un poco más alejadas del óptimo ($v_{Trp}^* \geq 95\%$), hay un total de 16 diseños, en los cuales el flujo a través del FCS4 en *Buchnera* es cero, entre los cuales se encuentra la solución Δ_{TRPS2}^{BCc} & Δ_{ANS}^{SCc} & Δ_{SHK3Dr}^{SCc} , que corresponde al diseño encontrado en consorcio del áfido del cedro. Los restantes 16 diseños mínimos exhiben una disminución en el valor máximo de producción de triptófano de entre 25% y 50%, respecto a lo predicho para el caso del consorcio ancestral. Esto indica, que el diseño de la complementación que exhibe el consorcio del áfido del cedro es la solución mínima respecto al número de genes, y eso no afecta en mayor medida el rendimiento en la producción del triptófano (ver tabla suplementaria S.5.6).

Al considerar los resultados obtenidos relajando la restricción sobre el balance de protones en el compartimento que corresponde a *Buchnera*, éstos muestran cierta similitud con los descritos al exigir el balance estricto de protones, pero presentan ciertas diferencias relevantes. En primer lugar, ningún diseño en el que se eliminen genes alcanza el 100% del óptimo. En efecto, en la solución óptima exhibida por $iBSCc_{Ances}$ cada endosimbionte produciría el triptófano *de novo*, y de forma independiente. Además, el subconjunto de genes cuya pérdida implican la menor disminución en la producción de triptófano corresponde al FCS1 de *Serratia*. Este resultado es consistente con el consorcio real, donde dicho endosimbionte ha perdido los genes que codifican las actividades para la síntesis del shiquimato a partir de la E4P. Por otra parte, existe un subconjunto de 16 diseños cuyo valor $v_{Tri}^* \geq 85\%$, que incluyen la solución Δ_{TRPS2}^{BCc} & Δ_{ANS}^{SCc} & Δ_{SHK3Dr}^{SCc} , algo similar a los resultados obtenidos cuando se considera el balance de los protones. En efecto, dentro de estos 16 diseños solo hay 4 casos en los que *Buchnera* conserva las actividades necesarias para sintetizar triptófano a partir del antranilato (FCS4). Con respecto, a la antranilato sintasa en *Serratia*, la mitad de los diseños no presentan flujo a través de la misma.

Si bien para los experimentos optimizando la meta-biomasa los resultados no muestran diferencias claras en la eficiencia de producción (tanto si se considera el efecto de los protones como si no), al considerar la producción de triptófano como objetivo, los resultados sí muestran que el diseño que se observa en el consorcio de *C. cedri* presenta ciertas ventajas. O dicho de otra manera, el diseño estequiométrico de la complementación en la ruta del triptófano presente en el dicho consorcio, se correspondería al caso donde se pierden el máximo número de genes, sin que esto afecte de forma drástica las capacidades en la producción de triptófano, la cual se mantiene a un 95% del óptimo exhibido por el consorcio ancestral $iBSCc_{Ances}$.

5.3.6 Complementación metabólica: condiciones necesarias y optimalidad

Los resultados presentados en el apartado previo muestran como distintos diseños estructurales (o estequiométricos) de complementación en una ruta metabólica exhiben gran variabilidad en la eficiencia de la síntesis. Sin embargo, además de la estructura es muy posible que otras propiedades del sistema también afecten la eficiencia del proceso. Por ejemplo, la cinética de las reacciones y el coste asociado a la producción de las enzimas (proteínas) deben suponer restricciones y constricciones adicionales que afectaran la eficiencia del proceso de síntesis. Concretamente, al igual que en el caso de la mayoría de los aminoácidos, la ruta de síntesis del triptófano suele encontrarse regulada mediante un *feedback* negativo en el cual el producto final, *i.e.* el triptófano, se une de forma inhibe la primera enzima de la ruta (Kwak et al., 1999; Rouhbakhsh et al., 1997, 1996; Spraggon et al., 2001). Dado que un mecanismo de este tipo se encuentra diseñado para mantener el flujo de producción ajustado a las necesidades del organismo, este debe representar un obstáculo si se intenta incrementar el valor de dicho flujo para sobre-producir y compartir el producto (Stephanopoulos and Vallino, 1991). Como consecuencia, se esperaría que los endosimbiontes tenga cierta propensión a perder tales mecanismo, de la misma manera que se observa como pierden los mecanismos de regulación transcripcional (Baumann et al., 2006; Wilcox et al., 2003). Por ejemplo, se han descrito casos de enzimas, como la pefenato deshidratasa, donde se cree que la inhibición se ha perdido como consecuencia de mutaciones puntuales en el sitio alostérico (Jiménez et al., 2000). Sin embargo, estudios realizados sobre la conservación de *trpE* en distintas cepas de *B. aphidicola*, muestran que los residuos clave en el sitio de regulación alostérico

de la enzima antranilato sintasa se encuentran conservados en la secuencia del gen (Lai et al., 1994; Rouhbakhsh et al., 1996). No obstante, dichos estudios no incluyen a la secuencia de *trpE* de *B. aphidicola* BCc. Por lo tanto, para evaluar el nivel de conservación del sitio alostérico de la antranilato sintasa en el caso de *B. aphidicola* BCc, se realizó un alineamiento múltiple incluyendo la secuencia del gen *trpE* presente en *B. aphidicola* BCc, su homólogo en otras once cepas de *B. aphidicola*, junto con los homólogos en *E. coli* y *S. marcescens* (ver el apartado 5.2.6 de la sección Materiales y Métodos).

La figura 5.11 muestra una sección del alineamiento múltiple, donde se puede ver que los residuos P21 y S40, así como el motivo LLES del sitio alostérico, descritos como importantes para la unión del triptófano (Caligiuri and Bauerle, 1991; Kwak et al., 1999; Tang et al., 2001), se encuentra mejor conservados en *B. aphidicola* BCc que en las demás cepas de *Buchnera*. Además, cambios como la sustitución Ser40 → Thr40, presentes en algunas cepas de *Buchnera*, podrían no tener consecuencias en la función, ya que ambos aminoácidos presentan propiedades fisicoquímicas similares. Esto sugiere que el mecanismo de inhibición puede aún estar activo en el endosimbionte y, de ser así es posible que aumentar el flujo requiera un aumento de la concentración de la enzima y de su coste asociado.

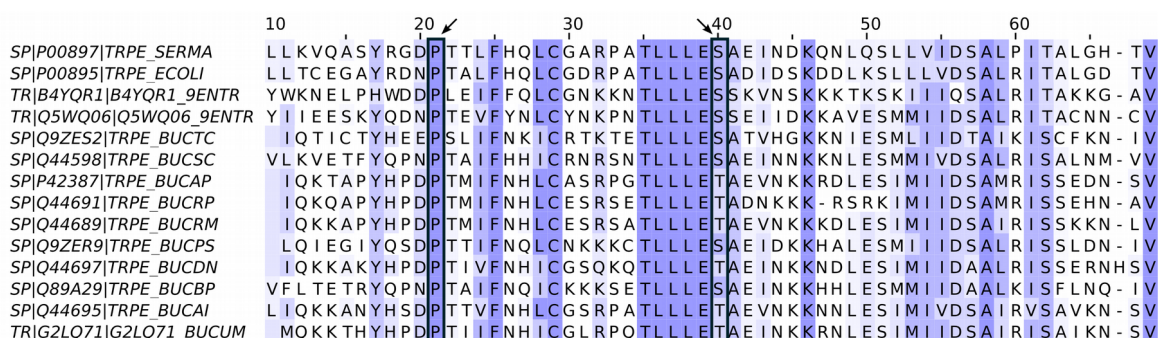


Figura 5.11. Alineamiento múltiple del gen *trpE* presente en doce cepas de *B. aphidicola*, junto a sus homólogos en *E. coli* K12 y *S. marcescens*. Todas las secuencias son aproximadamente de la misma longitud (~515 residuos) y la figura está enfocada en la región que corresponde al sitio alostérico. La primera, segunda y tercera filas (de arriba a abajo), corresponden a las secuencias de *S. marcescens*, *E. coli* y *B. aphidicola* BCc, respectivamente. Las demás secuencias pertenecen a distintas cepas de *Buchneras* (ver apartado 5.2.6). Las flechas en las columnas 21 y 40, señalan residuos descritos como claves del sitio alostérico (Caligiuri and Bauerle, 1991; Kwak et al., 1999; Tang et al., 2001).

Dado que el endosimbionte no solo produce el aminoácido para su propio consumo, sino que también debe proveer al hospedador del mismo, se espera que, como consecuencia del aumento en la demanda del aminoácido, los simbiontes deban incrementar su flujo de producción. Ahora bien, si se supone que el mecanismo de inhibición por producto que regula dicho flujo sigue funcionando en los endosimbiontes, el efecto inhibitorio causado por el aumento en la concertación del triptófano debería ser compensado de alguna forma, e.g. incrementado la concentración de la antranilato sintasa (Lai et al., 1994). En efecto, en diversas cepas de *B. aphidicola* se ha encontrado que los genes *trpEG* se encuentran amplificados, o bien como múltiples copias en el genoma, o mediante la presencia de plásmidos que incluyen una o más copias de estos (Baumann et al., 2006; Rouhbakhsh et al., 1997). Por ejemplo, la cepa de *Buchnera* del áfido *A. pisum*, contiene un plásmido que incluye una copia de *trpE* y dos de *trpG*. En el caso de *B. aphidicola* BCc, esta también contiene un plásmido (pTrp-Bcc: NG_036609) pero este incluye solo una única

copia de cada gen (Gosalbes et al., 2008; Lamelas et al., 2011a). La presencia de múltiples copias del gen que codifica para la enzima de la ruta regulada mediante inhibición por producto, sugiere una posible adaptación para incrementar el flujo de producción de triptófano (Lai et al., 1994; Rouhbakhsh et al., 1997). Si se considera esta evidencia junto al hecho de que el sitio de unión alostérica se encuentra bien conservado en las distintas cepas de *Buchnera*, es razonable suponer que este mecanismo de regulación sigue siendo funcional. En tal caso, aumentar la concentración de la enzima para compensar el efecto de la inhibición podría significar un coste elevado para el endosimbionte, en términos de la síntesis de enzimas (proteínas).

Ahora bien, el hecho de que en diversos consorcios la ruta del triptófano se encuentre distribuida en dos tipos de endosimbiontes diferentes, segregando en compartimientos diferentes la actividad enzimática regulada, de la actividad que sintetiza el producto final, podría tener implicaciones directas sobre la cinética del proceso. De acuerdo con lo anterior, es muy posible que la concentración de triptófano sea muy baja en el compartimento donde se encuentran la antranilato sintasa (*i.e.* *B. aphidicola* BCc) y como consecuencia de esto se atenúe el efecto de la inhibición sobre la enzima regulada. Este diseño podría permitir aumentar el flujo de síntesis de antranilato sin que esto represente un gran incremento en la concentración de la enzima. Para poder abordar esta pregunta en términos teóricos se requiere que las concentraciones de las enzimas y los metabolitos sean variables explícitas del modelo. Realizar dichas simulaciones empleando el modelo *iBSCc* requeriría conocer los mecanismos de reacción y los parámetros cinéticos de todas las enzimas de la red. Sin embargo, dicha aproximación es algo que de momento se encuentra lejos de ser viable, debido a la falta de datos experimentales para el sistema y a la dificultad de construir modelos cinéticos a escala genómica. Sin embargo, es posible adoptar una aproximación, que no requiere conocer tantos detalles del sistema y que permita entender aspectos más generales del problema. Esto implica construir un modelo sencillo que solo retenga las propiedades y características que, se entiende, dan cuenta del fenómeno de la complementación metabólica.

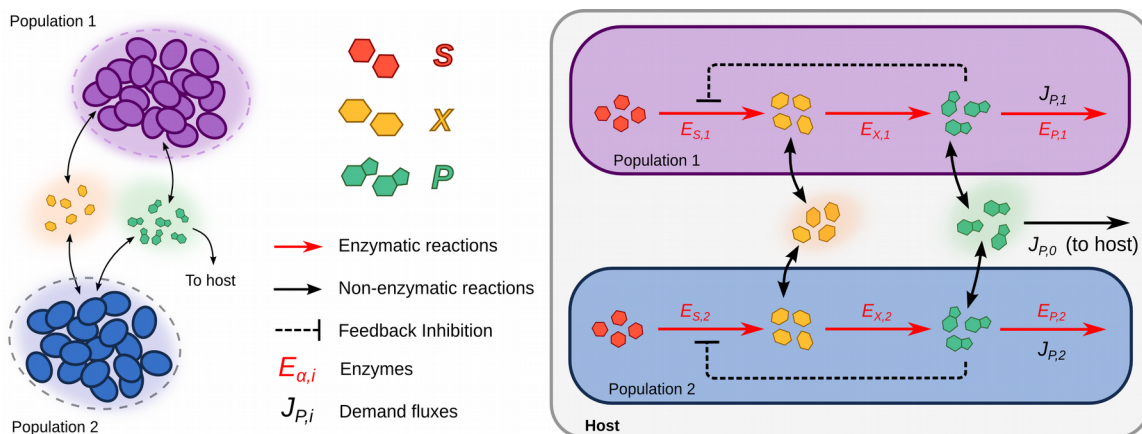


Figura 5.12. Esquema del modelo cinético de complementación metabólica. El modelo representa dos tipos de endosimbiontes (óvalos violetas y azules) que codifican la misma ruta hipotética que permite sintetizar el producto P a partir del sustrato S , a través de un intermediario X . Cada actividad de la ruta es catalizada por una enzima. Además, tanto el intermediario (X) como el producto final (P) pueden difundir a través de las membranas celulares. Tanto los endosimbiontes como el hospedador requieren de P para poder crecer; esto se representa mediante los flujos de consumo: $J_{P,0}$, $J_{P,1}$ y $J_{P,2}$. El modelo representado en la figura, es una abstracción del esquema presentado en la figura 5.9.

Con el objetivo de analizar los efectos que la cinética tiene sobre el fenómeno de la complementación metabólica, se desarrolló un modelo cinético simple para analizar la complementación metabólica de un consorcio endosimbionte. El modelo describe un consorcio compuesto por dos especies de bacterias endosimbiontes que codifican una ruta biosintética putativa, en el caso en que estas pueden intercambiar intermediarios de dicha ruta (ver figura 5.12). Además, se supondrá que cada una de las poblaciones de bacterias está representada por un gran número de células idénticas (N_i con $i=1,2$). En el modelo, cada una de las dos poblaciones de endosimbiontes es representada mediante un único compartimento y estos a su vez están embebidos en un compartimento extracelular a través del cual pueden difundir los compuestos intercambiados⁸. Ambos endosimbiontes codifican para la misma ruta biosintética lineal que permite la conversión del sustrato S (producido por la bacteria), a través del intermediario X en un producto final P , el cual se considera esencial para todos los miembros del consorcio. Además, tanto el metabolito X como P , se supone que pueden difundir fuera o dentro de la célula. Dado que P es definido como un nutriente esencial para los tres miembros del sistema, *i.e.* los dos endosimbiontes y su hospedador, el modelo incluye tres flujos que consumen dicho metabolito. Si bien en el modelo no hay ecuaciones de biomasa explícitas estos flujos representan la incorporación de P en las biomásas respectivas. Finalmente, dado que la inhibición por producto puede estar presente en algunos endosimbiontes, es de interés evaluar qué papel puede tener dicho mecanismo de regulación, en la emergencia de complementación metabólica⁹. Por tanto, el modelo incluye un mecanismo de inhibición por producto que regula la producción de P .

Las dos reacciones que componen la ruta de biosíntesis, $S \rightarrow X$ y $X \rightarrow P$, con flujos respectivos V_S y V_X , son modeladas empleando un cinética reversible de tipo Michalis-Menten, siguiendo la aproximación de Noor y colaboradores (2014). Por lo general, en la presencia de inhibición por producto, el paso regulado suele ser la primera enzima de la ruta (*e.g.* el caso del triptófano). Como consecuencia, la ecuación que representa la primera reacción ($S \rightarrow X$) incluye un término adicional para representar el efecto de la inhibición por producto. Si bien existen diversos mecanismos de inhibición enzimática, que incluyen la inhibición alostérica, en este caso se ha elegido uno de los mecanismos más simples (en términos de su formulación matemática¹⁰) que es el mecanismo de inhibición incompetitiva que da lugar a la siguiente ecuación:

$$V_S = \frac{k_{cat}[E_S]}{1+[P]/K_I} \cdot \frac{([S]-[X]/K_{eq})/K_M^+}{1+[S]/K_M^+[X]/K_M^-} \quad (\text{Ec. 5.2})$$

en donde k_{cat} es la constante catalítica de la enzima; K_M^+ y K_M^- son las constantes de Michaelis para la reacción directa y reversa, respectivamente; K_{eq} es la constante de equilibrio de la reacción $S \rightarrow X$; y K_I es la constante de la inhibición causada por la unión de P a la enzima¹¹.

8 La estructura en compartimentos, es equivalente a la utilizada en el modelo de consorcio *iBSCc* presentado en el apartado 5.3.2.

9 Por ejemplo, la antranilato sintasa de *B. aphidicola* BAp exhibe un alto grado de conservación en su sitio de unión alostérico (ver figura 5.11).

10 El caso de la inhibición competitiva será analizado en el apartado más adelante, en esta misma sección.

11 Se puede ver que en el caso en que la inhibición sea muy pequeña, *i.e.* $1/K_I \rightarrow 0$, la ecuación se aproxima a una cinética de Michaelis-Menten reversible.

Para representar el posible intercambio de metabolitos entre ambos endosimbiontes, el modelo incluye reacciones de transporte que permiten el flujo de los metabolitos X y P entre los distintos compartimentos. Dado que una característica común de los endosimbiontes es la falta de sistemas de transporte (Charles et al., 2011), los flujos de transporte de X y P son modelados empleando una ecuación de difusión de primer orden:

$$U_{\alpha} = D_{\alpha}([\alpha]_{in}-[\alpha]_{out}) , \quad \alpha = X, P \quad (\text{Ec. 5.3})$$

Donde U_{α} es el flujo correspondiente al metabolito α y D_{α} es su coeficiente de difusión a través de la membrana. En este caso, los flujos serán considerados positivos cuando el metabolito es excretado por cualquiera de los endosimbiontes y negativos, cuando estos sean importados. Por último, como se menciona al inicio del apartado, el producto P se supone necesario para el crecimiento de cada miembro del sistema, de manera que para representar dicho consumo se emplea una cinética saturable:

$$V_P = k_{cat}[E_P] \frac{[P]}{[P]+K} \quad (\text{Ec. 5.4})$$

Donde V_P es el flujo de consumo de P por parte de cada endosimbionte, y $[E_P]$ corresponde a la proteína necesaria para integrar el compuesto en la biomasa del organismo. Dado que se supone que P es esencial para el crecimiento, los flujos de consumo $V_{P,i}$ deberán ser mayores que cero para que el sistema pueda crecer.

Por último está la consideración de la masa relativa de endosimbiontes respecto al hospedador, *i.e.* lo que en el modelo *iBSCc* sería representado mediante la estequiometría de la meta-biomasa (ver apartado correspondiente en el [apartado 5.3.2](#)). A diferencia del análisis estequiométrico, al realizar un estudio dinámico del sistema, las concentraciones de los intermediarios, así como de las enzimas, son variables de estado del sistema, y deben ser tomadas en cuenta. Entonces, para facilitar la interpretación de los resultados, las poblaciones absolutas de los endosimbiontes, N_1 y N_2 son normalizadas por la población total $N=(N_1+N_2)$. Esto da como resultado $n_1=N_1/N$ y $n_2=N_2/N$. Al igual que en el estudio del consorcio del áfido del cedro, se supondrá que las poblaciones de ambos endosimbiontes son iguales, *i.e.* $n_1=n_2=1/2$. Tomando en cuenta el modelo completo, las ecuaciones de balance que describen el comportamiento dinámico del sistema se pueden escribir de la siguiente forma:

$$\begin{aligned} \frac{d[X_i]}{dt} &= V_{S,i} - V_{X,i} - U_{X,i} , \quad i=1,2 \\ \frac{d[P_i]}{dt} &= V_{X,i} - V_{P,i} - U_{P,i} , \quad i=1,2 \\ \frac{d[X_0]}{dt} &= N(n_1 U_{X,1} + n_2 U_{X,2}) \\ \frac{d[P_0]}{dt} &= N(n_1 U_{P,1} + n_2 U_{P,2} - J_{P,0}) \end{aligned} \quad (\text{Ec. 5.5})$$

En el caso de los metabolitos intracelulares, es importante notar que las ecuaciones de balance son iguales para ambos endosimbiontes, y que las concentraciones de estos no dependen de los tamaños de las poblaciones de las bacterias. Por otra parte, las ecuaciones de balance de los

metabolitos extracelulares sí dependen del tamaño de las poblaciones de simbioses y, en caso de P_0 , la concentración de este también depende del flujo de demanda de hospedador $J_{P,0}$. Finalmente, se considera que la producción del sustrato S es realizada por cada célula endosimbionte y su concentración se supone fija dentro de cada célula (como consecuencia no es necesario incluir su ecuación de balance). Por simplicidad, se considerará el valor de todos los parámetros y constantes cinéticas igual a uno, a menos que se indique de forma explícita un valor diferente. Además la concentración intracelular del sustrato S se considerará igual para ambos tipos de endosimbioses y su valor se fijará en $[S]=10$. Por último, y como ya fue mencionado, las poblaciones relativas de endosimbioses se fijan de la siguiente forma: $n_1=n_2=1/2$.

El criterio de optimización

Al igual que al realizar la comparación de los distintos diseños estructurales de la complementación metabólica con el modelo de consorcio *iBSCc*, para poder comparar distintos esquemas con del modelo cinético es preciso definir un criterio de optimización o un *fitness*, Las bacterias endosimbioses involucradas en interacciones nutricionales deben sobre-producir ciertos compuestos esenciales, como aminoácidos y/o vitaminas, para poder satisfacer la necesidad del hospedador (Douglas, 1992). Además, como ya fue mencionado, el incremento en la producción de dichos compuesto puede realizarse aumentando la concentración de las enzimas involucradas en la ruta, incluso en aquellos casos en que la inhibición por producto dificulta la modulación del flujo (Baumann et al., 1999; Lai et al., 1994; Rouhbakhsh et al., 1997). En este sentido, cualquier mejora en la eficiencia de la sobre-producción, en términos de la economía en la síntesis de proteína, podría conferir alguna ventaja al sistema completo, como se ha discutido en diversos estudios (Dekel and Alon, 2005; Flamholz et al., 2013; Goelzer and Fromion, 2011; Kafri et al., 2015; Mori et al., 2016; Ponce de León et al., 2008; Shachrai et al., 2010).

De acuerdo con lo anterior en este apartado se estudiará el compromiso entre la maximización del flujo a través de una ruta y la eficiencia en la distribución de las concentraciones de las enzimas. Empleando métodos de optimización no-lineal sujeta a restricciones, es posible modelar dicho compromiso, o *trade-off*, como dos problemas de optimización complementarios: i) maximización del flujo de producción de P sujeto a una restricción en la concentración total de enzima; y ii) minimización de la concentración total de enzima sujeto a una restricción en el flujo de interés, que garantice que la demanda es satisfecha. En lo que sigue se adoptará la segunda formulación del problema, *i.e.* la minimización del total de enzima necesaria para sostener un flujo de demanda dado, del producto P ¹².

De acuerdo con lo anterior, el problema de optimización se define en términos de: i) las variables de decisión que corresponden a las concentraciones de las enzimas e intermediarios; ii) las restricciones, que incluyen el balance de masas sobre cada metabolito; y iii) la función a optimizar, *i.e.* la concentración total de enzima. La función objetivo se define como la minimización de la concentración total de enzima (normalizada por la densidad poblacional de cada tipo de endosimbionte):

$$C = \sum_{i=1,2} \sum_{\alpha=S,X,P} n_i [E_{\alpha,i}] \quad (\text{Ec. 5.6})$$

¹² Puede mostrarse de forma teórica, que en el modelo analizado y posiblemente en general, ambos problemas convergen al mismo valor óptimo.

Donde C corresponde al total de enzima invertida por los dos endosimbiontes para producir una cierta cantidad de P que satisfaga la demanda total del sistema, *i.e.* de cada endosimbionte y del hospedador. La restricción que impone la demanda puede representar mediante una desigualdad sobre los flujos que representan el consumo de P :

$$V_{P,i} \geq J_{P,i} \quad \forall i = 0, 1, 3 \quad (\text{Ec. 5.7})$$

Donde $J_{P,i}$ son parámetros que representan la demanda de cada miembro del sistema. En forma adicional, se supondrá que las concentraciones de todos intermediarios cumplen el balance de masas, y esto se realiza introduciendo una restricción por cada ecuación de balance en la que se exige un balance cero, de forma que se garantiza que se cumple la condición de estado estacionario (ecuación 5.5).

De acuerdo con lo planteado, una solución particular de dicho problema consistirá en los valores concretos de concentraciones de cada metabolito y enzima del sistema (a partir de los cuales se pueden calcular los flujos a través de cada reacción) compatibles con las restricciones de estado estacionario y demandas impuestas. Además, en caso de tratarse de un óptimo, esta solución será la que emplee la concentración de enzima mínima necesaria para satisfacer los flujos de demanda. Dado que el interés se centra en poder comparar los perfiles de concentraciones que adoptan las enzimas en distintas soluciones, se introducirá una medida que permita cuantificar el grado de asimetría entre ambas células. El coeficiente de asimetría A_α se define para cada enzima $\alpha = S, X, P$, mediante la siguiente expresión:

$$A_\alpha = \frac{[E_{\alpha,1}] - [E_{\alpha,2}]}{[E_{\alpha,1}] + [E_{\alpha,2}]}, \alpha = S, X, P \quad (\text{Ec. 5.8})$$

Esta cantidad refleja la diferencia relativa en la concentración de cada enzima, entre las dos células. De manera que A_α valdrá cero cuando la concentración de la enzima α sea igual en ambas células, mientras que el coeficiente adoptará los valores extremos 1 o -1 cuando la concentración de la enzima sea cero en alguno de los dos tipos de células, o dicho de otra forma, cuando la enzima esté ausente en uno de los dos endosimbiontes. De esta manera se pueden representar los distintos estados que adopta el sistema al modificar parámetros de interés. Finalmente, dada la simetría del problema, *i.e.* ambos tipos de células son idénticos en términos de sus ecuaciones y parámetros cinéticos, solo serán considerados las soluciones en que $A_S \geq 0$. Los detalles de implementación se describen en el [apéndice C](#).

Emergencia de complementación metabólica

Empleando el modelo y el criterio de optimización introducidos en el apartado previo, se realizaron simulaciones para analizar la distribución óptima de las enzimas, al variar parámetros de interés que representan distintos escenarios. El primer escenario analizado no considera la demanda de P por parte del hospedador, de forma que el flujo de demanda correspondiente se fija en cero, *i.e.* $J_{P,0} = 0$, mientras que la demanda correspondiente a cada tipo de endosimbiontes se fija en $J_{P,i} = 1$. La [figura 5.13](#) muestra la distribución óptima de las enzimas, en términos del coeficiente de asimetría, para distintos valores de las constantes de difusión, D_X y D_P , y de la constante de inhibición K_I . Como se puede intuir, el caso extremo en que $D_X = D_P = 0$, los endosimbiontes no

pueden intercambiar metabolitos y como consecuencia la complementación metabólica no es posible. Sin embargo, en la medida que esta restricción se relaja y se permite que los metabolitos permeen, es posible la interacción entre ambos endosimbiontes y entonces pueden emerger soluciones asimétricas (*i.e.* complementación).

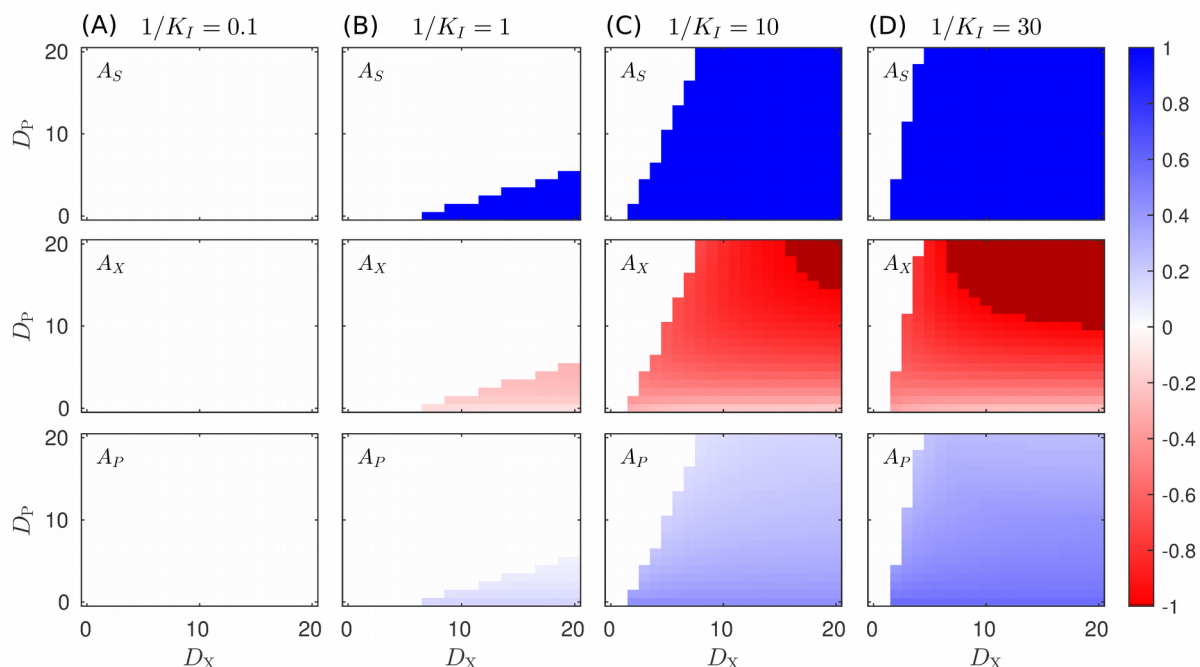


Figura 5.13. Coeficiente de asimetría A_α en distribuciones de enzimas óptimas en distintos escenarios. La figura representa distintas soluciones óptimas expresadas en términos del coeficiente de asimetría de cada enzima como función de las constantes de difusión D_X y D_P para el caso de la inhibición acompetitiva. Cada columna A-D representa las soluciones obtenidas para distintos valores de $1/K_I$ y cada panel de una columna corresponde a una de las tres enzimas consideradas en el modelo. La escala de color corresponde a los valores del coeficiente de asimetría de las distintas enzimas. Las zonas de color oscuros en los paneles de la enzima A_X correspondientes a $1/K_I=10$ y $1/K_I=30$ representan el rango de parámetros de difusión donde la ruta se encuentra completamente dividida entre ambas células, *i.e.* $[E_{S,2}]=[E_{X,1}]=0$.

Los resultados de las simulaciones indican que si el efecto de la inhibición es muy bajo (*e.g.* $1/K_I \leq 0.1$) las soluciones que minimizan la concentración total de enzima son simétricas en la distribución de todas enzimas, *i.e.* $A_\alpha = 0, \forall \alpha \in \{S, X, P\}$, con independencia de los valores de las constantes de difusión D_X y D_P . Contrariamente, cuando la constante de inhibición es lo suficientemente grande (*e.g.* $1/K_I \geq 1$), las soluciones asimétricas emergen como la estrategia óptima, para ciertos rangos de valores de las constantes de difusión (figuras 5.13 B-D). En efecto, para un valor dado de $1/K_I$, las constantes de difusión definen una línea de transición que divide el plano (D_X, D_P) en dos zonas que corresponden a los rangos de parámetros para los cual la solución simétrica (o asimétrica) es la estrategia óptima.

Cuando las condiciones favorecen la complementación (estrategia asimétrica), por lo general todas las enzimas exhiben cierto grado de asimetría. En particular, en el caso de la enzima E_S , el coeficiente de asimetría solo toma los valores extremos, $A_S = 0$ o $|A_S| = 1$ (indicando que la enzima

es expresada en un único endosimbionte). En cambio, en las demás enzimas el grado de la asimetría depende de los valores de las constantes de difusión (figuras 5.13B-D). Además, los resultados muestran de forma clara que el incremento en D_X favorece la emergencia de complementación, mientras que en ciertos escenarios (e.g. $1/K_I < 10$), el incremento D_P favorece las soluciones simétricas. Dicho de otra forma, las soluciones asimétricas surgen cuando la relación D_X/D_P supera cierto umbral, que depende de la constante de inhibición K_I . En definitiva, el modelo predice que la distribución asimétrica de las enzimas, *i.e.* la complementación, emerge como una estrategia óptima en presencia de inhibición de producto y cuando el producto final P , difunde más lento que el intermediario X . Esto último se debe al hecho de que si P difunde muy rápido, su concentración en ambas células será muy similar, y por lo tanto, el efecto de la inhibición será el mismo en ambas células. Contrariamente, si P difunde más lento que X , la concentración del primero en la célula que expresa la enzima E_S será baja, y su contribución como inhibidor será más baja, lo que permitirá sostener el mismo flujo neto de producción de P con una menor cantidad de enzima. En aquellas regiones donde la solución óptima se corresponde a una distribución asimétrica, se tiene la siguiente relación respecto a los coeficientes de asimetría de las tres enzimas: $A_S=1$, $A_X < 0$ y $A_P > 0$. Esto que implica que en las enzimas E_X y E_P , la asimetría no es completa. De manera que para poder ver de forma más clara el efecto de las constantes de difusión en la transición de las soluciones simétricas a las asimétricas, se estudió dicho cambio para un único valor de la constante de inhibición ($1/K_I=10$).

Para poder evaluar la eficiencia de cada estrategia se realizó un experimento, imponiendo como restricción que el coeficiente de asimetría valga $A_S=0$ y $A_S=1$, para analizar el caso simétrico y el asimétrico, respectivamente. De esta manera se calculó la distribución de enzimas óptima para distintos valores de las constantes de difusión D_X y D_P , en el caso simétrico y asimétrico. La figura 5.14A muestra que para la estrategia asimétrica, el costo total en proteínas crece cuando los metabolitos no pueden cruzar la membrana con facilidad, *i.e.* para valores bajos de D_X y D_P (ver la superficie amarilla). Por otra parte, la distribución asimétrica se vuelve óptima para valores de la difusión lo suficientemente altos, en cuyo caso la asimetría de la enzima A_X se aproxima al valor -1 (figura 5.14B). En estas condiciones, la ruta metabólica queda distribuida entre ambas células, con la primera enzima (E_S) operando en una célula y la segunda (E_X) en la otra, lo que da origen a la complementación metabólica (ver figura 5.14C). Además, la asimetría completa en la primera enzima, *i.e.* $A_S=1$, implica en todos los casos la excreción del intermediario X por parte de la primera célula, el cual es a su vez importado por la segunda célula, *i.e.* $U_{X,1} = -U_{X,2} > 0$ (figura 5.14D). Notoriamente, la excreción de X se produce incluso en condiciones donde el producto P no puede difundir a través de las membranas, *i.e.* $D_P=0$ (figura 5.14E). En tales condiciones, la segunda población de endosimbiontes comenzará a aprovechar el recurso X , sin brindar nada a cambio, lo que podría representar cierto grado de parasitismo u oportunismo (al menos respecto a los costos de producción de P). Sin embargo, si el producto P es capaz de difundir, el segundo endosimbionte comparte dicho compuesto con el primero y de esta forma se establece la relación de complementación. Finalmente, si el valor de la constante de difusión D_P es muy elevado ($D_P > D_X$), las concentraciones de P alcanzan valores muy similares en ambas células, y de esta manera no se reduce el efecto de la inhibición, ni siquiera cuando la ruta está distribuida entre ambos

organismos. De manera que en esta situación, las células adoptan una estrategia “indiferente”, la cual corresponde a la distribución totalmente simétrica de las enzimas en la que cada organismo produce su propio P , sin que halla intercambio neto de metabolitos entre ellos.

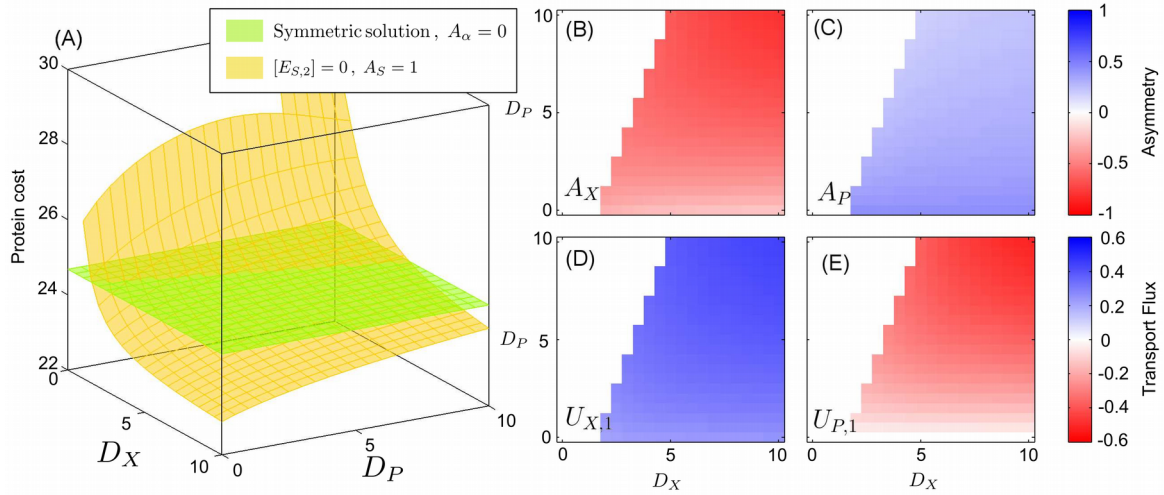


Figura 5.14. Soluciones óptimas (mínima concentración de enzima) para el caso de la inhibición acompetitiva. La figura muestra el costo en proteína en función de los parámetros de difusión D_X y D_P para un valor fijo de la constante de inhibición $1/K_I=10$. En (A) se representa el valor de la función de coste (Ec. 5.6) en función de D_X y D_P ; la superficie verde corresponde a la solución simétrica que se obtiene mediante la imposición de que ambas células tengan las mismas concentraciones de cada enzima; la superficie amarilla corresponde a la solución asimétrica que se obtiene fijando en cero la concentración de la enzima E_S en una de las células. (B) Coeficiente de asimetría para la enzima E_X ; (C) Coeficiente de asimetría para la enzima E_P ; (D) y (E) flujos de transporte de X ($U_{X,1}$) y P ($U_{P,1}$) en cada solución óptima. La escala de colores en B-C y D-E corresponde, al valor del coeficiente de asimetría de las enzimas y a la dirección de los flujos de transporte, respectivamente.

Evaluando el impacto de considerar un mecanismo de inhibición alternativo

En este apartado se busca extender los resultados sobre las condiciones necesarias para la emergencia de complementación metabólica evaluando el efecto de considerar un mecanismo de inhibición diferente al caso acompetitivo (Ec. 5.2). El objetivo es evaluar si los patrones de asimetría en la distribución de las enzimas en condiciones óptimas obtenidos previamente dependen de un mecanismo de inhibición específico o si son independientes del mismo. Para esto se considerará el mecanismo de inhibición competitiva, descrito por la siguiente ecuación:

$$V_S = k_{cat}[E_S] \cdot \frac{([S] - [X]/K_{eq})/K_M^+}{1 + [S]/K_M^+ + [X]/K_M^- + [P]/K_I} \quad (\text{Ec. 5.9})$$

La diferencia entre este mecanismo y el caso acompetitivo, es que en la inhibición competitiva el sustrato compiten por el sitio activo y esto afecta al grado de saturación de la enzima. En la inhibición acompetitiva se modifica la constante catalítica (Cornish-Bowden, 2013). Si bien ambos mecanismos implican un formulación diferente, los resultados que se obtienen son cualitativamente los mismos en ambos casos (ver figura 5.15).

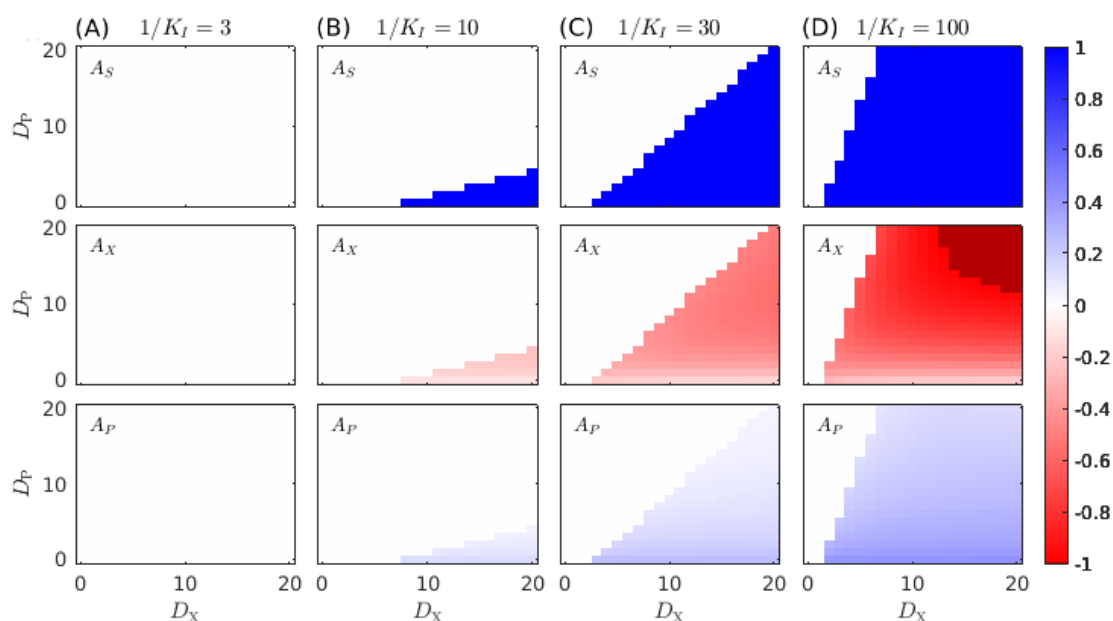


Figure 5.15. Coeficiente de asimetría A_α en distribuciones de enzimas óptimas en el caso de inhibición competitiva en la enzima. La figura representa distintas soluciones óptimas expresadas en términos del coeficiente de asimetría de cada enzima como función de los parámetros D_x y D_p . Las columnas corresponden a simulaciones con distintos valores en la constante de inhibición ($1/K_I=3, 10, 30, 100$). Las zonas de color oscuro para el caso de $1/K_I=100$ representan soluciones donde la ruta se encuentra completamente dividida entre ambas células, i.e. $[E_{S,2}] = [E_{X,1}] = 0$.

Si se comparan las figuras 5.13 y 5.15 se puede apreciar que las diferencias más notorias se dan a nivel de los rangos de parámetros en los que las soluciones asimétricas comienzan a emerger, así como la forma de la frontera del espacio de parámetros que separa las soluciones simétricas, de las asimétricas. Por ejemplo, para un valor de la constante inhibición $1/K_I=10$ en el caso competitivo, solo un rango muy reducido de valores de las constantes de difusión se obtienen soluciones óptimas asimétricas (figura 5.15B), mientras que en el caso de inhibición acompetitiva, dicho rango es mucho más amplio (figura 5.13B). De estos resultados se puede concluir que, si bien la presencia de la de inhibición por producto es una condición necesaria para la emergencia de complementación metabólica, este fenómeno no depende de un mecanismo específico.

Efecto de la variación de las demandas en los patrones de complementación

En este apartado se analizan las soluciones óptimas en la distribución de enzimas que se obtienen al considerar: i) la demanda de producto P por parte del hospedador ($J_{P,0} > 0$); ii) que las poblaciones de bacterias tengan demandas de P diferente ($J_{P,1} \neq J_{P,2}$); y iii) que las relaciones entre los tamaños poblaciones relativos de cada endosimbionte sean diferentes ($n_1 \neq n_2$). Para evaluar el efecto de la demanda del hospedador, se añadió una restricción sobre el flujo $J_{P,0}$, equivalente a las restricciones impuestas sobre los flujos de consumo de cada endosimbionte (Ec. 5.7). El valor de la demanda se varía entre cero y uno, y para cada valor se minimiza la concentración total de enzima. Los resultados del experimento muestran, que si el efecto de la inhibición es considerable, la distribución óptima de las enzimas exhibe complementación metabólica (figuras 5.16A-D). Además, los resultados solo cambian ligeramente como consecuencia de la demanda del hospedador. En efecto, al variar la demanda $J_{P,0}$ de cero (que corresponde al caso analizado al principio del apartado) hasta uno las distribución normalizada de la enzimas no cambia (figuras 5.16B-D). Claramente, el incremento en la demanda $J_{P,0}$ trae consigo un aumento en la concentración total óptima de enzimas, ya que la demanda implica un incremento necesario en la producción de P . No obstante, los patrones en la asimetría no exhiben cambios cualitativos.

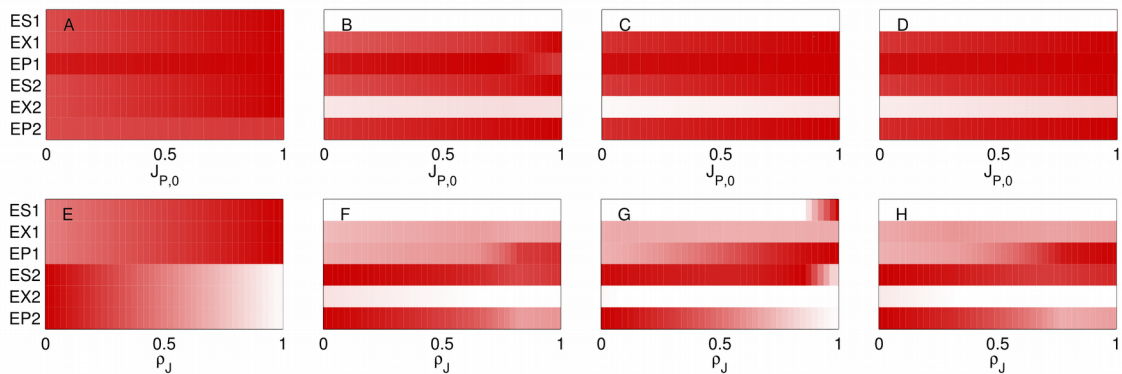


Figura 5.16. Asimetría de las enzimas en soluciones óptimas variando la demanda del hospedador y los endosimbiontes. Cada panel representa los valores de las concentraciones de enzima, normalizados por el valor máximo en la solución óptima. En el eje x se representa el valor del parámetro de control. De (A-D) se representan las soluciones óptimas para distintos valores de $J_{P,0}$. De (E-H) se representan las soluciones para distintos valores del coeficiente de asimetría $\rho_J = (J_{P,1} - J_{P,2}) / (J_{P,1} + J_{P,2})$, pero sin considerar la demanda del hospedador (i.e. se mantuvo $J_{P,0} = 0$). Los parámetros empleados en los experimentos son los siguientes: (A,E) $K_I = 0$, $D_X = 2$, $D_P = 2$; (B,F) $K_I = 1/25$, $D_X = 2$, $D_P = 2$; (C,G) $K_I = 1/25$, $D_X = 10$, $D_P = 10$; (D,H) $K_I = 1/25$, $D_X = 20$, $D_P = 4$. En todos los casos se representan las soluciones en las que $A_5 \geq 0$.

Debido al hecho de que hay dos especies de endosimbiontes, es esperable que cada población de bacterias tenga demandas diferentes del producto, debido a diferencias en la composición de su biomasa. De manera que es de interés evaluar como se ven afectados los patrones óptimos en la distribución de las enzimas frente a demandas diferentes por parte de cada población de endosimbiontes, i.e. $J_{P,1} \neq J_{P,2}$. Para analizar el efecto del cambio relativo en las demandas, los se introduce un nuevo coeficiente de asimetría que es definido sobre los flujos de demanda $\rho_J = (J_{P,1} - J_{P,2}) / (J_{P,1} + J_{P,2})$. Al calcular los patrones óptimos variando el coeficiente ρ_J entre cero

y uno, los resultados muestran como la asimetría en la demanda de P promueve, de forma más acentuada, las soluciones óptimas con asimetría en la distribución de las enzimas, *i.e.* la complementación metabólica (figura 5.16E-H). En efecto, para valores suficientemente altos de la constante inhibición ($K_i > 1/25$), la complementación metabólica en la que la ruta está completamente dividida entre ambas células (*i.e.* $[E_{S,1}] = [E_{X,2}] = 0$), emerge de forma robusta para un amplio rango de valores de ρ_j (figura 5.16F-H). En particular, la solución con $A_S < 0$ ($[E_{S,1}] = 0$) es óptima siempre que $\rho_j > 0$. Es importante destacar que, debido a la forma en que se ha definido ρ_j , al considerar el rango de valores $[0-1]$, sólo se están considerando los casos $J_{P,1} > J_{P,2}$; es decir, que para $\rho_j > 0$, la simetría entre las dos soluciones óptimas $A_S < 0$ y $A_S > 0$ se rompe. Sin embargo, la simetría se da respecto a $\rho_j = 0$, de forma que $\rho_j < 0$ implica que en la solución óptima $A_S > 0$.

Finalmente, para completar el análisis se considera la posibilidad de que ambas poblaciones de bacterias tengan tamaños relativos diferentes, *i.e.* $n_1 \neq n_2$. En este caso, dado que las poblaciones ya están normalizadas, basta con modular n_1 (con $n_2 = 1 - n_1$) para evaluar su efecto en los patrones óptimos. Los resultados indican que a medida que n_1 se aleja de 0,5 (con $n_2 = 1 - n_1$), la distribución asimétrica de las enzimas se ve favorecida. Este resultado indica que desde un punto de vista cualitativo, la asimetría en la distribución de las enzimas en las soluciones óptimas no parece depender de la relación entre los tamaños poblacionales. Sin embargo, para incluir el efecto de los tamaños poblacionales de forma más adecuada, se debería extender el modelo para las ecuaciones de crecimiento de cada población de forma que los tamaños fuesen variables y no parámetros del sistema.

Análisis del punto de corte en la complementación en un modelo con múltiples intermediarios

Como muestran los resultados de las simulaciones realizadas con el modelo, la emergencia de patrones de complementación metabólica está relacionados con la capacidad de difusión de los intermediarios de la ruta. Para considerar esta posibilidad se extendió el modelo añadiendo cinco nuevos intermediarios, X_1, \dots, X_5 , y las cinco enzimas EX_1, \dots, EX_5 , que catalizan las reacciones de la ruta hipotética de biosíntesis. Cada enzima EX_i , con $i=1..5$, cataliza la reacción reversible $X_i \rightarrow X_{i+1}$, excepto en el caso de EX_5 que cataliza $X_5 \rightarrow P$. De esta, forma la versión extendida del modelo permite analizar los distintos patrones de complementación que emergen como consecuencia de distintos perfiles de permeabilidad en los intermediarios de la ruta. Los experimentos incluyen los casos donde un único intermediario y el producto pueden difundir, y casos más complejos donde se permite que difundan diversos intermediarios y el producto final.

En el primer caso donde un único intermediario es permeable, los resultados predichos por el modelo extendido son consistentes y análogos a los obtenidos para el caso del modelo inicial que solo incluye un intermediario (figuras 5.17A-D). Es decir, que en los distintos casos en que sólo se permite el intercambio de un único intermediario y de P , la solución óptima frente a la presencia de inhibición es la complementación metabólica. Considerando únicamente las soluciones de tipo $A_S > 0$, el resultado cualitativo es siempre el mismo: la primera célula convierte el sustrato S en el intermediario más permeable, el cual es excretado (casi completamente), e importado por la segunda célula, la cual completa la síntesis del producto P que es en última instancia compartido

con la primera célula.

Sin embargo, si se consideran escenarios más complejos donde múltiples intermediarios pueden difundir a través de la membrana de forma equivalente, se obtienen una gama más amplia de comportamientos. Por ejemplo, en el caso extremo en que todos los metabolitos son permeables a la membrana, la estrategia que minimiza la concentración total de enzima exhibe cierto grado de asimetría en casi todas las enzimas (figuras 5.17EF). Si bien ambas células mantienen cierta concentración de cada enzima, en la primera enzima (E_5) y en la enzima que produce el producto P la asimetría es completa, lo que muestra que si el efecto de la inhibición es considerable, la estrategia óptima para minimizar la concentración total del enzima implica que la enzima que está sujeta a inhibición por producto y la enzima que cataliza la síntesis de dicho producto deben situarse en compartimentos diferentes. Esto puede ser una de las causas que promueve la complementación en el caso organismos o consorcios que deben sobre-producir un compuesto esencial, cuyo flujo se encuentra regulado mediante un mecanismo de inhibición por producto.

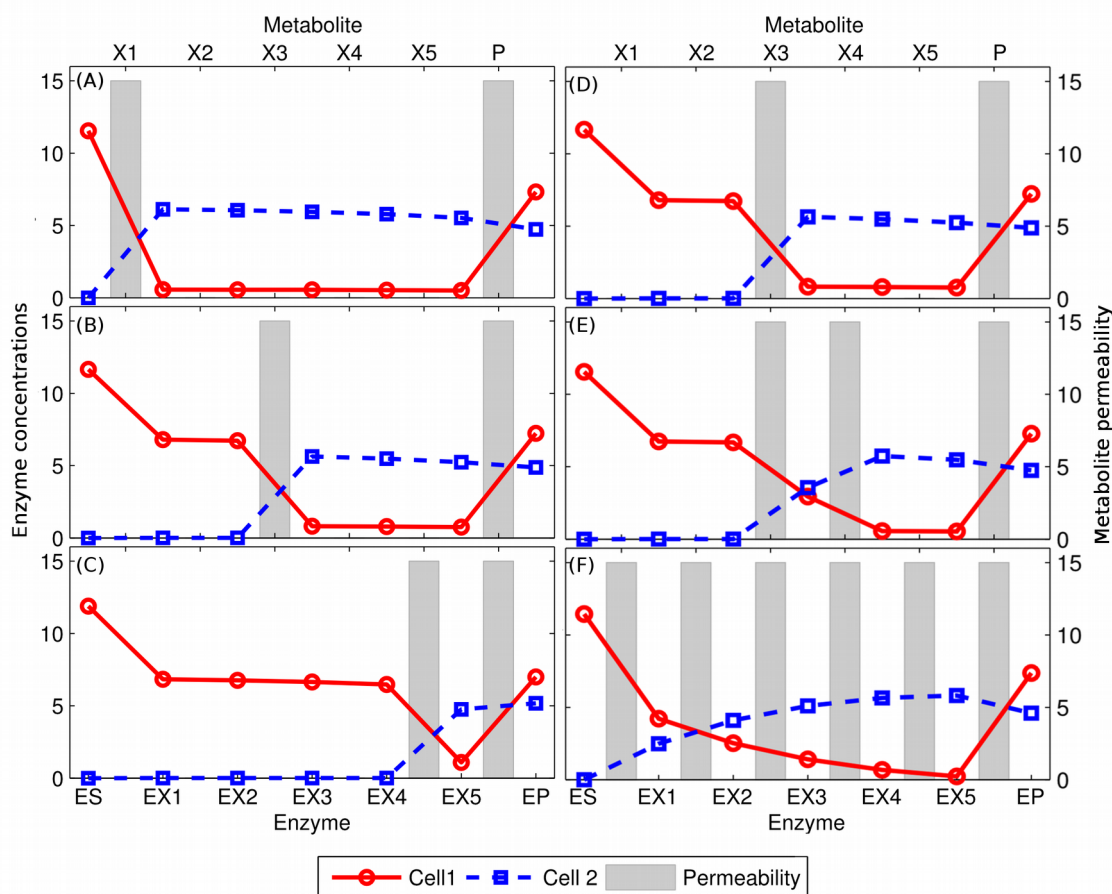


Figura 5.17. Distribuciones óptimas para el caso de distintos perfiles de permeabilidad en el modelo extendido. Las curvas rojas y azules representan las concentraciones óptimas de enzima en los dos tipos de células; las barras grises corresponden a las constantes de difusión ($D_i=15$) de cada uno de los intermediarios; en el caso de los intermediarios sin barra gris, se asume una constante de difusión mínima de $D=10^{-3}$. En cada panel (A-F) se representa la solución óptima frente a distintos perfiles de permeabilidad; P se supone permeable en todos los casos y los otros parámetros de la simulación son $K_i=1/10$ y $[S]=20$.

5.3.7 El papel de la permeabilidad

Los experimentos de reducción genómica *in-silico* presentados en el apartado 5.3.5 indican que la forma en que se organiza la complementación metabólica desde el punto de vista puramente estequiométrico, es un factor relevante en la eficiencia de la producción del triptófano. Además, los resultados también indican que el diseño que exhibe el consorcio del áfido del cedro es una solución muy próxima al óptimo que se alcanza al considerar el caso del ancestro hipotético en el cual la ruta está codificada en forma redundante en ambos endosimbiontes; dicha solución tiene la peculiaridad de que es mínima, *i.e.* que solo requiere una copia de cada uno de los genes. Por otra parte, las predicciones del modelo cinético indican, que en presencia de un mecanismo de regulación de tipo inhibición por producto, la complementación metabólica emerge como la estrategia que requiere la concentración total de enzimas mínima necesaria para sostener un flujo de producto que abastezca a todos los miembros del consorcio, incluyendo al hospedador. Además, este modelo también predice que si los intermediarios son intercambiados vía difusión, el punto de corte óptimo (*i.e.* el intermediario intercambiado) se corresponde con el intermediario más permeable de la ruta. Esta predicción puede ser contrastada con el caso del triptófano en el consorcio de *C. cedri*, para lo cual primero es necesario medir o estimar la permeabilidad de la membrana a los distintos intermediarios. Utilizando distintos predictores de permeabilidad basados en propiedades fisicoquímicas de los metabolitos, se estimó la permeabilidad de los intermediarios de la ruta de biosíntesis del triptófano (ver apartado 5.2.5 de la sección de Materiales y Métodos). Empleando dichos descriptores se identificaron aquellos intermediarios para los que se predice alta permeabilidad. Los resultados, resumidos en la tabla 5.7 muestran que de todos los intermediarios de la ruta del triptófano, el antranilato, el corismato y el indol parecen ser los más permeables, y como consecuencia, son los candidatos más probables a ser el metabolito intercambiado en la complementación de la ruta (los estimadores indican que los demás intermediarios no son en principio muy permeables).

Tabla 5.7. Propiedades fisicoquímicas, sus estimadores y predictores de la permeabilidad de la membrana a los intermediarios de la ruta del triptófano. Para cada compuesto se presenta su masa molecular (MM), un estimador del coeficiente de partición octanol/agua (AlogP), la superficie topológica polar accesible (TPSA), y la predicción de dos clasificadores basados en reglas (1PR y 3PR). Para detalles sobre los cálculos ver el apartado 5.2.5, y para los resultados completos ver tabla S.5.6 del apéndice B.

Paso	Metabolito	MM (g/mol)	AlogP	TPSA (Å ²)	1PR*	3PR*
1	Corismato	224,17	0,41	109,72	MA	A
2	Antranilato	136,13	0,24	66,15	MA	A
3	N-(5-fosforibosil)-antranilato	346,21	-0,54	184,08	MB	MB
4	1-(o-carboxifenilamino)-1-deoxiribulosa-5-P	346,21	-0,6	191,92	MB	MB
5	(1S,2R)-1-C-(indol-3-yl)glicerol 3-P	285,19	0,12	138,48	MB	MB
6	Indol	117,15	2,29	15,79	A	A
7	L-triptófano	204,22	-0,52	83,56	MA	MB

* Abreviaciones: alta (A); media alta (MA); media baja (MB)

En primer lugar, el indol es una molécula empleada por organismos, como *E. coli*, para la

señalización (Martino et al., 2003), y que se sabe puede permear fácilmente a través de una bicapa lipídica (Piñero-Fernandez et al., 2011). Sin embargo, dado que se sabe que el metabolito intercambiado entre *B. aphidicola* BCc y *S. symbiotica* BCc es el antránilato y no el indol, se podría preguntar entonces, si existe alguna razón para esta opción, o si ambos intermediarios son soluciones “equivalentes”. Una posible respuesta a esta pregunta se puede encontrar en la estructura del complejo triptófano sintasa. Este es un tetrámero $\alpha_2\beta_2$, donde la subunidad α , codificada por *trpA*, cataliza la síntesis del indol a partir del indol-3-glicerol-fosfato; a continuación, el indol es canalizado de la subunidad α a la subunidad β , codificada por el gen *trpB*, donde es empleado para sintetizar el triptófano (Dunn, 2012; Miles, 2013). Como consecuencia de la canalización entre ambas subunidades, el indol no tiene posibilidad de difundir en forma libre fuera de la célula. Además, se ha mostrado de forma experimental que si la subunidad α no está presente, la actividad catalítica de la subunidad β (*i.e.* la conversión de indol en triptófano), se reduce de 1-2 órdenes de magnitud (Kirschner et al., 1991). Esto implica que separar las subunidades del complejo triptófano sintasa tiene consecuencias muy negativas en la eficiencia de la catálisis. Esta restricción estructural conlleva a que el indol no pueda difundir libremente y por lo tanto no puede ser el intermediario de la complementación metabólica. Entonces, si el indol no es un candidato factible restan el corismato y el antranilato. El primero carece de sentido como punto de corte ya que no resuelve el problema de separar la antranilato sintasa y la síntesis de triptófano en compartimientos diferentes. Finalmente, el antranilato no presenta la restricción del indol, es un intermediario predicho como muy permeable y permite separar en compartimientos diferentes la enzima regulada y la síntesis del metabolito que actúa como inhibidor. En base a lo anterior se puede deducir que el antranilato posiblemente sea el punto más favorable para establecer una complementación metabólica en la ruta del triptófano.

Finalmente y a modo de corolario, se discutirá otro caso paradigmático de complementación metabólica en endosimbiontes; este es el caso de las actividades transaminasa que se han perdido en los distintos linajes de *B. aphidicola* (Hansen and Moran, 2011; Macdonald et al., 2012; Poliakov et al., 2011), así como en otros endosimbiontes tales como *Tremblaya* y *Portiera* (Baumann et al., 2006; McCutcheon and von Dohlen, 2011; Zelezniak et al., 2015; Zientz et al., 2004). Las transaminasas son actividades enzimáticas de importancia crítica en la síntesis de los aminoácidos; en el caso de los aminoácidos ramificados (*i.e.* Leu, Ile y Val), así como los aromáticos a excepción del triptófano (*i.e.* Phe y Tyr), la transaminación es la última actividad de la ruta. Además, las rutas de los tres aminoácidos ramificados comparten la misma actividad enzimática (EC 2.6.1.42) y lo mismo ocurre en el caso de los dos aromáticos mencionados (que comparten la actividad EC 2.6.1.57). En muchos linajes de endosimbiontes, como la mayoría de las *Buchnera*, se han perdido los genes que codifican para las dos transaminasas antes mencionadas, mientras que se han retenido los genes que codifican las demás actividades. Datos de transcriptómica y proteómica indican que ambas transaminasas se encuentran sobre-expresadas en los bacteriocitos (Hansen and Moran, 2011; Poliakov et al., 2011). Además existen pruebas experimentales que, de forma directa, muestran que es el hospedador quien cataliza estas reacciones (Russell et al., 2013).

Estos eventos de complementación no implican a dos endosimbiontes, sino que la relación es entre un endosimbionte y su hospedador. Aún así, estos casos cumplen con las suposiciones del

modelo, *i.e.* la demanda de un metabolito, la ausencia de transportadores y la inhibición por producto¹³. Por lo tanto, en este caso también se puede verificar si el punto de corte de las complementaciones metabólicas coincide con el intermediario más permeable de la ruta. Empleando los predictores de permeabilidad ya descritos, se analizaron los perfiles de permeabilidad de los distintos intermediarios en las rutas correspondientes. Los resultados indican que en todos los casos, el punto de intercambio de la complementación metabólica corresponde con el intermediario predicho como el más permeable (ver [tabla S.5.6](#) del apéndice B); por ejemplo, en el caso de las rutas de síntesis de la valina y la leucina, los intermediarios predichos como más permeables son el 4-metil-2-oxopentanoato y el 3-metil-2-oxobutanoato, respectivamente. En ambos casos, el intermediario se corresponde con el metabolito intercambiado entre *Buchnera* y su hospedador (Russell et al., 2013). El mismo patrón se repite en el caso de la ruta de síntesis de la fenilalanina. Este resultado sugiere, por un lado, que los puntos donde se producen eventos de complementación suele coincidir con un intermediario permeable; por otra parte, que la complementación metabólica en las rutas de los aminoácidos ramificados, así como en el caso del a fenilalanina podrían haber emergido como una solución o estrategia para poder responder al incremento en la demanda de dichos aminoácidos.

5.4 Discusión

Reconstrucción y análisis metabólico de un consorcio endosimbionte

La reconstrucción y el análisis metabólico de *B. aphidicola* BCc y *S. symbiotica* SCc, endosimbiontes del áfido del cedro ha permitido, en primer lugar, realizar una revisión y refinamiento de la anotación de los genomas de dichos organismos. En efecto, como consecuencia de la revisión manual de las redes, realizada empleando la aproximación de los módulos inconexos, fue posible corregir errores en las anotaciones de distintas actividades enzimáticas, e identificar genes metabólicos previamente no anotados. Además, las simulaciones realizadas con los modelos revisados han permitido refinar las predicciones *in-silico* sobre las condiciones de crecimiento y capacidades metabólicas de los endosimbiontes respecto a las previamente inferidas directamente a partir de la anotación de los genomas de dichos organismos.

Por otra parte, el análisis de redes metabólicas muy reducidas, como es el caso de *B. aphidicola* BCc pone de manifiesto problemas metodológicos como el caso del balance de protones. No es fácil encontrar en la literatura menciones a este problema, aunque existe un apartado dedicado a dicho problema en la publicación de la reconstrucción del GEMM de *E. coli* K12 iJR904 (Reed et al., 2003). En dicho estudio, en el que se realizó un análisis de sensibilidad para evaluar el efecto del balance global de los protones sobre las predicciones de crecimiento, del modelo los autores encuentran que el balance neto de protones puede ser positivo o negativo dependiendo de la fuente de carbono con que se simule el crecimiento. De acuerdo con lo anterior, si el balance neto de protones fuese positivo el crecimiento del organismo contribuiría a disminuir el pH del ambiente, mientras que si ese fuese negativo, el ambiente se volvería más básico. Sin embargo, los autores concluyen que no se dispone de datos experimentales para validar las predicciones del modelo.

¹³ Al igual que en el caso del triptófano, la síntesis de estos aminoácidos se encuentra regulada a diferentes niveles que usualmente incluyen un mecanismo de inhibición por producto (White et al., 2012).

Más allá del estudio mencionado, el problema del balance de protones en los modelos metabólicos reconstruidos a escala genómica no ha recibido mayor atención, posiblemente debido a que al considerar redes metabólicas de mayor tamaño, el efecto que dicha restricción impone sobre las capacidades de la red se “diluya”. Pero al considerar el caso de endosimbiontes, o en general organismos con metabolismos muy reducidos, este tipo de problemas se vuelven evidentes. Además, como se vio en el estudio realizado con *iBCc98*, si la restricción es tratada de una forma u otra, las predicciones del modelo pueden variar notoriamente. Por ejemplo, si el efecto del balance de los protones predicho *in-silico* tuviese lugar *in-vivo*, la versatilidad del metabolismo de *B. aphidicola* BCc se vería considerablemente reducida, ya que procesos metabólicos presumiblemente independientes, como la síntesis de ATP y el ciclo de los folatos, quedarían acopladas dando lugar a sobre-producciones de aminoácidos acoplados al crecimiento. Si bien este diseño sería claramente desventajoso para un organismo de vida libre, en el caso de un endosimbionte, cuya principal función es la sobre-producción de aminoácidos, esta red metabólica podría presentar ventajas al sistema insecto-simbiontes. En efecto, un comportamiento similar ha sido también descrito y discutido por Calle-Espinosa y colaboradores en el caso del endosimbionte de la mosca blanca del tabaco, *P. aleyrodidarum* (Calle-Espinosa et al., 2016). En dicho estudio, las predicciones obtenidas con el modelo metabólico del endosimbionte *iJC86*, indican que el crecimiento del organismo se encontraría acoplado a la sobre-producción de diversos aminoácidos. Los autores concluyen que dichos acoplamientos son consecuencia de las limitadas capacidades energéticas del endosimbionte.

Si bien los fenómenos difieren en los detalles moleculares, es notoria la similitud en cuanto a tratarse de sistemas similares hasta cierto punto (*i.e.* insecto albergando endosimbiontes que le proveen de aminoácidos esenciales) pero con historias evolutivas independientes. Este tipo de fenómenos podría jugar un papel relevante en la evolución de organismos endosimbiontes. Por el contrario, el efecto del balance de protones podría representar un artefacto, consecuencia de falta de información en la formulación del modelo, por ejemplo el desequilibrio de protones podría acoplarse a procesos de transporte en contra de gradiente. Lo más probable es que en el caso de los endosimbiontes el problema del balance de protones tenga efectos en la fisiología de la célula, pero que también aún falten piezas para entender este fenómeno con mayor detalle.

Por otra parte, las simulaciones con *iSCc236* indican que *S. symbiotica* SCc presenta cierta semejanza con *Blattabacterium*, en la forma en que dos rutas de biosíntesis se encuentran completadas con el hospedador. Como ya fue mencionado, en ambos sistemas el endosimbionte ha perdido los genes que codifican las primeras actividades de las rutas de síntesis de la coenzima A y del sirohemo y, como consecuencia, los respectivos hospedadores deben proveer a sus endosimbionte de los intermediarios necesarios para completar la síntesis de los cofactores. Esto podría representar un caso de convergencia que posiblemente se deba a una combinación de factores que incluyan las restricciones de permeabilidad de los componentes. Además, las simulaciones realizadas con el modelo reconstruido del consorcio (*iBSCc*) han permitido reproducir los casos de complementación metabólica ya descritos en la literatura, pero también predicen un posible caso de complementación metabólica previamente no descrito, a saber la síntesis del THF, el cual sería sintetizado por *S. symbiotica* SCc a partir del shiquimato provisto por *B. aphidicola* BCc.

Emergencia de complementación metabólica en consorcios endosimbiontes

La emergencia de la complementación metabólica es un fenómeno complejo y poco comprendido desde un punto de vista teórico. Para poder entender algunas propiedades de dicho fenómeno se realizaron algunos estudios siguiendo dos aproximaciones complementarias. Por un lado, se utilizó el modelo reconstruido a escala genómica del consorcio endosimbionte *iBSCc* para simular en detalle las propiedades estructurales en el diseño de la complementación metabólica en el caso de la ruta de biosíntesis del triptófano. El estudio consistió en comparar todos los posibles diseños de la complementación en dicha ruta, empleado como criterios la maximización de la producción de meta-biomasas y la maximización de la producción de triptófano. Si bien no se encontraron grandes diferencias con respecto a la eficiencia en la producción máxima de meta-biomasa, los distintos diseños sí exhiben gran variabilidad en el flujo máximo de producción de triptófano. Esto indica que la forma en que se distribuye o divide una ruta entre dos organismos tiene efectos directos sobre la eficiencia del proceso. Notablemente, los resultados muestran que el diseño de la complementación en la ruta del triptófano exhibido por el consorcio del áfido del cedro es mínimo, en término del número de genes requeridos, y permite sostener un flujo de producción de triptófano que es tan solo un 5% menor al predicho para el caso del hipotético consorcio ancestral en que ambos simbiontes codifican la ruta completa. Además, los restantes diseños mínimos en el número de genes exhiben una disminución en el valor máximo de producción de triptófano de entre 25% y 50%, respecto a lo predicho para el caso del consorcio ancestral. Los resultados antes expuestos indican que la complementación que exhibe el consorcio de *C. cedri* es la configuración de genes mínima que permite alcanzar un flujo de triptófano cuasi-óptimo respecto al ancestro hipotético. Además de la estructura, es muy posible que otros factores condicionen también la eficiencia de la complementación metabólica. Por ejemplo, en el caso de la ruta del triptófano en el consorcio de *C. cedri*, se ha encontrado que la enzima de ruta inhibida por el triptófano tiene un alto grado de conservación en el sitio de unión del inhibidor. Esto sugiere que la inhibición se encuentra activa.

La consideración de los aspectos cinéticos, como la inhibición antes mencionada, no es abordable a escala genómica, debido a la complejidad asociada a tal modelo, además del elevado número de parámetros y la falta de datos experimentales para estimarlos. Por este motivo, para poder estudiar los efectos de la cinética sobre el fenómeno de la complementación metabólica, se desarrolló un modelo más sencillo que solo retuviera las características fundamentales necesarias para describir el fenómeno. Empleando dicho modelo fue posible estudiar de forma más general, las posibles ventajas y desventajas de la complementación, e identificar condiciones necesarias para su emergencia. Partiendo de la suposición que el sistema evolucionará hacia el estado en donde la concentración total de enzima sea mínima y que permita satisfacer la demanda de P , se ha encontrado que existen ciertas condiciones necesarias para la emergencia de complementación metabólica. En primer lugar, el producto de la ruta debe ser esencial y la ruta encontrarse codificada de forma redundante (*i.e.* que más de un miembro del consorcio codifica las actividades enzimáticas). Además, la presencia de inhibición por producto en la ruta es clave para que la complementación implique una reducción en la concentración total de enzima, respecto al caso donde la ruta es redundante. Esto se debe a que la presencia de inhibición de P sobre la enzima E_s tiene como consecuencia que para incrementar el flujo de producción de P haya que elevar en gran

medida la concentración de la enzima del paso inhibido E_S . Cuando la ruta se encuentra repartida en dos organismos se necesita menos cantidad de enzimas para sostener una misma demanda que en el caso en que cada organismo exprese la ruta completa .

Otra condición necesaria para que sea posible la complementación en una ruta metabólica es que al menos uno de los intermediarios y el producto final de la ruta se puedan intercambiar entre los miembros del consorcio. Además, dado que no suele haber transportadores asociados a los intermediarios de una ruta, el compuesto intercambiado deberá presentar ciertas características fisicoquímicas que le permitan difundir a través de las múltiples membranas celulares que deberá a travesar. De acuerdo con el modelo, si algún intermediario de la ruta exhibe tales características, existe la posibilidad de que emerja complementación en dicha ruta metabólica. Si se contrasta este resultado con el caso del triptófano, se encuentra que el punto donde se divide la ruta del triptófano coincide con el intermediario más permeable de la ruta, que permite separar la enzima inhibida por el triptófano de la enzima que lo produce. Esto muestra que la permeabilidad de la membrana a los intermediarios impone restricciones fisicoquímicas sobre la emergencia de la complementación. La presencia de fenómenos como la canalización entre dos enzimas de una ruta puede suponer restricciones adicionales a una posible complementación metabólica. Las predicciones del modelo mostraron ser robustas a distintas variaciones introducidas en la formulación del mismo. Por ejemplo, distintos mecanismos de regulación en la enzima inhibida producen resultados muy similares. Además, el análisis de sensibilidad realizado sobre los flujos de demanda, así como la influencia de los tamaños poblacionales, muestra que las predicciones del modelo no dependen de estos parámetros y por tanto son, hasta cierto punto, generales. Sin embargo, para poder considerar de forma más rigurosa el efecto de los tamaños poblacionales, el modelo debería ser extendido para incluir las ecuaciones de crecimiento de las poblaciones.

Consideraciones finales

Los resultados y predicciones obtenidos con los modelos metabólicos a escala genómica de los endosimbiontes, además de su valor intrínseco, muestran la utilidad que esta clase de modelos tienen como herramienta para refinar las anotaciones genómicas y el propio. En esta línea, los modelos también permiten hacer predicciones más rigurosas sobre las interacciones nutricionales entre los miembros de un consorcio o comunidad microbiana. Por otra parte, los modelos pueden servir como marco teórico para interpretar y entender patrones complejos de co-evolución, como la complementación metabólica. Además, estos permiten generar hipótesis concretas, que pueden ser verificadas de forma experimental. Finalmente, si bien los modelos a escala genómica son una herramienta muy valiosa, los modelos más sencillos permiten analizar fenómenos complejos desde un punto de vista más amplio y, por lo tanto, pueden dar lugar a conclusiones más generales.

Capítulo 6

6 Construcción y análisis de Metamodelos

6.1 Introducción

Más allá de los avances recientes en la generación automática de redes metabólicas, la obtención de modelos metabólicos reconstruidos a escala genómica (GEMM) precisos y consistentes con resultados experimentales, requiere aún de una intensa supervisión por parte de expertos. El proceso mediante el cual se resuelven las inconsistencias presentes en GEMM es referido como revisión o refinado de redes (Thiele and Palsson, 2010). La revisión de un GEMM implica la identificación y resolución de las inconsistencias presentes en la formulación del modelo. Como se ha discutido en los capítulos previos, las inconsistencias suele ser consecuencia de errores en la definición de las tablas de asociación gen-proteína-reacción (GPR). La detección de inconsistencias en un GEMM se puede realizar de forma exhaustiva empleando el CBM. Una vez que se han identificado las inconsistencias, el refinado consiste en un proceso de toma de decisiones en el se deben encontrar las reacciones candidatas para completar los distintos *gaps*.

El problema de predicción *in-silico* de reacciones candidatas para completar *gaps* en un GEMM es referido como *gap-filling*. Por una lado, se han propuesto aproximaciones con un enfoque genómico que se centran en la detección de genes candidatos a codificar las actividades ausentes (Green and Karp, 2004; Vitkin and Shlomi, 2012). Por otra parte, se han propuesto diversos métodos basados en el CBM que plantean el *gap-filling* en término de problemas de optimización (Brooks et al., 2012; Kumar et al., 2007; Kumar and Maranas, 2009; Latendresse, 2014; Reed et al., 2006b; Vlassis et al., 2014). Estas aproximaciones permiten identifica conjuntos de reacciones que permiten completar los *gaps* presentes en un GEMM. En general, los métodos mencionados requieren de una base de datos metabólica donde buscar reacciones candidatas para rellenar o completar los *gaps*. Sin embargo, en las aproximaciones basadas en el CBM y que emplean métodos de optimización, la propia base de datos metabólica es tratada como una red metabólica global. De esta manera el GEMM de un organismo que se está reconstruyendo es tratado como si fuese una subred embebida en el metamodelo (Thiele et al., 2014). Además, dado que las bases de datos metabólicas incluyen, por lo general, reacciones que se expanden a través de todo el árbol de la vida, estas pueden ser consideradas metamodelos del metabolismo (Pah et al., 2013).

Del mismo modo que ocurre en modelos metabólicos particulares, el estudio de la consistencia de un metamodelo, mediante el uso del CBM, puede permitir la detección de inconsistencias estructurales y/o termodinámicas, así como elucidar sus causas. La presencia de cualquier clase de inconsistencia en un metamodelo se propagará a cualquier GEMM que se derive a partir de él, en forma análoga a como los errores de anotación se pueden propagarse al realizar transferencia de información por homología de secuencia (Gilks et al., 2002; Valencia, 2005). Por estos motivos, el análisis de consistencia de un metamodelo es relevante para poder reconstruir GEMM consistentes. Por ejemplo, Pah y colaboradores han analizado las propiedades topológicas de una red metabólica global inferida a partir de la base de datos KEGG (Pah et al., 2013). En dicho estudio, los autores muestran como la resolución de inconsistencias en la red global permite mejorar la reconstrucción de las redes de organismos individuales. Sin embargo, en dicho estudio el análisis de la red global está basado en la representación del metabolismo como un grafo. Esta representación simplifica la estructura del metabolismo al no considerar la estequiometría y las restricciones que esta impone, lo cual en muchas situaciones conduce a resultados incorrectos (Arita, 2012, 2004; Brohée et al., 2008; Croes et al., 2006; Zhou and Nakhleh, 2011). No obstante, la misma idea podría realizarse aplicando el CBM en una red metabólica global, o metamodelo. El análisis de consistencia realizado de esta manera puede permitir la detección y resolución de inconsistencias de forma más rigurosa, ya que se consideran restricciones como el balance de masas o la reversibilidad de las reacciones. Además, técnicas como el análisis de acoplamiento de flujos (Burgard et al., 2004), se pueden utilizar sobre un metamodelo para extraer información estructural como acoplamiento estequiométrico entre reacciones como el caso de los subconjuntos enzimáticos o de reacciones, *i.e.* conjuntos de enzimas que deben operar juntas en cualquier estado estacionario (Pfeiffer et al., 1999; Poolman et al., 2007). Esta información de la organización del metabolismo a escala global puede ser de gran utilidad en la reconstrucción y revisión de modelos de organismos particulares. Por ejemplo, si a nivel de un metamodelo metabólico se detecta que un conjunto de reacciones involucradas en una ruta metabólica conforman un subconjunto enzimático, esta información puede ser utilizada para completar dicha ruta durante la reconstrucción del metabolismo de un organismo particular.

El objetivo principal del presente capítulo consiste en la reconstrucción y revisión de un metamodelo consistente del metabolismo bacteriano, esto es un modelo metabólico que incluya la mayor parte de la bioquímica descrita en bacterias. Un metamodelo con tales características puede emplearse para reconstruir y refinar modelos de organismos individuales. Además un metamodelo puede emplearse para estudiar propiedades generales de la organización de la bioquímica y evolución del metabolismo (Barve and Wagner, 2013; Hosseini et al., 2015; Samal et al., 2010). Para este fin, se seleccionó un conjunto de 130 modelos metabólicos de bacterias previamente publicados que fueron reconstruidos empleando la herramienta de reconstrucción Model SEED (Henry et al., 2010; Overbeek et al., 2005). La elección de dicho conjunto de datos se debe a dos motivos. Por un lado, el *pipeline* del RAST-SEED permite la reconstrucción del GEMM de un microorganismo de forma automática, a partir del genoma sin anotar (Overbeek et al., 2014). Además, uno de los pasos del *pipeline* implica *gap-filling* que produzca un primer borrador del modelo que prediga la producción de biomasa (Henry et al., 2010). Por otra parte, el conjunto de datos posiblemente represente la colección de GEMMs de mayor tamaño tamaño, producida en forma automática. Posiblemente por este motivo ha sido empleado en diversos estudios sobre la

organización del metabolismo en bacterias (Chiu et al., 2014; Freilich et al., 2011; Ish-Am et al., 2015; Yamada et al., 2012; Zelezniak et al., 2015).

Previo a la construcción del metamodelo se realizará un análisis de consistencia a gran escala sobre los 130 modelos con el fin de detectar errores recurrentes producidos durante la reconstrucción metabólica automática. A continuación, empleando los 130 GEMMs se reconstruirá una primera versión de metamodelo sobre la cual se realizó un análisis para detectar la presencia reacciones bloqueadas producidas por la presencia de *gaps* y *deadends* globales. Para realizar el análisis y el refinados se empleará la aproximación de módulos inconexos (ver sección 3.6). Además emplearán las siguientes bases de datos como referencia de la bioquímica conocida: SEED (Overbeek et al., 2005), KEGG (Wixon and Kell, 2000), MetacCyc (Caspi et al., 2014) y BiGG (Schellenberger et al., 2010). Posteriormente, el metamodelo se empleará como red de referencia para resolver reacciones bloqueadas y metabolitos *gap* en los GEMMs individuales, utilizando una versión modificada del algoritmo *fastcore* (Vlassis et al., 2014). Dicho proceso permitirá analizar el efecto de la consistencia del metamodelo en la predicción del algoritmo de *gap-filling*. Por último, se realizará un breve análisis del proceso de extensión de un metamodelo, mediante la incorporación de nueva información bioquímica.

6.2 Materiales y Métodos

6.2.1 Modelos metabólicos reconstruidos a escala genómica

A través del portal Model SEED (<http://seed-viewer.theseed.org/>) se obtuvieron 236 modelos metabólicos reconstruidos a escala genómica. El conjunto inicial fue filtrado como se explica a continuación: i) 41 modelos fueron descartados porque no se encuentran disponibles en formato SBML; ii) se descartaron dos modelos correspondientes a organismos no perteneciesen al dominio bacteria¹; Dado que el presente estudio se centra en el análisis de metabolismos bacterianos, los dos modelos de organismos no pertenecientes al dominio Bacteria fueron también excluidos del análisis; y iii) para aquellos modelos en que se dispone de más de una versión, se eliminaron las versiones antiguas (27 en total) de y se retuvieron las versiones más recientes. Como consecuencia del filtrado se redujo el conjunto inicial a un total de 166 modelos metabólicos correspondientes a bacterias de un amplio rango taxonómico (ver tabla 6.1).

El conjunto de 166 modelos seleccionados incluye a los 130 modelos previamente publicados por Henry *et al.* (2010), aunque estos presentan pequeñas diferencias con los originales que se detallan al final del apartado (ver tabla S.6.1B en el apéndice B). Además, 22 de estos GEMMs han sido refinados y validados empleando datos experimentales sobre esencialidad de genes y condiciones de crecimiento (Henry et al., 2010). Dado que los 130 modelos antes mencionados han recibido cierto grado de revisión adicional durante el proceso de reconstrucción, estos fueron empleados en la primera fase del estudio. Por otra parte, los 36 modelos restantes (no incluidos en la publicación original) fueron reservados para estudiar la extensión de un metamodelo (ver tabla S.6.2C).

1 Un arquea (*Methanosarcina barkeri* str. Fusaro) y un eucariota (*Ostreococcus tauri*)

Tabla 6.1. Distribución taxonómica de las bacterias incluidas en el conjunto de datos empleados en el análisis de consistencia y construcción del metamodelo. El código de colores indicado en la primera columna corresponde al código empleado en la publicación original (Henry et al., 2010). En la cuarta columna se indica el total de modelos pertenecientes a una clase, para los conjuntos de 130 y 36 modelos, respectivamente, comentados en el texto

Color	Phylum (o grupo)	Clase	#Modelos
	<i>Actinobacteria</i>	<i>Actinobacteria (clase)</i>	10 / 1
	<i>Bacteroidetes</i>	<i>Bacteroidetes</i>	5 / 2
	<i>Firmicutes</i>	<i>Bacilli</i>	21 / 11
	<i>Firmicutes</i>	<i>Clostridia</i>	5 / --
	<i>Firmicutes</i>	<i>Mollicutes</i>	3 / 2
	<i>Proteobacteria</i>	<i>Alphaproteobacteria</i>	18 / 1
	<i>Proteobacteria</i>	<i>Betaproteobacteria</i>	12 / 1
	<i>Proteobacteria</i>	<i>delta/epsilon subdivisions</i>	11 / --
	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	33 / 15
	<i>Spirochaetes</i>	<i>Spirochaetes (class)</i>	3 / 1
	<i>Thermotogae</i>	<i>Thermotogae (class)</i>	1 / --
	<i>Aquificae</i>	<i>Aquificae (class)</i>	1 / --
	<i>Chlamydiae</i>	<i>Chlamydiae</i>	2 / --
	<i>Chloroflexi</i>	<i>Dehalococcoidetes</i>	1 / --
	<i>Deinococcus-Thermus</i>	<i>Deinococci</i>	2 / --
	<i>Fusobacteria</i>	<i>Fusobacteria (class)</i>	1 / --
	<i>Cyanobacteria</i>	<i>Phrochlorales</i>	-- / 1
	<i>unclassified Bacteria</i>	<i>candidate division TGI</i>	1 / --

Todos los modelos empleados fueron descargados en formato SBML de la página web del SEED (<http://www.theseed.org>). Como se comentó previamente, para algunos de los modelos se descargaron diversas versiones, que luego en las etapas de filtrado fueron descartadas, reteniéndose únicamente la versión más reciente. En particular, se encontraron versiones más recientes para tres de los modelos previamente publicados por Henry et al. (2010). En forma adicional, se encontró que el modelos de *Thiomicrospira crunogena* XCL-2 con identificador Seed39765.1, no se encontraba disponible; sin embargo, empleando el nombre del organismo se encontró un modelo con un identificador diferente (Seed317025.3), el cual fue empleado en lugar del modelo que no estaba disponible. Finalmente, los modelos se agruparon de acuerdo al rango taxonómico, utilizando el mismo esquema de colores presente en la figura 2 de la publicación original (Henry et al., 2010) y descrito en la tabla 6.1 (para más detalles sobre la taxonomía de los organismos puede consultarse la tabla S.6.1 en el apéndice B). La clasificación difiere levemente de la propuesta en dicha publicación debido a que esta presentaba leves inconsistencias que fueron corregidas². En forma adicional, al comparar las definiciones de las ecuaciones de biomasa se encontró que algunos modelos comparten la misma definición. Los 166 modelos incluye un total de 157 ecuaciones de

2 Un organismo pertenecientes al grupo de Actinobacteria (verde claro) y otro a los Bacteroidetes (azul oscuro), habían sido agrupado con las Gammaproteobacterias (amarillo)

biomasa y los grupos de organismos que comparten la misma definición de biomasa son los siguientes: 1) *Mycoplasma genitalium* G-37 y *Mycoplasma pulmonis* UAB CTIP; 2) *Salmonella typhimurium* LT2, *Escherichia coli* K12 y *Vibrio cholerae* O1 biovar eltor str. N16961; 3) *Shewanella frigidimarina* NCIMB 400, *Shewanella halifaxensis* HAW-EB4, *Shewanella woodyi* ATCC 51908, *Shewanella amazonensis* SB2B, *Shewanella pealeana* ATCC 700345 y *Shewanella sediminis* HAW-EB3, *Shewanella* sp. MR-4 (ver [tabla S.6.3D](#)).

6.3 Resultados

6.3.1 Análisis de consistencia de modelos metabólicos de bacterias

Con los 130 GEMMs de bacterias correspondientes al conjunto de datos publicado por (Henry et al., 2010) se realizó un análisis de consistencia sobre cada uno de los modelos para detectar: i) el conjunto de reacciones bloqueadas; ii) el conjunto de metabolitos *gap*. En la [figura 6.1a](#), se representa el número de reacciones bloqueadas encontradas en cada modelo, respecto el número total de reacciones del modelo. Si bien se puede apreciar cierta dispersión en los porcentajes de reacciones bloqueadas, variando entre 19-50%, la dispersión es menor que la que fue descrita en la publicación original (Henry et al., 2010). Esto es debido al hecho de que en el presente trabajo se emplearon las versiones más recientes de cada uno de los modelos (ver [apartado 6.2.1](#)).

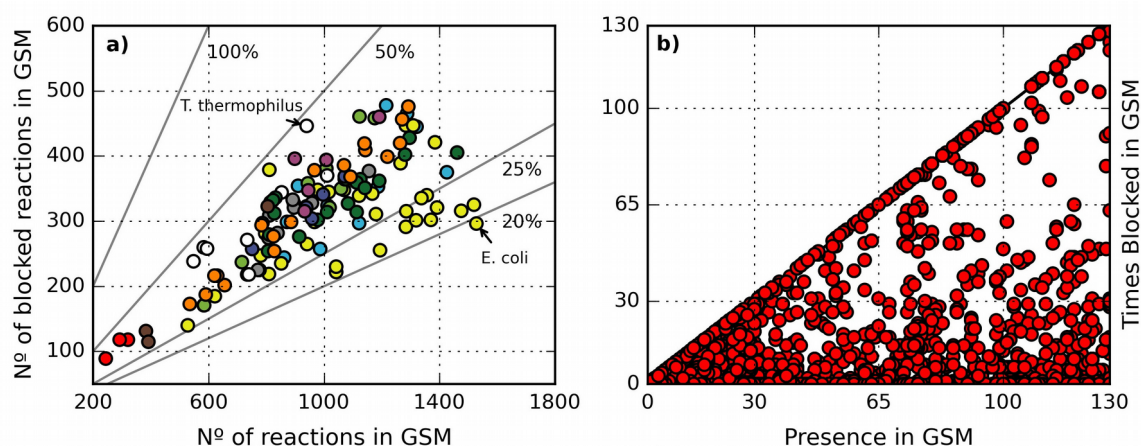


Figura 6.1. Resultados del análisis de consistencia realizado sobre 130 GEMMs. En a) se representa para cada GEMM, el número de reacciones bloqueadas encontradas respecto al número total de reacciones. Las líneas negras representan las iso-probabilidades (20%, 25%, etc). Los colores de los puntos se corresponden a los grupos taxonómicos de la figura 2-a de Henry et al. (2010), y que también se incluyen en la [tabla 6.1](#). En b) se representa para cada reacción, el número de modelos en los que fue encontrada bloqueada respecto al total de modelos en los que la reacción se encuentra presente. La diagonal principal indica las reacciones que se encontraron bloqueadas en todos los modelos en los que se encuentran presentes.

El modelo con el menor porcentaje de reacciones bloqueadas (19%) corresponde al GEMM de *E. coli* K12 (SEED ID: Opt83333.1) mientras que el porcentaje más alto (~47%) fue encontrado en el modelo de *Thermus thermophilus* HB8 (SEED ID: Seed300852.3). Más allá de la dispersión que presentan los resultados, es aún posible apreciar una clara tendencia entre el número de reacciones bloqueadas en un modelo y el tamaño del modelo (*i.e.* el total de reacciones en el modelo), como ya fue previamente descrito (Henry et al., 2010). Esta tendencia sugiere que la dificultad en la

reconstrucción de un modelo metabólico se incrementa con redes de mayor tamaño. Con los resultados del análisis se estimó la frecuencia y distribución de las reacciones bloqueadas y los metabolitos *gap* en los 130 modelos que componen el conjunto de datos. En la [figura 6.1-b](#) se muestra, para cada reacción, el número total de modelos en que fue encontrada bloqueada respecto al número total de modelos en los que la reacción se encuentra presente. Los resultados indican que un total de 608 reacciones (~28%) no fueron detectadas como bloqueadas en ningún modelo (puntos sobre el eje x). Este subconjunto de reacciones incluyen actividades enzimáticas involucradas en rutas metabólicas centrales, así como rutas de biosíntesis de compuestos esenciales para la formación de biomasa, tales como aminoácidos y nucleótidos (para más detalles ver [tabla S.6.2AB](#)). Los resultados también muestran que un total de 1107 reacciones (~50%) están bloqueadas en algunos modelo y activas en otros. Esto indica que ciertas rutas fueron reconstruías de forma correcta en ciertos GEMM, mientras que en otros casos la reconstrucción contiene *gaps*, que bloquean ciertas reacciones (ver [tabla S.6.2](#) en el apéndice B).

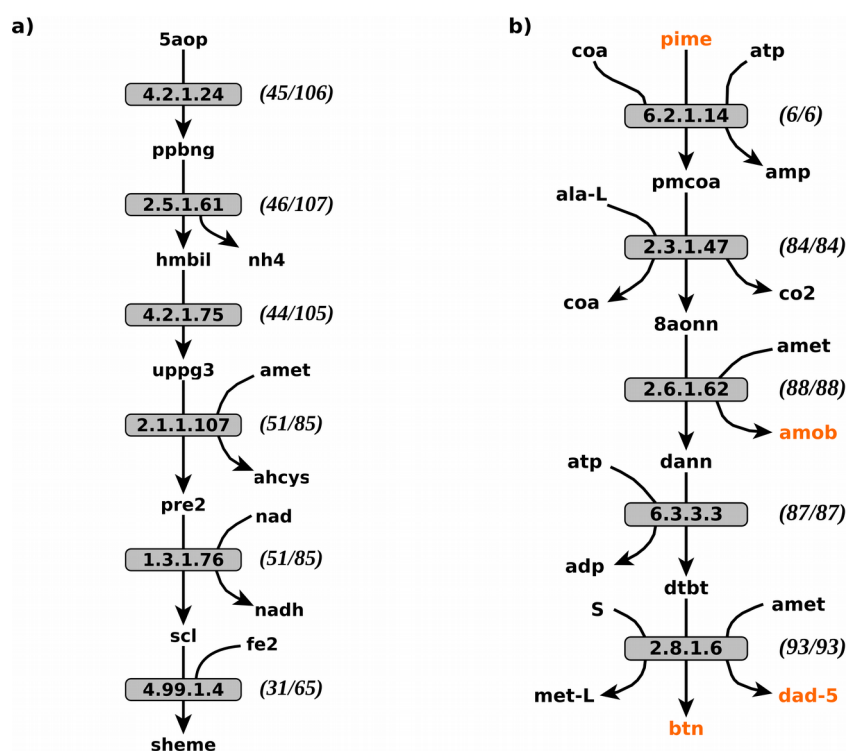


Figura 6.2. Ejemplos de dos rutas metabólicas cuya reconstrucción contiene inconsistencias. En a) ruta de biosíntesis del sirohemo, más común en procariotas. En b) ruta de biosíntesis de la biotina a partir del pimelato; en anaranjado se representan los metabolitos RNC. Los números entre paréntesis refieren al número total de GEMM en lo que una reacción fue encontrada como bloqueada/presente. Los nombres de los metabolitos se encuentran en la [tabla S.6.3C](#) del apéndice B.

Un ejemplo es el caso de la ruta de biosíntesis de sirohemo, un cofactor hallado en muchos procariotas y que participa de la asimilación de sulfitos y nitritos mediante su unión a ciertas reductadas³. Como se muestra en la [figura 6.2a](#), la ruta se encuentra conformada por tres actividades enzimáticas (EC 2.1.1.107, EC 1.3.1.76 y EC 4.99.1.4), y emplea el uroporfirinógeno III como precursor (Johansson and Hederstedt, 1999). La ruta se encuentra presente en al menos 65 GEMM en distintos grados de completitud (ver [figura 6.2a](#)): en 34 GEMMs (~40%) la ruta esta

3 Ver la discusión sobre la revisión de la ruta de síntesis del sirohemo en la [apartado 4.3.1](#).

completa y activa; en 28 modelos (~33%) la ruta está completa pero inactiva debido a que el producto final, *i.e.* el sirohemo, no se encuentra incluido en la ecuación de biomasa de dichos modelos; en 23 modelos (~27%) hay al menos una reacción no está presente, *i.e.* representa un *gap*, que causa el bloqueo de las reacciones restantes. Como muestran los resultados, la inferencia automática de la ruta de biosíntesis del sirohemo se encuentra lejos de ser trivial. Esto se deb, en entre otras cosas, a que el uroporfirinógeno III es precursor en la biosíntesis de al menos otros dos cofactores (adenosilcobalamina y el grupo hemo). Situaciones similares también fueron detectadas en rutas de síntesis de aminoácidos, *e.g.* la ruta de la histidina que se encontró reconstruida de forma completa e incompleta en 86 y 20 modelos, respectivamente.

Finalmente, un total de 480 reacciones (~22%) fueron detectadas como bloqueadas en cada uno de los modelos en que están presentes (ver diagonal principal en la [figura 6.1b](#)). Además, un total de 570 metabolitos fueron clasificados como *gaps* en todos los casos. Por ejemplo, la ruta de biosíntesis de la biotina a partir del pimelato incluye un total de cinco actividades enzimáticas ([figura 6.2b](#)). Notoriamente, ningún modelo incluye las reacciones de la ruta de síntesis del pimelato, la cual ha constituido un enigma hasta hace relativamente pocos años, en que se obtuvieron resultados experimentales que permitieron elucidar dos rutas alternativas de biosíntesis en *E. coli* y *B. subtilis* (Lin and Cronan, 2011). Como consecuencias un *gap* en el conocimiento de la bioquímica también puede ser una de las causas de las inconsistencias detectadas. La detección de reacciones siempre bloqueadas indica la existencia de inconsistencias recurrentes, lo cual sugiere que las mismas podrían ser intrínsecas al propio proceso de reconstrucción empleado para generar los 130 modelos. En ciertos casos, estas inconsistencias pueden ser consecuencia, por ejemplo, de errores sistemáticos en la anotación de los genomas que conducen, en última instancia, a la omisión y/o inclusión incorrecta de ciertas reacciones en los modelos. Por otra parte, la simple existencia de reacciones siempre bloqueadas podría indicar la presencia de *gaps* y/o inconsistencias en la propia base de datos metabólica de referencia empleada durante la reconstrucción de los modelos, como lo muestra el caso de la biotina y la síntesis del pimelato.

6.3.2 Construcción y curado del metamodelo

Con el objetivo de detectar y analizar inconsistencias globales, se construyó un metamodelo. La reconstrucción se realizó mediante la unión en una única red de los 130 GEMMs empleados en el apartado previo. El metamodelo resultante, referido como MM130.0 incluye un total de 2195 reacciones, 1811 metabolitos y 370 flujos de intercambio. Además, el metamodelo también incluye un total de 126 formulaciones de ecuación de biomasa diferentes⁴. Después de la construcción de MM130.0, se realizó un análisis de consistencia sobre dicho metamodelo. Los resultados indicaron que MM130.0 contiene un total de 417 reacciones bloqueadas y 520 metabolitos *gap*. El cálculo de los módulos inconexos dio como resultado que las reacciones bloqueadas y los metabolitos *gap* se distribuyen en un total de 229 UMs (ver en material suplementario la [figura S.6.1](#) y la [tabla S.6.3](#)).

4 Esta diferencia se debe a que los modelos de organismo filogenéticamente muy próximos comparten la misma definición de la biomasa, tal es el caso de dos grupos modelos (ver apartado 6.2.1).

Todas las reacciones previamente clasificadas como nunca bloqueadas (608), junto con las reacciones clasificadas como bloqueadas en algunos casos (1107), se encuentran todas activas en MM130.0⁵. Las 417 reacciones detectadas como bloqueadas en el metamodelo son un subconjunto de las 480 reacciones identificadas en el apartado previo como siempre bloqueadas. De esta misma forma, los 520 metabolitos clasificados como *gaps* en MM130.0, son un subconjunto de los 570 metabolitos clasificados como siempre *gap* en el análisis de consistencias de los 130 GEMMs. Estos resultados indican que 63 de las 480 reacciones halladas como siempre bloqueadas en los modelos (junto a 50 metabolitos clasificados como *gap*) se “reconectaron” como consecuencia de la construcción del metamodelo (ver [figura 6.3](#)). Para poder dilucidar las causas de la reconexión espontánea, las 63 reacciones y los 50 metabolitos se proyectaron en el correspondiente grafo bipartito. El grafo construido de esta forma contiene un total de 18 componentes conexos como se muestra en la [figura 6.3](#). El análisis visual de cada componente permitió identificar dos escenarios que explican la reconexión espontánea de reacciones y metabolitos inconsistentes. El primer escenario involucra el caso de rutas metabólicas incompletas en los GEMMs individuales, pero que al considerar la unión de los modelos en un metamodelo se completan ([figura 6.4a](#)). El segundo escenario corresponde a dos clases de artefactos: i) rutas lineales entrada-salida y ii) ciclos estequiométricos. Las rutas lineales entrada-salida, como el ejemplo de la [figura 6.4b](#), se corresponden a casos en los que un metabolito es importado y a continuación excretado, después de unos pocos pasos. Por otra parte, los ciclos corresponden a modos de flujo que violan el principio de conservación de energía y por lo tanto carecen de significado físico (ver [sección 3.3](#)). Un ejemplo se representa en la [figura 6.4c](#), que incluye dos formas alternativas de la ruta de biosíntesis de la vitamina B12 (adenosilcobalamina), que al ser consideradas en forma conjunta forman un ciclo estequiométrico, que viola el principio conservación de la energía⁶. Si bien las reacciones involucradas en las dos clases de artefactos descritos son consistentes desde el punto de vista del balance de masas, los modos de flujo que describen se corresponden a situaciones carentes de significado físico o fisiológico y por tanto representan inconsistencias que requieren la revisión correspondiente.

El tamaño y la distribución de los 229 UMs identificados en MM130.0 sugieren la presencia de tres grupos diferentes. Por un lado los tres UMs de mayor tamaño incluyen un total de 99 reacciones, las cuales se encuentran involucradas en la síntesis de lípidos de membrana y ácidos teicoicos. Tanto los lípidos de membrana como los ácidos teicoicos son polímeros cuya elongación se realiza a través un pequeño conjunto de reacciones iterativas o recursivas. En general, dado que en el análisis estequiométrico un polímero P de largo $P(n)$ y el mismo polímero con largo $P(n+1)$ se consideran dos especies químicas diferentes, esto mismo ocurre con las reacciones de elongación. Es decir, las reacciones de polimerización $P(n-1) \rightarrow P(n)$ y $P(n) \rightarrow P(n+1)$ son tratadas como reacciones diferentes, aunque la enzima que las catalice sea la misma. De manera que si en lugar de considerar las reacciones se consideran sus correspondientes actividades enzimáticas (*i.e.* considerando los números ECs asociados), el tamaño de los UMs se reduce mucho (ver [apartado 4.3.1](#)). Además, que debido a la gran diversidad que existe en la composición de las membranas y paredes celulares de los seres vivos, la reconstrucción de las rutas de síntesis de esta clase de componentes no es nada trivial. En el otro extremo se encuentran los UMs compuestos por una

5 Este resultado puede deducirse por las propiedades del FCA descritas en [la sección 3.7](#).

6 De acuerdo a la [definición 3.3](#) estas reacciones son *t-bloqueadas*.

única reacción. Un total de 179 UMs caen dentro de esta categoría y en muchos casos, la reacción que compone el UM se corresponde a un transportador cuyo sustrato no es metabolizado por ninguna de las reacciones presentes en el metamodelo. En medio de estos dos extremos se encuentra un tercer grupo de UMs, que son las que incluyen muchas rutas metabólicas relevantes desde un punto de vista bioquímico, así también como artefactos. Dada la diversidad presente en este grupo, la revisión y resolución de tales UM no presentan una solución trivial o que se pueda automatizar. Como consecuencia cada UM fue inspeccionada en forma visual y resueltas en forma manual, consultando las bases de datos metabólicas de referencia.

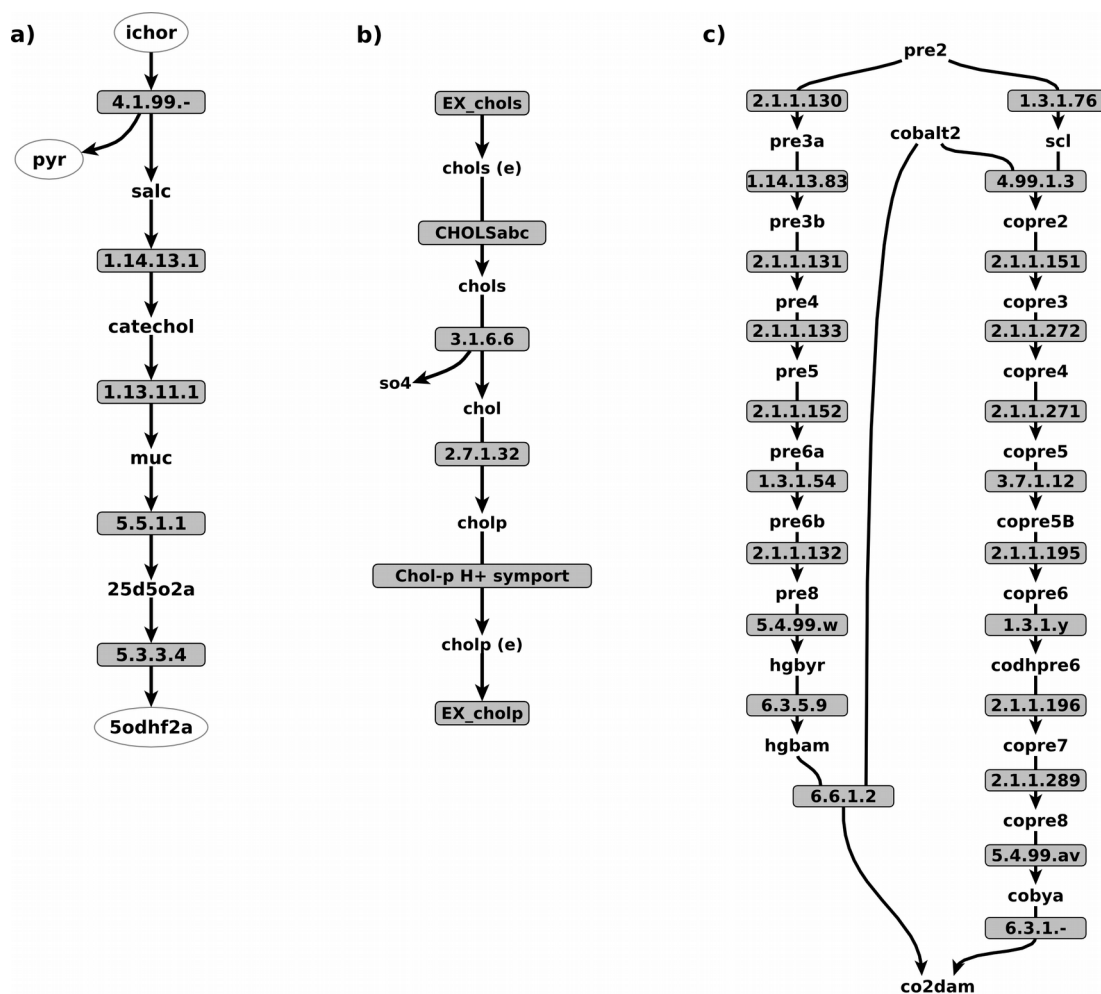


Figura 6.4. Tres ejemplos de rutas desbloqueadas como consecuencia de la construcción del metamodelo. En a) ruta de degradación aeróbica del catecol (los metabolitos encerrados en círculos se agregaron al diagrama para facilitar su interpretación). En b) importación del sulfato de colina, como ejemplo de ruta lineal entrada-salida. En c) dos variantes de la ruta de la biosíntesis de la vitamina B12 (incorporación tardía y temprana del cobalto a la izquierda y derecha, respectivamente) que forman un ciclo estequiometricamente cerrado. Los cofactores y metabolitos muy conexos fueron omitidos de las figuras. Los nombres de los metabolitos se pueden encontrar en la [tabla S.6.3C](#).

A modo de ejemplo, se describirá la curación del UM4 ([figura 6.2b](#)), que al igual que en el metamodelo, fue también detectada en la mayoría de los GEMMs, y que implica la transformación de pimelato en biotina. El UM contiene tres metabolitos *deadend* de tipo RNC: i) el S-adenosil-4-metilthio-2-oxobutanoato, subproducto de la adenosylmethionine-8-amino-7-oxononanoato

transaminasa EC 2.6.1.62⁷; ii) la 5'-deoxiadenosina, es un subproducto de la biotina sintasa EC 2.8.1.6⁸; y iii) la biotina, el producto final de la ruta. En el caso de los dos primeros *deadend*, dada la falta de información experimental sobre su destino metabólico (*e.g.* excreción, reciclaje, etc.) se añadieron dos flujos de intercambio que representan sumideros. Además, se añadieron las actividades necesarias para la síntesis del pimelato. Finalmente, la biotina, se debe incluir en la ecuación de biomasa, debido a que es un cofactor esencial para el crecimiento en diversas condiciones ambientales. De esta manera se representa la demanda del cofactor requerido por el organismo para poder crecer.

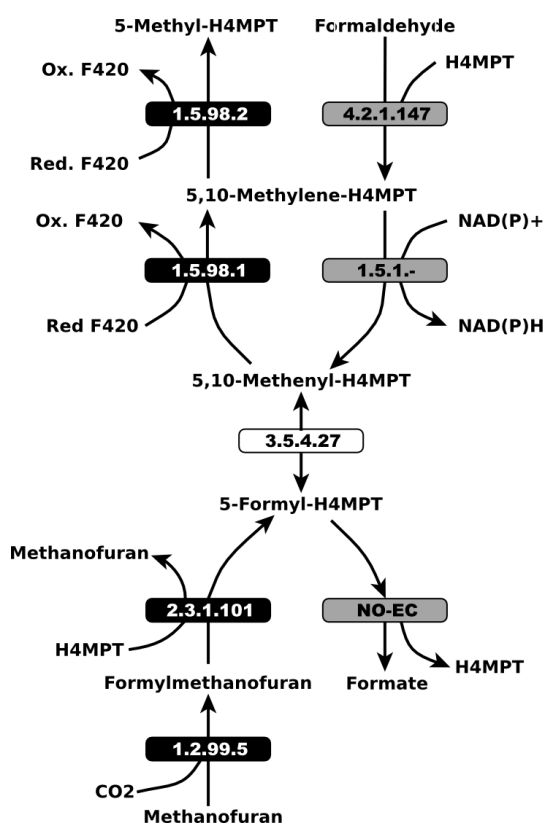


Figura 6.5. Relación entre las rutas metabólicas de la metanogénesis a partir del CO₂ y la denominada oxidación del formaldehído V (H4MPT). El color negro indica las reacciones que fueron eliminadas debido a que son exclusivas de la ruta metanogénica. El color blanco indica actividades involucradas en la metanogénesis y en la ruta de la oxidación del formaldehído. Las reacciones coloreadas con gris participan de la ruta de la oxidación del formaldehído y fueron agregadas de forma manual al metamodelo. Las abreviaciones de los metabolitos se pueden encontrar en la tabla S.6.3C.

Otro ejemplo ilustrativo es el caso del UM7, compuesto por 5 actividades enzimáticas propias de la metanogénesis, un proceso metabólico que es realizado exclusivamente por organismos miembros del dominio *Archaea* (ver figura 6.5). Las reacciones fueron únicamente encontradas en los GEMMs de tres proteobacterias: *Burkholderia xenovorans* LB400; *Methylococcus capsulatus* str. Bath; and *Methylobacillus flagellatus* KT. En base a lo antes expuesto, es razonable suponer que estas reacciones fueron añadidas a los modelos mencionados, como consecuencia de errores en la anotación funcional de los genomas correspondientes. Sin embargo, un análisis más detallado

7 Este compuesto fue previamente descrito como un *gap* universal (Mackie et al., 2013).

8 La 5'-deoxiadenosina ha sido descrito como un *deadend* en el modelo de *E. coli* iJO1366 Orth et al., 2011

sobre la biología de las tres bacterias mostró que estas codifican alguna variante de la ruta metilotrónica (Chistoserdova et al., 2009), la cual es, en cierto sentido, análoga a la metanogénesis, ruta con la cual comparte diversas enzimas (Ferry, 2010). Esto podría ser la razón de que actividades propias de la metanogénesis hayan sido anotadas en genomas bacterianos. Es decir, que un conjunto de actividades enzimáticas que cumplen papeles diferentes en distintos organismos puede conducir a la inferencia incorrecta de una ruta. Con el objetivo de refinar el metamodelo, las reacciones con números EC 1.2.99.5, 1.5.98.1, 1.5.98.2 y 2.3.1.101 que pertenecen exclusivamente a la ruta metanogénica, fueron eliminadas. Además, se añadieron las reacciones con ECs 4.2.1.147, 1.5.1.- y la formiltransferasa/hidrolasa (formación de formato a partir del 5-formil-H4MPT sin EC asignado), que junto a la actividad EC 3.5.4.27 (presente en el metamodelo) completan la denominada ruta de la oxidación del formaldehído V (H4MPT), descrita en la base de datos metabólica MetaCyc (Caspi et al., 2014). Estos cambios dieron como resultado la definición de una ruta en el metamodelo que se ajusta en mayor medida al fenotipo metilotrónico de ciertas bacterias.

Tabla 6.2. Resultados de la construcción y revisión manual de metamodelo. Las dos primeras columnas describen la versión inicial del metamodelo creada mediante la unión de 130 GEMM de bacterias. La tercera y cuarta columnas indican el número de elementos eliminados/agregados, respectivamente, como consecuencia de la curación manual. La última columna describe la versión final del metamodelo.

Componente	MM130.0 (Inicial)	Eliminados	Agregados	MM130.1 (Revisado)
Metabolitos	1811	222	77	1666
† Reacciones	1767	106	230	1891
Transportadores	428	55	23	396
Intercambios	328	52	19	295
* Sumideros	--	--	3	3

† Las ecuaciones de biomasa no están consideradas.

* Un sumidero es un tipo especial de flujo de intercambio introducido para permitir que aquellos metabolitos cuyo destino bioquímico es desconocido, puedan ser eliminados del sistema evitando así la formación de un “deadend”.

Durante el proceso de revisión y refinado se añadieron al metamodelo un total de 230 reacciones, 23 transportadores y 77 metabolitos. A su vez se introdujeron 19 flujos de intercambio y 3 sumideros. Las reacciones añadidas fueron extraídas de la base de datos SEED (Overbeek et al., 2014), excepto en 22 casos donde las reacciones que fueron añadidas de forma manual, y que fueron encontradas en MetaCyc (Caspi et al., 2014), en KEGG (Kanehisa and Goto, 2000) o en BiGG (Schellenberger et al., 2010). Además, se 29 metabolitos, no incluidos en ninguna de las 127 ecuaciones biomasa originales, fueron clasificados como componentes de la biomasa del algún organismo. Dichos conjunto de componentes incluye: 26 compuestos que corresponden a diversas clases de componentes de membrana y 3 cofactores entre los que se encuentra la biotina⁹. Estas modificaciones permitieron desbloquear un total de 259 reacciones. Por otra parte, se eliminaron del metamodelo un total de 161 reacciones y 222 metabolitos que no pudieron ser resueltos. Estas reacciones y metabolitos inconsistentes, denominados meta-bloqueadas y meta-gaps,

⁹ Como se verá más adelante, esta información puede ser empleada durante el proceso de resolución de inconsistencias de un GEMM individual.

respectivamente, fueron eliminadas del metamodelo por uno de los siguientes motivos: i) o bien no tienen lugar en bacterias (*e.g.* el caso de las reacciones metanogénicas); o ii) en la bibliografía consultada no se encontró información que permitiera establecer la forma en que las reacciones se encuentran conectadas al resto del metabolismo. Como resultado del proceso de revisado, se obtuvo un metamodelo consistente denominado MM130.1. Los cambios introducidos al metamodelo se resumen en la [tabla 6.2](#), y la [figura S.6.1](#) representa las UMs del metamodelo resueltas (o eliminadas). Todos los detalles correspondientes al proceso de revisión y resolución de inconsistencias en el metamodelo se encuentran en las [tablas S.6.3A-E](#).

6.3.3 Análisis de acoplamiento de flujos del metamodelo

Este apartado se centra en la aplicación del análisis de acoplamiento de flujos (FCA, ver [apartado A.2.3](#) del apéndice A). El objetivo del análisis es detectar relaciones funcionales entre las actividades enzimáticas a escala de un metamodelo. Como fue discutido en la sección 3.7, esta información puede ser de utilidad para asistir el proceso de reconstrucción y refinados de GEMMs. Además, esta clase de estudio puede ayudar a entender la organización del metabolismo a una mayor escala que cuando se analizan organismos individuales. El FCA fue realizado sobre el metamodelo revisado MM130.1 mediante empleando el algoritmo FFCA (David et al., 2011), como se describe en el [apéndice C](#). Los resultados del análisis muestra que de las 2287 reacciones (sin incluir flujos de intercambio), un total de 1405 (~61%) forman parte de al menos una relación de acoplamiento. En total se identificaron 8184 pares de acoplamientos que solo representan un 0.5% de todos los posibles pares de acoplamientos. En efecto, un importante número de reacciones (~39%) no presentan ningún tipo de acoplamiento. Esto se debe a que al considerar un metamodelo se incrementa la redundancia de la red lo que disminuye el acoplamiento entre las reacciones.

En muchos casos las relaciones de acoplamiento entre pares de reacciones son direcciones y no recíprocas ($i \rightarrow j$). Esto implica que la reacción i requiere la presencia de la reacción j para poder operar en estado estacionario, pero que j no requiere de la presencia i . También existen muchos casos donde un subconjunto de dos o más reacciones se implican mutuamente ($i \leftrightarrow j$) dando lugar a un conjunto de acoplamiento completo (FCS, full coupling set), o subconjuntos enzimáticos (Pfeiffer et al., 1999). Dado que los flujos de las reacciones dentro un FCS son linealmente dependientes, en el estado estacionario cada FCS representa una unidad funcional indivisible, y por tanto cada FCS puede ser tratado como una meta-reacción, ya que desde el punto de vista estequiométrico conforma una clase de equivalencia (ver [apartado A.2.3](#)). En el caso de MM130.1 las relaciones de acoplamiento recíproco definen un total de 364 FCS (ver [tabla 6.3](#)). Cada FCS está formado por un cierto número de reacciones involucradas, que en total incluyen 1071 reacciones. Dado que desde el punto de vista estequiométrico y funcional, cada FCS se comporta como si fueran una única reacción, las 1071 reacciones involucradas en algún FCS se corresponden a 364 meta-reacciones. Como consecuencia, el número de flujos “independientes” dentro del metamodelo es de 1580 en lugar de 2287. Este resultado indica que, pese al gran número aparente de reacciones incluidas en el metamodelo, las restricciones impuestas por la estequiometría de la red dan lugar a la presencia de múltiples FCS, que reducen en gran medida los grados de libertad del sistema.

Tabla 6.3. Descripción de los FCS presentes en MMI30.1. En la primera y segunda columnas se indica el número de reacciones que conforman el FCS y cuantos FCS de dicho tamaño hay, respectivamente. En la última columna de listan algunos subsistemas o rutas metabólicas asociadas a los FCS. Los números entre paréntesis indican el total de FCS de un tamaño dado, asociados al subsistema correspondiente.

#Reacciones	#FCS	Subsistema o ruta metabólica
34	1	Biosíntesis de lipopolisacáridos (1)
16	3	Biosíntesis de ácidos grasos (3)
12	1	Beta oxidación de ácidos grasos (1)
11	3	Biosíntesis: Vitamina B12 (1), Biotina (1) y Peptidoglicano (1)
10	2	Biosíntesis: Histidina (1) y Cobalamina (1)
9	2	Degradación de naftaleno (1) Biosíntesis de Vitamina B12 (1)
8	1	Biosíntesis de ácidos grasos (1)
7	3	Biosíntesis: Riboflavina (1) y UQ (1) Degradación aminoácidos aromáticos (1)
6	6	Biosíntesis: clorofila (1), cofactores (2), ácidos grasos (1), otros (2)
5	6	Rutas de degradación diversas (5) Biosíntesis de leucina (1)
4	37	Biosíntesis: aminoácidos (2), envoltura celular (10), cofactores (6) otros (19)
3	65	Biosíntesis: cofactores (9), aminoácidos (5), envoltura celular (11) otros (40)
2	234	Biosíntesis (41) Lípidos (59) Transporte (25) Otros (109)

La presencia de acoplamientos surge como consecuencias de las restricciones que la conservación de la masa impone sobre el universo bioquímico. Además, debido a las propiedades matemáticas de los acoplamientos se espera que las relaciones detectadas en el metamodelo, se encuentren presentes en GEMM de organismos particulares. Como consecuencia, la existencia de meta-acoplamientos podría tener consecuencias tanto en la organización y regulación, como en el origen y evolución del metabolismo. Concretamente, aquellas actividades enzimáticas que se implican mutuamente (*i.e.* las enzimas presentes en un FCS) seguramente tengan una regulación coordinada (Montagud et al., 2011; Notebaart et al., 2008), así como patrones evolutivos similares en términos de perfiles filogenéticos (Cohen et al., 2012; Pellegrini et al., 1999), y de la co-evolución de su sitios activos (Juan et al., 2008; Pál et al., 2005). Sin embargo, hay que tener en cuenta que la relación entre genes, proteínas y reacciones puede no ser de uno a uno. De manera que la correspondencia entre un FCS y un módulo definido en términos de proteínas no es trivial. Aunque, como se verá más adelante, existen casos donde hay una clara correspondencias entre FCSs y módulos a nivel de las proteínas.

La distribución de tamaños de los FCS (en número de reacciones) varía entre 2 y 34, con una media de ~3 reacciones. Si bien la desviación típica es de aproximadamente de 3 reacciones, la distribución es bastante dispersa. En efecto, la mayoría de los FCSs (~64%) están formados por dos reacciones, mientras que solo un ~8% incluyen 5 o más reacciones (ver [tabla 6.3](#)). El FCS de mayor tamaño incluye un total de 34 reacciones que se encuentran involucradas en la síntesis de lipopolisacáridos. Muchos de los FCS que incluyen más de 4 o 5 reacciones se encuentran también asociados a procesos iterativos de biosíntesis de polímeros (*e.g.* lípidos de membrana, componentes

de la pared celular, etc). En tales casos el número de actividades enzimáticas acopladas es considerablemente menor al número de reacciones que incluye el FCS¹⁰. De esta manera, si se proyectan las reacciones sobre los números EC que representan las actividades enzimáticas, este número se reduce en forma notable respecto al total de reacciones. Por ejemplo el FCS de mayor tamaño, que está formado por 34 reacciones asociadas a la síntesis de lipopolisacáridos, se reduce a 13 actividades cuando las reacciones son proyectadas sobre sus números ECs (contando cada EC una única vez), y lo mismo ocurre con los FCS de 16 reacciones relacionados con la síntesis de lípidos, que al proyectarlos en ECs se reducen a 3. Sin embargo, hay un número considerable de FCS donde sí hay correspondencia entre el número de reacciones y de actividades enzimáticas diferentes, y tales casos se corresponden en su mayoría con rutas de biosíntesis, mayoritariamente de cofactores como la riboflavina, el piridoxal o las vitaminas B7 y B12, pero también de algunos aminoácidos y rutas de degradación de diversos compuestos (ver [tabla 6.3](#)). Un ejemplo muy ilustrativo de relación entre FCS y ruta metabólica lo constituye el caso de la síntesis de histidina. En dicho caso, los 10 genes que codifican las 9 actividades enzimáticas que definen la ruta completa conforman un único FCS (ver [tabla 6.4](#)). Esto implica que desde el punto de vista estequiométrico esos 10 genes constituyen una unidad con una función muy concreta y exclusiva que es la síntesis de histidina. Además, al analizar en los modelos la posibilidad de que los genes codifiquen actividades involucradas en otras rutas se encontró que en la mayoría de los casos, los genes son específicos de la ruta de la histidina, a excepción de la actividad transaminasa (EC 2.6.1.9) codificada por *hisC*, que en algunos organismos también codifica la actividad EC 2.6.1.57 involucrada en la síntesis de fenilalanina.

Tabla 6.4. Genes y actividades enzimáticas de la ruta de síntesis de la histidina. La tabla incluye los 10 genes y las 9 actividades enzimáticas que estos codifican y que definen la ruta completa de síntesis de histidina. A su vez la tercer columna indica el total de modelos en que dicha actividad fue encontrada bloqueada / activa.

Genes	Número EC	Nº de GEMM con la actividad (bloqueada / activa)
<i>hisG</i>	2.4.2.17	15 / 86
<i>hisI</i>	3.6.1.31	13 / 86
<i>hisI</i>	3.5.4.19	13 / 86
<i>hisA</i>	5.3.1.16	15 / 86
<i>hisF</i> & <i>hisH</i>	2.4.2.-	16 / 86
<i>hisB</i>	4.2.1.19	15 / 86
<i>hisC</i>	2.6.1.9	16 / 86
<i>hisK</i>	3.1.3.15	6 / 86
<i>hisD</i>	1.1.1.23	15 / 86

El hecho de que las actividades de la ruta de la histidina conformen un FCS en el metamodelo sugiere un acoplamiento funcional a diversos niveles. Por ejemplo, el acoplamiento estequiométrico podría tener implicaciones sobre la evolución de las secuencias de las enzimas de

¹⁰ Esto se debe a que tales rutas están compuestas por un conjunto de actividades que operan en forma iterativa, e.g. la elongación de un lípido implica únicamente cuatro actividades enzimáticas (ver discusión de la UM4 en la [apartado 4.3.1](#)).

dicha ruta, ya que, la optimización del flujo de producción de histidina requeriría el cambio coordinado de las constantes cinéticas de las distintas enzimas. Esto implicaría a su vez, un proceso coordinado en la evolución de las enzimas involucradas. En efectos, estudios previos realizados sobre los genes de dicha ruta indican que las secuencias exhiben en cierta medida señal de co-evolución (Juan et al., 2008). Por otra parte, si una enzima participa en distintos procesos metabólicos, es posible la optimización del flujo en una de las rutas afecte de forma negativa el desempeño de la otro proceso. Como consecuencia, se espera que las enzimas con esta característica no presenten acoplamiento estequiométrico con otras actividades. De acuerdo con lo anterior, sería esperable que aquellas enzimas que se encuentran acopladas, puedan evolucionar de forma coordinada sin afectar al resto del sistema. Mientras que las enzimas que participan en muchas rutas metabólicas se encuentra sujetas a restricciones más estrictas respecto a su evolución, ya que una modificación que podría optimizar el flujo a través de cierta ruta podría ser en detrimento de otro proceso.

Desde el punto de vista de la reconstrucción metabólica, el hecho de que en un metamodelo todas las actividades de una ruta formen parte de un FCS sugiere que la reconstrucción de dicha ruta en un organismo particular debería ser, en principio, trivial. Esto se debe a que con encontrar algunos de los genes que codifiquen actividades de la ruta sería suficiente para inferir la presencia necesaria de los demás actividades. Además de los FCS, las relaciones de acoplamiento direccional detectadas en un metamodelo también pueden emplearse para inferir reacciones que son requeridas para que ciertas actividades puedan operar. De esta manera, el conocimiento de las relaciones de acoplamiento de un metamodelo facilitaría en gran medida la reconstrucción de ciertas rutas metabólicas (para un análisis más detallado ver la discusión de los meta-acoplamientos en la [sección 3.7](#)). Esta información podría ser empleada para dirigir la búsqueda de los genes, no detectados en el proceso de anotación y que generan los *gaps* en la estructura de la ruta. Sin embargo, de la misma manera en que algunos genes son suficientes para inferir la presencia de las demás actividades, si dichos genes se hubiesen anotado de forma incorrecta, esto conduciría a completar una ruta no presente en el organismo, lo que representaría un artefacto. Al analizar la presencia y consistencia de la ruta de la histidina en los 130 modelos se encontró que de los 106 modelos que contiene al menos una de las actividades de la ruta; en 21 de estos, dichas reacciones se encuentra bloqueadas debido a la presencia de *gaps*. Este resultado muestra que incluso en casos donde la reconstrucción de una ruta es en principio trivial, los algoritmos de *gap-filling* automáticos fallaron al completar la ruta, incluso en casos donde el modelo incluye 9 de las 10 reacciones implicadas en la vía, Esto indica que los algoritmos de *gap-filling* aún pueden ser mejorados. Finalmente, hay que considerar que los organismos con metabolismos reducidos y que exhiben complementación metabólica pueden presentar dificultades.

Para facilitar el análisis de los acoplamientos, la estructura de las relaciones de acoplamiento puede ser representada mediante un grafo dirigido, denominado el grafo de acoplamiento (Burgard et al., 2004). En dicha red, cada nodo representa una reacción y cada arista dirigida o arco con origen en un nodo i y destino en un nodo j representa un un acoplamiento direccional entre ambas reacciones, es decir $i \rightarrow j$. A diferencia de la representación del metabolismo a través del grafo de compuestos, en el cual los nodos son metabolitos y estos se conectan unos con otros siempre que participen de una misma reacción (Barabási and Oltvai, 2004; Wagner and Fell, 2001), en el grafo

de acoplamientos los nodos corresponden a actividades enzimáticas y los arcos indican dependencias funcionales entre las mismas, que son consecuencia de las restricciones estequiométricas impuestas por el balance de masa. Como consecuencia, esta representación del metabolismo, si bien parcial, esta exenta de los artefactos que se producen en el grafo de compuestos por la presencia de metabolitos altamente conectados como el ATP, H₂O, CO₂, NAD⁺, etc., que crean relaciones espúreas. El grafo de acoplamientos es dirigido, de forma que cada nodo *i* tendrá asociado un grado de salida k_{out} , que indica el número de reacciones implicadas, y un grado de entrada k_{in} que indica el número de reacciones que lo implican, es decir, aquellas reacciones que requieren de la presencia de *i* para poder funcionar. Cada reacción *i* podrá ser caracterizada por su grado de salida k_{out} , así como por un grado de entrada k_{in} . A su vez, como ya se ha mencionado previamente, siempre que existan un arco (*i*, *j*) y un arco (*j*, *i*), ambas reacciones formaran parte de un mismo FCS.

La figura 6.6a representa el grafo de acoplamientos del metamodelo MM130.1, que por simplicidad no incluye las 882 reacciones que no presentan acoplamientos con otras reacciones, *i.e.* que tienen un grado $k_{in} = k_{out} = 0$. Si se ordenan los nodos de acuerdo su grado de entrada k_{in} se observa que la reacción con mayor número de implicaciones corresponde a la actividad EC 2.7.7.39 que se encuentra implicada por otras 69 reacciones. Esta actividad cataliza la formación de CDP-glicerol, compuesto involucrado la síntesis peptidoglicano que, por tratarse de un polímero, implica una síntesis iterativa y no representa un número real de actividades enzimáticas, como se ha discutido más arriba. Por otra parte, al analizar el grado de salida k_{out} se encuentra que las reacciones con el mayor grado de salida ($k_{out} = 35$) corresponden a las actividades que conforman el FCS de mayor tamaño (ver tabla 6.3) y que participan de la ruta de biosíntesis de lipopolisacáridos. Como se puede deducir de este resultado, las relaciones de acoplamiento entre los miembros de un FCS se reflejan en el grado de los nodos ya que todas las reacciones de un FCS se conectan unas con otras. Dado que las relaciones de acoplamiento entre las reacciones son transitivas, si una reacción *i* perteneciente a un FCS está direccionalmente acoplada con una reacción *j* que no forma parte del FCS, todas las demás reacciones del FCS se encontrarán acoplados a *j*. Como consecuencia las relaciones de acoplamiento entre los miembros de un FCS implican información redundante. Una aproximación para eliminar dicha redundancia y simplificar la representación, consiste en condensar todos los miembros de un FCS en un único nodo. En la figura 6.6b se representa el grafo de acoplamientos una vez que se han condensando las reacciones de cada FCS en un único nodo que, en cierto sentido, representan de forma parcial (y en algunos casos completa) módulos funcionales. Si bien en esta representación se pierde la información de cuantas reacciones conforman cada FCS, la ventaja es que el diagrama se simplifica al reducir la información redundante, y es posible identificar en forma visual patrones de conectividad entre los módulos que representan los FCS. Al analizar la estructura del grafo de acoplamientos se ve que, incluso si se excluyen las reacciones con grado cero, la estructura es no conexas, y contiene diversos componentes de distintos tamaños (figuras 6.6a-b). El grafo de acoplamiento sin condensar contiene un total de 396 componentes (sin contar las reacciones aisladas), muchas de las cuales (334) corresponden a FCSs, que incluyen, por ejemplo, reacciones de rutas de degradación o síntesis de distintas clases de metabolitos (*e.g.* amino ácidos) así como también rutas de asimilación de compuestos como azúcares (*e.g.* sorbosa, galactosa, etc).

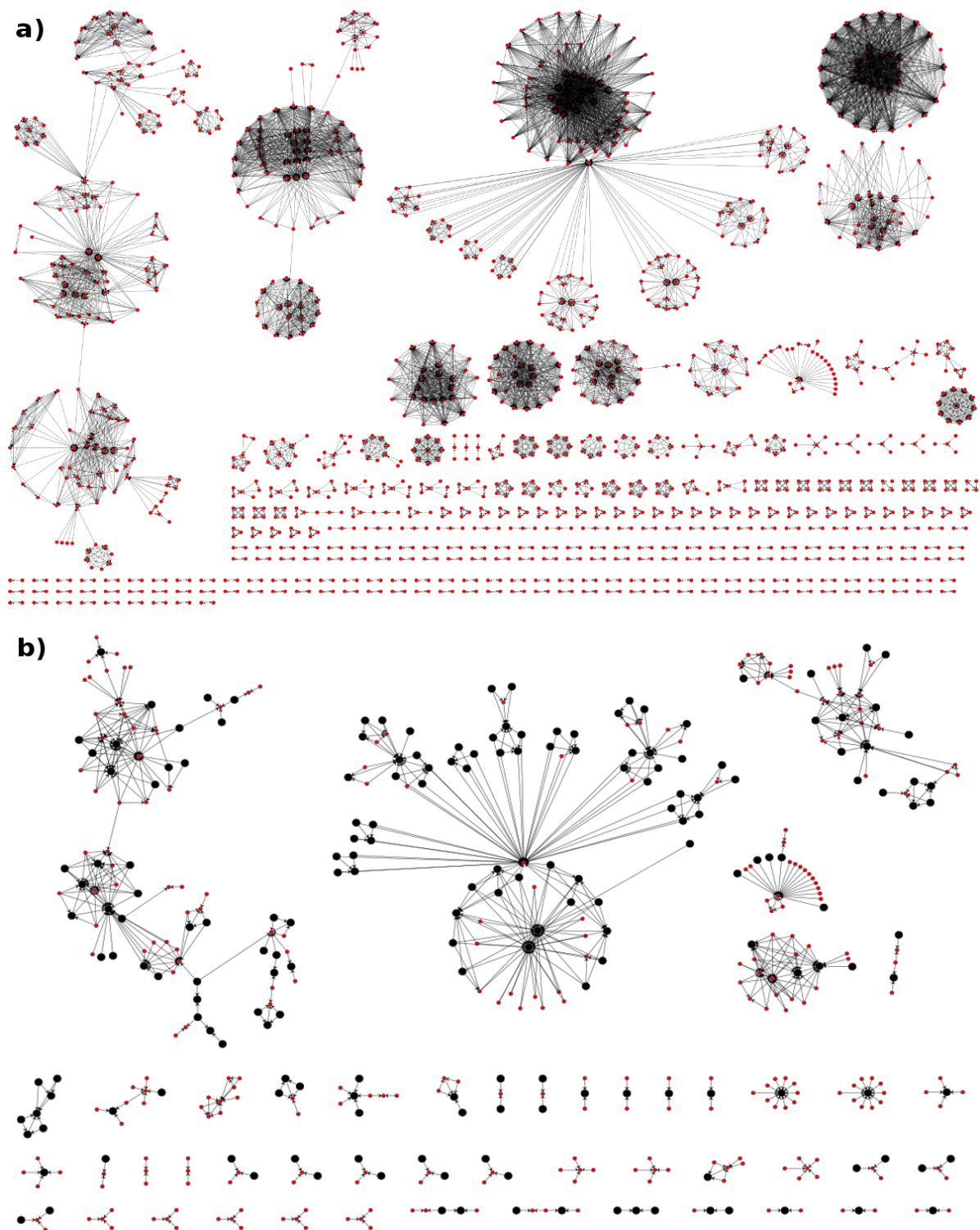


Figura 6.6. Grafo de acoplamiento del metamodelo MM130.1. En a) se representan todas las relaciones de acoplamiento entre los flujos de MM130.1. Cada nodo representa una reacción y las flechas indican que una reacción requiere a otra. En b) el grafo de acoplamientos pero las reacciones de un mismo FCS se representan con un único nodo de mayor tamaño y color negro. Los nodos rojos son reacciones individuales. En b) los FCS sin conexiones (334) no se han representado en la figura.

El grafo condensado por su parte incluye un total de 62 componentes (excluyendo las reacciones y FCS aislado). Las componentes de mayor tamaño incluyen reacciones de distintas rutas metabólicas, que en algún punto se encuentran ligadas estequiométricamente debido, por lo general, al hecho de compartir algún precursor. Por ejemplo, el componente de mayor tamaño incluye reacciones presentes en las rutas de síntesis de mevalonato, síntesis *de novo* de purinas, y el FCS que corresponde a la ruta de síntesis de la histidina (ver [tabla 6.4](#)) entre otras. En este caso, existe un acoplamiento direccional entre el FCS que representa la ruta de síntesis de histidina, y la ruta síntesis *de novo* de purinas. Tal acoplamiento se da a través del imidazol-4-carboxamida (AICAR). Dicho metabolito es un sub-producto de la biosíntesis de histidina y a la vez es un intermediario de la ruta síntesis *de novo* de purinas. Como el AICAR solo es consumido en la síntesis de purinas, el proceso de síntesis de histidina queda direccionalmente acoplada a la síntesis *de novo* de purinas, *i.e.* que siempre que se produzca histidina se produce AICAR y este sólo puede ser reutilizado en la síntesis de purinas. Como consecuencia, la producción de histidina implicará o la producción de purinas, o la acumulación de AICAR en la célula, o la eliminación del AICAR por excreción. Esta situación podría representar un caso de punto débil en el metabolismo (Meléndez-Hevia et al., 2009; Melendez-Hevia and Paz-Lugo, 2008), ya que si en cierto contexto la demanda de purinas es menor a la demanda de histidina la concentración de AICAR comenzará a aumentar y esto reducirá el flujo de producción de histidina. Por otra parte, la eliminación de AICAR representa un desperdicio importante para la célula y esto debería tener consecuencias selectivas negativas. Sin embargo, se han visto que ciertas bacterias endosimbiontes, con genomas muy reducidos, que han perdido la ruta de síntesis *de novo* de purinas pero que conservan la ruta de síntesis de histidina, *e.g.* *B. aphidicola* BCc, *P. aleyrodidarum*. Dado que en casos así, el metabolismo de hospedador y huésped se complementan, se ha propuesto que el AICAR podría ser excretado por el endosimbionte al citosol de las células del hospedador, en la cuales el compuesto sería reciclado produciendo purinas que a su vez son asimiladas en parte por los endosimbiontes (Calle-Espinosa et al., 2016; Lamelas et al., 2011a). Este ejemplo ilustra cómo las relaciones de acoplamiento permiten detectar conexiones no triviales entre procesos metabólicos, *a priori*, independientes como son la síntesis de aminoácidos y nucleótidos.

Por otra parte, la distribución de conectividad¹¹ del grafo de acoplamientos brinda información acerca del número total de reacciones $N(k_{out})$ que implican o requieren k_{out} reacción (grado de salida) así como el número de reacciones $N(k_{in})$ que son requeridas por k_{in} reacciones. La [figura 6.7](#) muestra, en escala logarítmica, la distribución de grado de entrada y salida, *i.e.* la frecuencia de nodos para distintos valores de k , en el grafo de acoplamientos sin condensar y con los FCS condensados. En los cuatro casos, las distribuciones fueron ajustadas a una distribución de ley de potencia. Como se muestra en la [figura 6.7](#), si bien el ajuste de la regresión es más claro para el caso del grafo condensado (mayor r^2), dado que el número de nodos es menor también lo es la significancia estadística (menor p -valor). Sin embargo, en los cuatro casos el p -valor calculado indica que los ajustes son significativos. Resultados similares fueron descritos en estudios previos donde se vio que para los casos de las redes metabólicas de *H. pylori*, *E. coli* y *S. cerevisiae*, la distribución de conectividad del grafo de acoplamientos sigue una distribución potencia con

11 $N(k)$ es la distribución de grado que indica la probabilidad o frecuencia de nodos con k conexiones. En el caso de los grafos dirigidos se pueden definir dos distribuciones de grado, la distribución del el grado de salida $N(k_{out})$ y la distribución del grado de entrada $N(k_{in})$.

exponentes similares al encontrado para el caso del metamodelo (Burgard et al., 2004). Es remarcable el hecho de encontrar en el metamodelo una relación de escala en grado de los acoplamientos entre las reacciones, similares a las previamente encontradas en organismo individuales. Como discuten los autores, esto podría sugerir cierto principio general en la organización del metabolismo.

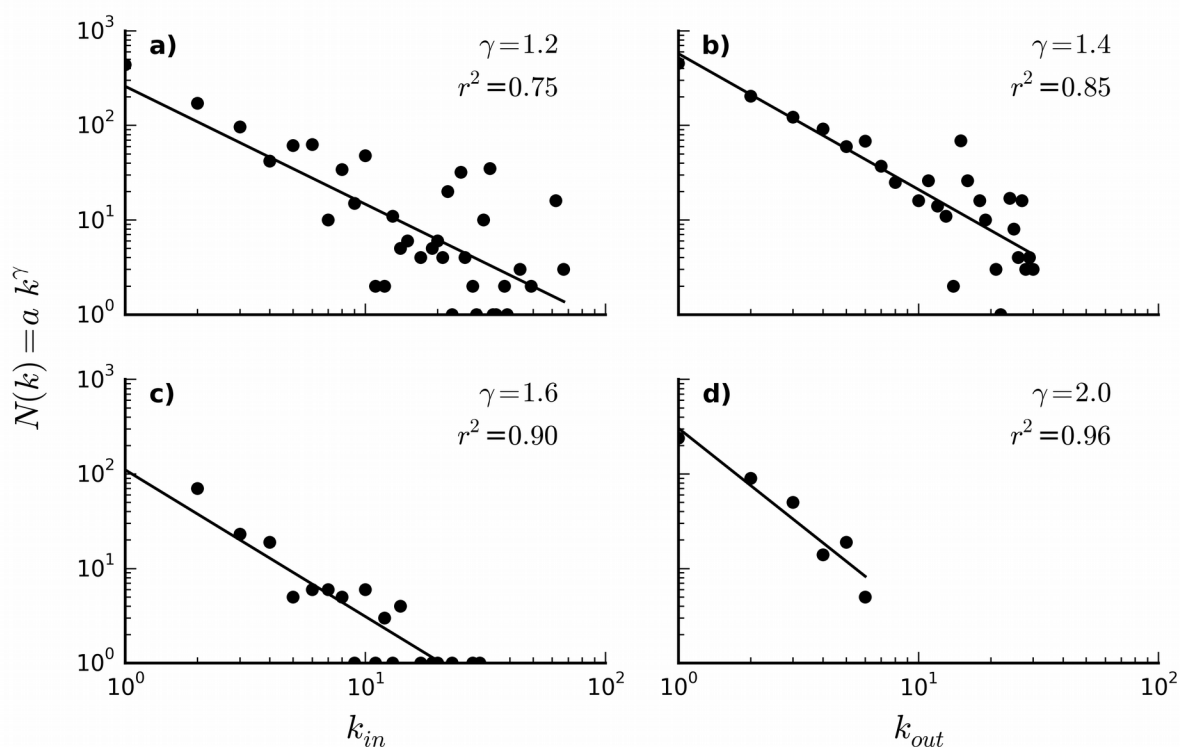


Figura 6.7. Distribuciones de grado del grafo de acoplamiento de MM130.1. Para el grafo de acoplamiento sin condensar (ver figura 6.6a), se representa en a) el número de reacciones $N(k)$ que son implicadas por otras k reacciones y en b) el número $N(k)$ de reacciones que implican a otras k reacciones. En c) y d) se muestran las mismas representaciones que en a) y b), respectivamente, pero los cálculos se realizan sobre el grafo de acoplamiento en el cual los FCS han sido condensados en un único nodo (ver figura 6.6b). Los p -valores correspondientes para las regresiones representadas en a), b), c) y d) son 2×10^{-7} , 3×10^{-8} , 2×10^{-7} , 2×10^{-3} , respectivamente.

Considerando que los exponentes reportados por los autores aumentan conforme los modelos son más grandes, es esperable que el exponente obtenido con el metamodelo sea mayor. En efecto, el exponente estimado para la distribución del grado de salida en el grafo condensado, es $\gamma=2.0$, valor levemente mayor que valores previamente obtenidos para organismo singulares: $\gamma=1.2, 1.3, 1.4$, para los casos de *H. pylori*, *E. coli* y *S. cerevisiae*, respectivamente. A diferencia del estudio realizado por Burgard et al. donde únicamente se analiza la distribución del grado de salida (Burgard et al., 2004), *i.e.* el número reacciones requeridas por una cierta reacción, en el presente estudio también se ha realizado el análisis de la distribución del grado de entrada de las reacciones, es decir el número de reacciones que implican a un reacción determinada. En este sentido se ha de destacar que no se encontró correlación alguna entre el grado de entrada y salida de los nodos en ninguna de las dos representaciones del grafo de acoplamientos (datos no mostrados). Además, es notorio el hecho de que en el grafo de acoplamientos sin condensar los exponentes de las distribuciones de los grados de entrada y salida son similares, lo que indica que

dichas propiedades tienen un factor de escala muy similar. Por otra parte, es notable la diferencia que hay en el exponente de la distribución del grado de entrada (o salida) estimado sobre el grafo sin condensar y condensado. El exponente estimado es mayor en el grafo condensado, tanto para el grado de salida como el de entrada. Esta relación de escala encontrada en el grafo de acoplamientos del metamodelo (también descrita en organismos individuales), podría sugerir cierta universalidad en la estructura de los acoplamientos en las redes metabólicas.

Por otra parte, los estudios previos realizados empleando la representación del metabolismo mediante el grafos de compuestos, muestran que los compuestos más centrales del metabolismo (*e.g.* ATP/ADP, NAD⁺/NADH, piruvato, etc.) se corresponden con los nodos más conexos del grafo de compuestos (Barabási and Oltvai, 2004; Jeong et al., 2000; Wagner and Fell, 2001). De acuerdo con este resultado, podría esperarse que las reacciones que se encuentran muy implicadas por otras reacciones en el grafo de acoplamientos se correspondan a funciones centrales del metabolismo, y posiblemente ancestrales. Sin embargo, el análisis del grafo de acoplamientos indica lo opuesto, es decir que las actividades enzimáticas involucradas en el metabolismo central (*e.g.* glicólisis, ciclo TCA, o ruta de las pentosas fosfato) no presentan casi acoplamientos. En concreto, todas las actividades presentes en la glicólisis, la ruta de las de pentosas fosfato o el ciclo TCA, tienen un grado tanto de salida como de entrada igual a cero. Este resultado sugiere la gran versatilidad que existe en el metabolismo central, como se ha comentado brevemente al analizar el efecto de *knockout* dobles en el metabolismo de *Blattabacterium* Bge (ver apartado 4.3.5). Por otra parte, las actividades involucradas, tanto en rutas de biosíntesis, como en rutas de degradación, tienden a presentar cierto grado de acoplamiento, llegando a casos extremos como la síntesis de la histidina donde toda la ruta conforma un único FCS.

Finalmente, desde el punto de vista de la reconstrucción metabólica, las relaciones de acoplamiento pueden ser empleadas para predecir reacciones que deben estar presentes en el modelo, dada la presencia de otras reacciones. Esto es posible ya que de acuerdo con las propiedades matemáticas de los acoplamientos, estas relaciones entre las reacciones de una red siguen siendo válidas en todas aquellas subredes consideradas como subconjuntos de reacciones (para más detalles sobre las propiedades matemáticas de las relaciones de acoplamiento ver la sección 3.7). Entonces, al reconstruir el metabolismo de un organismo se puede combinar el listado de actividades enzimáticas que se extrae de la anotación del genoma, con la información de los acoplamientos a nivel del metamodelo. Esto puede ayudar a elucidar o inferir otras reacciones necesarias en el modelo, no detectadas durante la anotación. Por ejemplo, el caso de la ruta de síntesis de la histidina donde todas las reacciones conforman un FCS, la detección de genes de la ruta sería suficiente para inferir los genes que faltan para completar la misma. No obstante, esta aproximación se vería limitada para el caso de la reconstrucción del metabolismo central debido a su bajo grado de acoplamiento, pero dado que suele ser la parte del metabolismo más conservada, su reconstrucción no debería presentar mayores dificultades. En esta misma dirección, así como las relaciones de acoplamiento de un metamodelo se pueden emplear para asistir o facilitar el proceso de reconstrucción metabólica, estas también pueden ser empleadas durante el proceso de revisión y refinado de un modelo, como se verá en el próximo apartado.

6.3.4 Aplicación del *gap-filling* al conjunto de 130 GEMMs

La versión inicial del metamodelo (MM130.0) y su versión revisada (MM130.1) fueron empleados en forma alternativa, como redes de referencia para realizar *gap-filling* sobre cada uno de los 130 GEMMs. En este sentido, la comparación entre modelos completados con los diferentes metamodelos permite evaluar el impacto que la consistencia y completitud de un metamodelo tienen al realizar *gap-filling* sobre un GEMM. Además, los resultados del análisis de consistencia del metamodelo permiten identificar en los modelos individuales reacciones meta-bloqueadas y que por tanto carecen de sentido en el modelo. En la [figura 6.8](#) se representa en forma esquemática el proceso de detección, clasificación y curado de las reacciones bloqueadas de un modelo. El proceso de refinado se dividió en cuatro pasos o etapas que incluyen el pre-procesamiento de los modelos, el *gap-filling* o completado de los *gaps* y post-procesamiento. En la etapa de pre-procesamiento se identifican y eliminan las reacciones meta-bloqueadas y se calculan los pesos que se emplearán para penalizar la adición de reacciones externas al modelo por parte de *fastcore*¹². Los pesos se ajustan teniendo en cuenta si el incorporar una reacción al modelo implicaría asumir la presencia de uno o más genes codificantes para la actividad que no identificadas en la anotación del genoma.

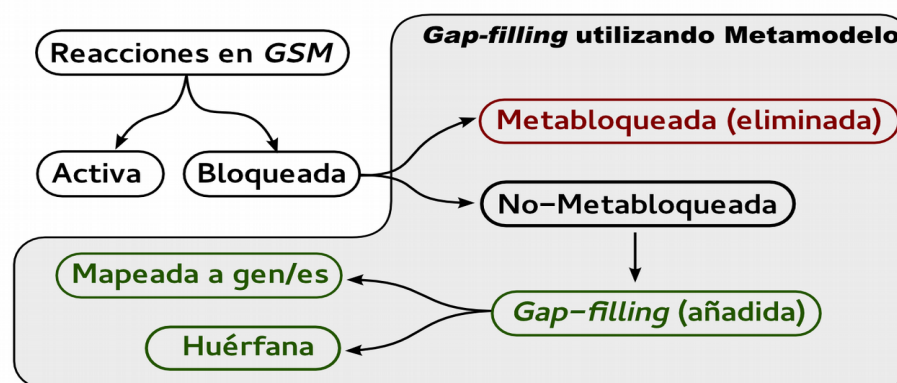


Figura 6.8 Diagrama de flujo del proceso de *gap-filling*. Cada una de las reacciones bloqueadas es clasificada como no-meta-bloqueada o meta-bloqueada. Las meta-bloqueadas (rojo) son aquellas reacciones que no pudieron ser resueltas en el metamodelo y por lo tanto serán eliminadas de los modelos individuales. Las reacciones no-meta-bloqueadas, son aquellas que se encuentran bloqueadas en algunos modelos y activas en otros (inclusive el metamodelo) y por lo tanto pueden ser resueltas. Las reacciones *gap-filling* (verde) son el conjunto de reacciones que se necesita añadir a un modelo para desbloquear todas las reacciones bloqueadas no-meta-bloqueadas. Las reacciones *gap-filling* agregadas un GEMM pueden ser posteriormente asociadas a uno o más genes codificantes (mapeadas) mientras que otras permanecerán huérfanas (sin gen asociado).

La etapa de auto-completado o *gap-filling* se divide en dos pasos. En el primero se emplean las relaciones de acoplamiento del metamodelo para inferir reacciones no presentes en el modelo pero que son requeridas por una o más reacciones bloqueadas del modelo. Por ejemplo, en el apartado previo se discute el hecho de que las reacciones involucradas en la síntesis conforman un FCS y en consecuencia en un modelo alguna de las reacciones se encuentran ausente, toda la ruta estará bloqueada. Entonces, para cada reacción bloqueada i presente en un modelo, se identifican todos los meta-acoplamientos, *i.e.* todas las reacciones cuya ausencia producen el bloqueo de i , en el metamodelo. Los meta-acoplamientos que no se encuentran presentes en el modelo son añadidos.

¹² El algoritmo fue re-implementado en lenguaje Python a partir del código en Matlab distribuido con la publicación original (Vlassis et al., 2014). Para detalles de la implementación ver [apéndice C](#).

Dado que en muchos casos la adición de los meta-acoplamientos, si bien necesaria, no es suficiente para resolver las inconsistencias, el segundo paso del auto-completado implica la aplicación de un algoritmo de *gap-filling* basado en optimización. El algoritmo de *gap-filling* empleado es una versión modificada de *fastcore*, en la que es posible incluir un vector de pesos lineales que permite penalizar en forma diferenciada las reacciones candidatas a añadirse al modelo para resolver *gaps*. En concreto, los pesos se ajustan de la siguiente forma: *i*) reacciones espontáneas, sumideros y flujos de intercambio, tienen un peso de 0 ya que tanto una reacción espontánea como un flujo de intercambio no implica suponer la presencia de un gen codificante; *ii*) las reacciones no presentes en el modelo que tienen número EC igual a alguno de los EC de las reacciones presentes en el modelo también se les asigna un peso de 0, debido que el modelo ya incluye un gen que potencialmente podría catalizar dicha reacción; *iii*) a las demás reacciones no presentes en el modelo se les asigna un peso genérico de 10, valor que mejora el desempeño de *fastcore*, como se ha descrito previamente (Thiele et al., 2014). En la etapa de post-procesamiento para cada reacción añadida en el auto-completado, su número EC se compara con el de las reacciones originales del modelo y si hay coincidencias con alguna reacción, entonces la regla de asociación gen-proteína-reacción de dicha reacción es transferida a la nueva reacción. El proceso completo de refinado se resume en los siguientes tres pasos:

1. **Pre-procesamiento:** Eliminar todas las reacciones meta-bloqueadas y los metabolitos meta-*gap* del modelo y pre-computar los pesos para *fastcore*
2. **Auto-completado de *gaps* (*gap-filling*):**
 - a) Completado mediante meta-acoplamientos:
 - b) *gap-filling* basado en optimización utilizando *fastcore*, empleando el metamodelo como la red global de referencia y utilizando los pesos pre-computados en 1.
3. **Post-procesamiento:** para cada nueva reacción añadida se busca un gen codificante de una enzima presente en el modelo, que sea capaz de catalizar la reacción.

El procedimiento descrito fue aplicado a cada uno de los 130 GEMMs presentes en el conjunto de datos empleando en forma alternativa MM130.0 y MM130.1 como red de referencia. Esto resultó en dos versiones consistentes de cada modelo. Si bien, el procedimiento descrito ha permitido realizar el curado los 130 modelos individuales, en este apartado sólo se presentan y discuten los resultados globales y cualitativamente más significativos con respecto a la consistencia del metamodelo empleado como red de referencia. El porcentaje promedio de reacciones auto-completadas empleando meta-acoplamientos y *fastcore*, es en el promedio de los 130 modelos del ~50%. Es decir, que aproximadamente la mitad de las reacciones necesarias (aunque no suficientes) para resolver las inconsistencias en un modelo se pueden inferir a partir de las relaciones de acoplamiento del metamodelo. Esto indica que gran parte de las reacciones ausentes que generan los *gaps* en los modelos pueden ser inferidas de las relaciones de acoplamiento de un metamodelo. Además, siempre que no se modifique el metamodelo, los meta-acoplamientos solo se deberán computar una única vez. Como consecuencia, el auto-completado de un modelo basado en meta-acoplamiento no requiere tiempo de cómputo y entonces el tiempo global de cálculo para curar un modelo se reduce, en relación al tiempo requerido al auto-completar el modelo empleando únicamente *fastcore*. Si bien la diferencia en los tiempos de cómputo no es muy notoria, esta será

más apreciable si en lugar de emplear *fastcore*, se emplean métodos de optimización basados en programación entera mixta, como los previamente propuestos (Kumar et al., 2007; Latendresse, 2014; Reed et al., 2006b).

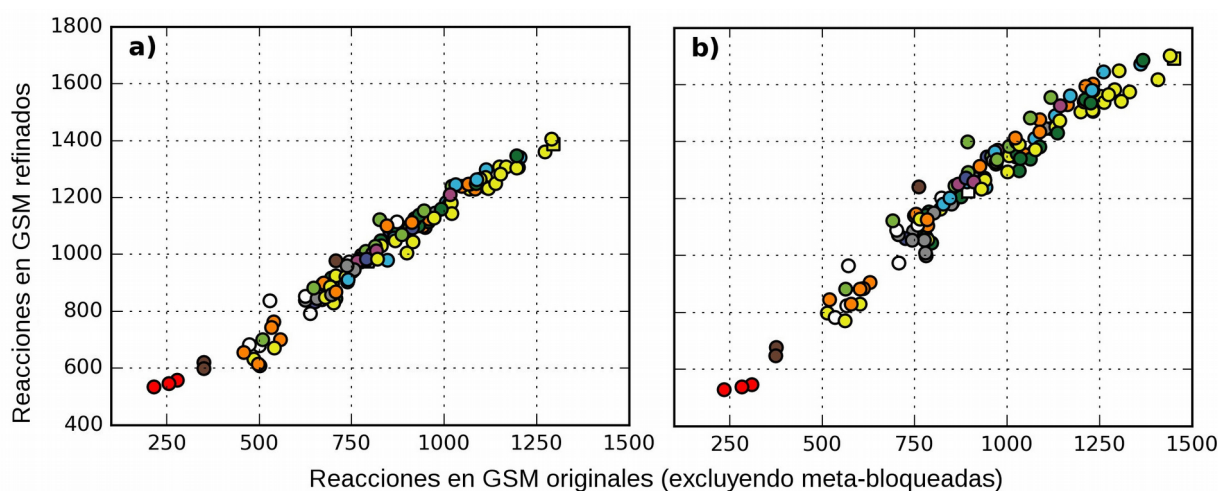


Figura 6.9 Relación entre los modelos antes y después del gap-filling. En a) y b) se presentan los resultados del gap-filling empleando MM130.0 y MM130.1, respectivamente. Ambas gráficas representan la relación entre el número total de reacciones en los modelos refinados, con respecto al número de reacciones (excluyendo las reacciones meta-bloqueadas) presentes en los modelos originales. Los cuadrados blancos y amarillos corresponden a *E. coli* y *T. thermophilus*, respectivamente.

Por otra parte, al comparar los modelos curados respecto a su versión original, los resultados muestran que el tamaño de los modelos completados tiende a incrementarse notoriamente. Este resultado es independiente de la versión de metamodelo empleado como referencia. Como se muestra en las figuras 6.9a-b, la relación de tamaños entre los modelos, antes y después de resolver las inconsistencias, sigue una relación lineal. Además, se encuentra una tendencia similar si al considerar el tamaño de los modelos originales se consideran también las reacciones meta-bloqueadas (las cuales no se incluyen en las figuras 6.9a-b). Como era de esperar, el tamaño de los modelos obtenidos luego de aplicar el procedimiento de *gap-filling* es mayor para la versión que se obtiene al emplear MM130.1. Sin embargo, para ciertos modelos no parece haber diferencias en el resultado obtenido al emplear una u otra versión del metamodelo. Tal es el caso de cinco organismos, por un lado *M. genitalium* G-37, *M. pulmonis* y *Phytoplasma* OY (puntos rojos), y por otro *T. pallidum* y *B. burgdorferi* B31 (puntos marrones). Estos cinco organismos comparten la característica de ser patógenos y de poseer los genomas más pequeños de los organismos del conjunto de datos. Como consecuencia, la gran mayoría de las rutas metabólicas que se encontraban bloqueadas en MM130.0, resueltas en MM130.1, no están codificadas en los genomas de los cinco patógenos y por lo tanto, en dichos casos, el emplear un metamodelo u otro, no afecta el resultado del *gap-filling*.

Al comparar en cada modelo el número total de reacciones añadidas que no tienen genes asociados (huérfanas) respecto el número total de reacciones bloqueadas que son curadas, se ve, que para la mayoría de los modelos, esta relación es menor que uno (ver figuras 6.10a-b). En tales casos la resolución de una reacción boqueada implica, en promedio, la adición de una reacción huérfana al modelo. Las figuras 6.10a-b también permiten ver que al emplear un metamodelo

consistente un mayor número de modelos queda por debajo de la recta $x=y$, lo que indica que requieren la adición de menos reacciones huérfanas en relación a las reacciones que son resueltas. Sorprendentemente, en el caso de *T. thermophilus*, el modelo con el mayor porcentaje de reacciones bloqueadas, la resolución de todas sus inconsistencias requiere la adición de un número relativamente pequeño de reacciones huérfanas. Sin embargo, en muchos otros casos el número total de reacciones huérfanas añadidas excede de forma significativa el total reacciones bloqueadas resueltas. Esto sugiere que en dichos casos muchas de las reacciones bloqueadas podrían representar artefactos y que por lo tanto debería ser “podadas” en lugar de desbloqueadas mediante la adición de reacciones huérfanas. Esta tendencia es muy clara para el caso de los cinco organismos patógenos con genomas reducidos mencionados previamente.

Para evaluar el efecto que tiene el emplear distintas versiones de un metamodelo en los resultados del algoritmo de *gap-filling*, se comparó el contenido de reacciones de cada par de modelos refinados con MM130.0 y MM130.1. En un total de 121 modelos, la versión que fue completada empleando MM130.1 contiene al menos el 90% de las reacciones de la versión completada empleando MM130.0. Esto sugiere que, en la mayoría de los casos el modelo consistente que se obtuvo al emplear MM130.0 como referencia se encuentra casi completamente incluido en la versión obtenida al emplear MM130.1. Sin embargo, en el caso de cinco modelos el porcentaje de inclusión fue menor al 80%. Estos cinco modelos, que se corresponden con los patógenos antes mencionados, tienen aproximadamente el mismo número total de reacciones huérfanas añadidas, y sin embargo difieren en forma significativa, al menos respecto a los demás modelos, cuando se comparan los respectivos conjuntos de reacciones.

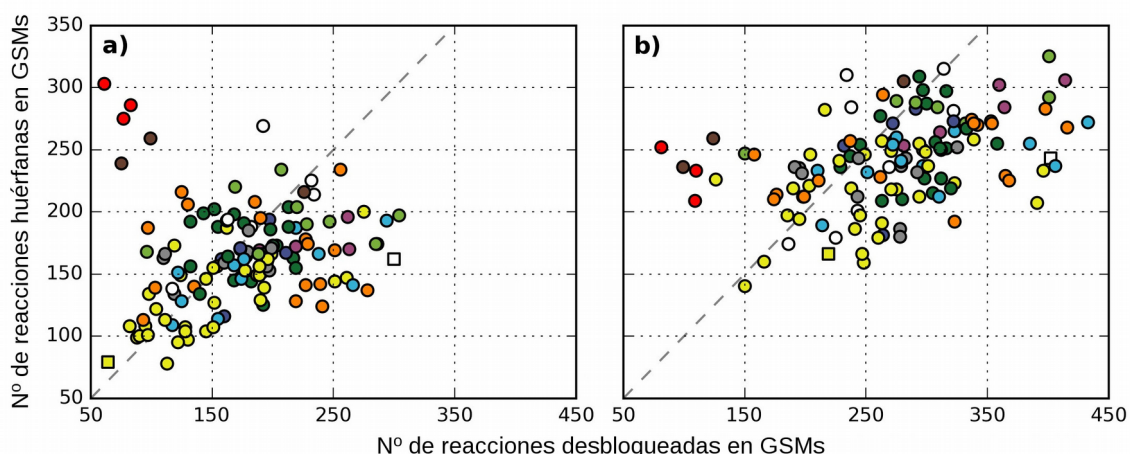


Figura 6.10 Relación entre las reacciones desbloqueadas y las reacciones añadidas. En a) y b) se presentan los resultados del *gap-filling* empleando MM130.0 y MM130.1, respectivamente. Ambos gráficos representan la relación entre el número de reacciones huérfanas que se añaden a los modelos, respecto al número de reacciones inconsistentes que son resueltas como consecuencia de añadir las reacciones huérfanas. La línea punteada representa la función $y=x$. El cuadrado blanco y el amarillo corresponden a *E. coli* y *T. thermophilus*, respectivamente.

Las rutas que fueron resueltas a mano en MM130.1 se completaron correctamente en aquellos modelos que incluían alguna de dichas rutas reconstruidas de forma inconsistente. Por ejemplo, la ruta de síntesis de la biotina fue completada en 97 modelos. Resultados similares fueron obtenidos para el caso del sirohemo y de la vitamina B12. Sin embargo, como es de esperar, ninguna de estas rutas puede ser completada en los modelos al realizar el *gap-filling* cuando se emplea la versión

inconsistente del metamodelo (MM130.0). Además, se encontró que al emplear MM130.0 como red de referencia, muchos de los artefactos descritos (ver apartado 6.3.2) se propagan a los modelos al aplicar el algoritmo automático de *gap-filling*. Por ejemplo, para todos aquellos modelos que incluyen al menos una reacción de la ruta de síntesis de la vitamina B12, el correspondiente modelo refinado incluye las dos variantes de dicha ruta (ver figura 6.4c), que dan lugar a un artefacto ya que es la única forma en que dichas reacciones pueden operar en MM130.0.

6.3.5 Extendiendo un metamodelo con nueva información bioquímica

Las bases de datos metabólicas como el SEED o el BioCyc, al igual que las bases de datos de secuencias, son actualizadas periódicamente. Las actualizaciones incluyen la incorporación de nuevas reconstrucciones metabólicas, así como nueva información bioquímica, como por ejemplo una nueva actividad enzimática, o la resolución de una ruta metabólica no descrita previamente. En el presente estudio, al realizar la construcción del metamodelo MM130 no se incluyeron ciertos modelos presentes en el SEED. Tales modelos pueden contener información no presente en MM130.1 como compuestos, reacciones, o incluso rutas metabólicas completas. Como consecuencia, un metamodelo debería ser actualizado con cierta frecuencia. Por lo tanto, para evaluar el impacto de extender un metamodelo con nueva información bioquímica en dos sentidos. Por un lado se estudiará su impacto en las propiedades estructurales del metamodelo, *i.e.* como se reorganizan los acoplamientos. Por otra parte, se analizará el efecto en la resolución de los modelos individuales. Para extender el metamodelo se emplearon los 36 GEMMs de bacterias no considerados en la construcción de MM130 (verla descripción del conjunto de datos en el apartado 6.2.1). El primer paso para realizar la extensión del metamodelo fue identificar todas las reacciones y metabolitos presentes en los 36 GEMM, que no estuviesen incluidas en MM130.1. En total se encontraron 331 reacciones y 243 metabolitos y estos fueron añadieron al metamodelo. La extensión dio lugar a una nueva versión del metamodelo denominada MM166.0. Esta primera versión de MM166.0 incluye un total de 1909 metabolitos y 3047 reacciones. Al analizar la consistencia del metamodelo extendido se detectó que 143 reacciones añadidas se encontraban bloqueadas. De este conjunto, 119 están incluidas en la lista de reacciones clasificadas como meta-bloqueadas y por lo tanto fueron descartadas (ver apartado 6.3.2). Las restantes 24 reacciones bloqueadas fueron resueltas de forma manual, a través de la resolución de los módulos inconexos correspondientes. Durante la resolución de dichas reacciones, se añadieron de forma manual 12 nuevas reacciones y dos flujos de intercambio.

Tabla 6.5. Resultados de la extensión del metamodelo con 36 nuevos GEMMs. La tabla resume el número de nuevos metabolitos y reacciones añadidos a la versión refinada de MM166.1

Componente	MM130.1	Añadidos	MM166.1 (consistente)
Metabolitos	1666	113	1779
† Reacciones	1891	141	2032
Transportadores	396	55	451
Intercambios	295	30	325

† Sin considerar ecuaciones de biomasa.

La extensión de MM130.1 con 36 nuevos modelos y la subsecuente revisión y refinado, dieron como resultado una versión consistente del metamodelo extendido, denominada MM166.1, que se encuentra descrita en la [tabla 6.5](#) (el [archivo S.6.1](#) incluye el metamodelo MM166.1 en formato SBML). Esta nueva versión incluye un total de 113 nuevos metabolitos, así como 196 nuevas reacciones, de las cuales 55 son transportadores. La mayor parte de las nuevas reacciones intracelulares incluidas, participan en rutas metabólicas de degradación de compuestos, como por ejemplo la putresina o hidrocarburos como el octano, así como rutas de degradación de distintos azúcares (*e.g.* ruta alternativa de degradación de la lactosa). Además, como se indica en la [tabla 6.5](#), MM166.1 incluye 30 nuevos flujos de intercambio, lo que se traduce en 30 nuevos compuestos que pueden ser importados o excretados por determinados modelos.

6.3.6 Análisis del impacto de la extensión de un metamodelo

El impacto de extender un metamodelo fue analizado en dos sentidos. En primer lugar se estudiaron los cambios en las propiedades estructurales de la red que surgen al añadir nuevas reacciones. Concretamente se estudió la reorganización de los subconjuntos de acoplamiento completo, es decir, como se reducen o fraccionan los FCS presentes en MM130 al añadir nuevas reacciones y cómo se organizan las nuevas reacciones en FCS. En segundo lugar se analizó el impacto que la extensión de un metamodelo puede tener al realizar *gap-filling*. Este segundo análisis implicó realizar un *gap-filling* como el descrito en el [apartado 6.3.4](#), pero empleando MM166.1, para luego comparar las nuevas versiones refinadas con las obtenidas previamente empleando MM130.1 como modelo de referencia. El análisis de los cambios en la organización del FCS se realizó de la misma forma que la comparación de los modelos *Blattabacterium* Bge iCG238 y iMP240, es decir, comparando los FCS mediante el índice de Jaccard (ver [apartado 4.3.3](#)).

La [tabla 6.6](#) presenta los resultados de la comparación en la organización de los FCS entre las dos versiones del metamodelo, *i.e.* MM130.1 y MM166.1. En primer lugar se destaca que las 196 nuevas reacciones, 105 se encuentran agrupadas en 32 FCS, donde el de mayor tamaño incluye 14 reacciones involucradas en la síntesis de ácidos grasos. Otros FCSs se corresponden con las ya mencionadas rutas de degradación de la putresina y la lactosa. Si bien, de acuerdo a las propiedades de los acoplamientos, al expandir un modelo mediante la incorporación de nuevas reacciones, las relaciones de acoplamiento entre las relaciones iniciales pueden relajarse, es notable el hecho de que la mayoría de los FCS presentes en MM130.1 no se ven afectados como consecuencia de la extensión. Esto indica que la mayoría de las reacciones añadidas pertenecen a subsistemas metabólicos independientes a los presentes en MM130.1 y como consecuencia no modifican en gran medida las relaciones de acoplamiento entre las reacciones originales. No obstante, algunos FCSs sí se ven afectados por la expansión, indicando que algunas reacciones añadidas aportan modos alternativos de funcionamiento que relajan la estructura de los FCSs previamente identificados.

Tabla 6.6. Reorganización de los FCS al extender el metamodelo. La primera columna indica el número de reacciones que conforman el FCS. La segunda y tercera columnas indican el número de FCS de cierto tamaño presentes en MM130.1 y MM166.1, respectivamente. La cuarta indica cuántos pares de FCS de un tamaño dado son iguales entre ambos modelos. En la quinta y sexta columnas, el número por fuera de los paréntesis indica cuantos FCS con un índice de Jaccard mayor a 0.75 y 0.5 han sido encontrados, respectivamente, tomando como referencia de la comparación del modelo MM166. El número entre paréntesis indica la diferencia en reacciones para un par de FCS.

# Reacciones	MM130	MM166	$J_{1.0}^*$	$J_{0.75}$	$J_{0.50}$
2	234	250	213	--	12 (-1), 4 (-2)
3	65	67	52	9 (-1)	1 (-2)
4	37	30	25	1 (-1)	1 (-2)
5	6	9	4	1 (-1)	--
6	6	5	4	1 (-1)	--
7	3	3	2	--	--
8	1	3	1	--	1 (-4)
9	2	2	2	--	--
10	2	2	1	1 (-1)	--
11	3	2	2	--	--
12	1	0	--	--	--
14	0	1	--	--	--
16	3	3	3	--	--
33	0	1	--	1 (-1)	--
34	1	0	--	--	--

Como se menciona al inicio del apartado, para evaluar el impacto de la extensión de un metamodelo en el resultado del *gap-filling*, se aplicó el procedimiento descrito en la [apartado 6.3.4](#) a los 130 GEMM. El procedimiento de *gap-filling* fue también aplicado a los 36 nuevos modelos empleados para la extensión del metamodelo (empleando MM130.1 y MM166.1). De manera que para cada uno de los 166 modelos metabólicos se obtuvieron dos versiones refinadas: una empleando MM130.1 y la otra empleando MM166.1. De forma similar al análisis realizado en [apartado 6.3.4](#), la comparación entre cada versión de modelo refinado se basó en dos factores: 1) la eficacia del *gap-filling*, es decir el número de reacciones desbloqueadas respecto al número de reacciones añadidas; y 2) las diferencias entre los conjuntos de reacciones añadidas a un modelo al emplear cada versión del metamodelo. La eficacia promedio para el caso de los modelos refinados con MM130.1 y MM166.1 fue de ~ 0.87 y ~ 0.93 , respectivamente para los 130 del conjunto de datos inicial. Mientras que para los 36 nuevos modelos la eficacia fue de ~ 0.9 con MM130.1 y ~ 0.99 al emplear MM166.1. El incremento en la eficacia al emplear el metamodelo extendido se puede explicar en base al hecho que el algoritmo *fastcore* identifica un modelo consistente mínimo partiendo del conjunto de reacciones incluidas en el modelo con *gaps*. La forma de identificar dicho mínimo es minimizando el número total de reacciones que se deben añadir al modelo para que todas las reacciones sean consistentes. Por lo tanto, si cuando se aplica *fastcore* (o un procedimiento alternativo de *gap-filling* basado en optimización) se emplea como red de referencia

un metamodelo más grande, que incluya nuevas rutas y reacciones, es esperable que surjan formas de resolver las inconsistencias que impliquen un menor número de reacciones. Como consecuencia, al emplear metamodelos más completos se espera un incremento en la eficacia del completado. Sin embargo, a la hora de evaluar las diferencias entre las soluciones obtenidas con los dos metamodelos, la métrica de eficacia definida no brinda ninguna información acerca de la similitud entre los conjuntos de reacciones obtenidos.

Para poder comparar las soluciones obtenidas para cada modelo, en término de las reacciones que son añadidas, se empleó el índice de Jaccard. El índice fue calculado para cada par de modelos comparando los conjuntos de reacciones añadidas en cada caso. El valor medio de los índices de Jaccard medidos sobre los 130 modelos fue de ~ 0.71 , y el mismo valor se obtuvo para los 36 modelos adicionales. Esto indica que en promedio cada par de versiones curadas comparte aproximadamente el 85% de las nuevas reacciones añadidas (tabla S.6.4F). Sorprendentemente, no se aprecian mayores diferencias entre los valores del índice de Jaccard calculado para los conjuntos de 130 y 36 modelos metabólicos, lo que sugiere que la mayor parte de la información bioquímica necesaria para curar los modelos se encuentra presente en MM130.1, y que las 196 reacciones adicionales añadidas en el proceso de extensión pertenecen a unos pocos modelos. Además, dado que la mayoría de las nuevas reacciones añadidas durante la extensión del metamodelo, eran en su mayoría consistentes en los modelos individuales, es de esperar que su presencia no afecte los resultados del *gap-filling*.

6.4 Discusión

6.4.1 Detección de inconsistencias estructurales en modelos metabólicos

El análisis de consistencia realizado sobre el conjunto de 130 modelos metabólicos de distintas bacterias ha revelado que en la mayoría de los modelos, al menos el 25% de las reacciones se encuentran bloqueadas. Los resultados también indican que la reconstrucción automática de un organismo no-modelo es más sencilla cuando este se encuentra filogenéticamente próximo a organismos más caracterizados y para los que se dispone de una reconstrucción manual. Esto se ve reflejado en el hecho de que los modelos con menos de un 25% de sus reacciones bloqueadas pertenecen a las clase de las γ -proteobacterias, que incluyen a *E. coli*. En efecto, el modelo con el menor porcentaje de reacciones bloqueadas ($\sim 19\%$) corresponde a *E. coli* K12 (Opt83333.1). Como punto de referencia, el porcentaje de reacciones bloqueadas en la última versión de la reconstrucción manual del GEMM de *E. coli* K12 iJO1366 es de $\sim 10\%$.

Si se analiza el otro extremo, *i.e.* el modelo con mayor porcentaje de reacciones bloqueadas, nos encontramos con que el GEMM de *T. thermophilus* (Seed300852.3) tiene un total de $\sim 47\%$ de sus reacciones bloqueadas. Para el caso de este organismo existe también una reconstrucción manual disponible denominada, *T. thermophilus* iTT548 (Lee et al., 2014) cuyo porcentaje de reacciones bloqueadas es del 27%. Este valor muestra, por un lado, que la reconstrucción metabólica de *T. thermophilus* presenta dificultades incluso cuando el modelo es curado de forma manual y ajustado empleando datos experimentales. Por otra parte, al igual que para el caso de *E. coli*, las reconstrucciones curadas de forma manual reducen de forma muy notable el número de inconsistencias presentes en el modelo.

6.4.2 Análisis de consistencia global empleando un metamodelo

El análisis comparativo de las inconsistencias halladas en los 130 modelos muestra que un total de 480 reacciones (~22%) se encuentran bloqueadas en cada uno de los modelos en que están presentes (*i.e.* siempre bloqueadas). Esto sugiere que un número importante de inconsistencias estequiométricas están presentes a nivel de la base de datos metabólica empleada como referencia para realizar las reconstrucciones. Al agrupar las 480 reacciones “siempre bloqueadas” por subsistema¹³, se encuentra que el 25% de las mismas se encuentra involucradas en biosíntesis de componentes de la pared celular así como de lípidos de membrana (ver [tabla S.6.2D](#)). Esto muestra de que la inferencia automática de las rutas síntesis de componentes de la envoltura celular es altamente compleja, lo que se debe, en gran medida, a la naturaliza iterativa de dichos procesos.

La gran diversidad de componentes de la envoltura celular que existe en general y la falta de información experimental sobre la composición de membrana de la mayoría de los organismos analizados dificulta aún más la tarea. Además, un 17% de las reacciones bloqueadas se encuentran asociadas a subsistemas relacionados la biosíntesis de diversos cofactores y vitaminas, como por ejemplo las rutas de síntesis del sirohemo, la biotina y la vitamina B12. En el caso de la biotina, la ruta se encontraba bloqueada debido a que estaba incompleta en todos los modelos. La razón que se encontró es que cuando se reconstruyeron los GEMM del conjunto de datos, aún no se había resuelto la estructura bioquímica de dicha ruta. Esto representa un ejemplo ilustrativo de como los modelos deben evolucionar en paralelo con el desarrollo del conocimiento bioquímico. En otros casos, ciertas rutas se encontraron completas pero las reacciones estaban bloqueadas, como por ejemplo, la ruta de síntesis de vitamina B12. Esta se encuentra reconstruida de forma completa en diversos modelos y sin embargo, dado que estos modelos no incluyen reacciones de consumo para la vitamina (*e.g.* ecuaciones de biomasa), la ruta se encuentra bloqueada.

La detección de *gaps* recurrentes motivó un análisis de consistencia global, *i.e.* el análisis de consistencia de la red metabólica global o metamodelo que surge como la unión de diversos modelos individuales. Con este propósito, se reconstruyó un metamodelo denominado MM130.0 que incluye los 130 GEMM de metabolismos bacterianos analizados. El análisis de consistencia y la inspección visual del conjunto de UMs identificadas en MM130.0 permitió: i) detectar reacciones incluidas de forma errónea durante la reconstrucción de los modelos (*e.g.* reacciones pertenecientes a la metanogénesis); ii) detectar y completar los *gaps* en rutas metabólicas incompletas, mediante la adición de nuevas reacciones (*e.g.* el caso de la biosíntesis del ácido pimélico, precursor en la síntesis de la biotina); iii) la detección y resolución de *deadends* globales como el caso del S-adenosil-4-metiltio-2-oxobutanoato y de componentes no incluidos en las ecuaciones de biomasa de ningún GEMM.

6.4.3 Acoplamiento de flujos y módulos en el metamodelo

El análisis de acoplamiento a nivel de un metamodelo permite la identificación de módulos funcionales, *i.e.* conjuntos de reacciones que se requieren para poder realizar ciertas funciones. Parte del interés o motivación de realizar un FCA sobre el metamodelo consiste en la posibilidad de

13 De acuerdo con la definición empleada en la ontología del SEED, un subsistema es un conjunto de roles funcionales (*e.g.* actividades enzimáticas) que un anotador ha decidido que se encuentran relacionados. Los distintos subsistemas definidos suelen corresponderse con la colección de roles funcionales que componen una ruta metabólica, un complejo (*e.g.* el ribosoma), etc.

detectar “acoplamientos universales”. Estas relaciones estructurales sugieren la existencia de restricciones en el diseño de las rutas metabólicas y que por tanto se encuentren conservadas a lo largo del árbol de la vida. Por ejemplo, los resultados expuestos en el apartado 6.3.6 muestran que las actividades involucradas, tanto en rutas de biosíntesis, como en rutas de degradación, tienden a presentar cierto grado de acoplamiento. Un caso extremo lo representa la ruta de síntesis de la histidina, donde todas las reacciones conforma un FCS. Esto sugiere que muchas para rutas metabólicas la evolución solo ha encontrado un “solución” química. Además, los intermediarios de dichas rutas no tienen conexión con otras rutas metabólicas. Tales situaciones dan lugar a módulos cuya interfase con el resto de la bioquímica son el sustrato inicial y el producto final. El caso concreto de la histidina indica que dentro del universo bioquímico conocido, no solo existe una única ruta para la síntesis de un aminoácido fundamental como la histidina, sino que todas las actividades enzimáticas que participan en su síntesis conforman una unidad funcional, en principio, indivisible. Además, el hecho que la ruta de la histidina tenga como producto intermediario al AICAR, genera un acoplamiento direccional entre este proceso y las síntesis *de novo* de purinas. Esto podría tener consecuencias funcionales debido a que como ya fue mencionado, este acoplamiento representa un caso de punto débil en el metabolismo (Melendez-Hevia and Paz-Lugo, 2008). Este ejemplo ilustra además, cómo el FCA permite identificar acoplamientos estequiométricos entre procesos metabólicos, *a priori*, independientes y que al ser identificados en un metamodelo plantean cierta universalidad.

Contrariamente a los casos de rutas de degradación y biosíntesis, la mayor parte de las actividades enzimáticas involucradas en el metabolismo central (*e.g.* glicólisis, ciclo TCA, o ruta de las pentosas fosfato) prácticamente no presentan acoplamientos. Este resultado sugiere la gran versatilidad que existe en el metabolismo central. Además, si bien hay que tener en cuenta que al realizar el análisis de un metamodelo hay una yuxtaposición de muchas variantes de las rutas (lo que puede tener como efecto el que los acoplamientos se relajen), a diferencia de otras rutas metabólicas, la estructura del metabolismo central del carbono se encuentra muy conservadas a lo largo del árbol de la vida (Morowitz et al., 2000; Peregrín-Alvarez et al., 2009; Romano and Conway, 1996). Como consecuencia, el bajo acoplamiento detectado entre las actividades enzimáticas del metabolismo central no se debe al efecto antes mencionado si no que refleja, tanto de la robustez, como de la versatilidad funcional que exhibe el conjunto de dichas actividades enzimáticas (Barve et al., 2014; Noor et al., 2010). En efecto, este patrón en el cual el metabolismo central se encuentra muy desacoplado, mientras que las rutas de degradación y síntesis presentan un mayor nivel de acoplamiento se corresponde en buena medida con el concepto de arquitectura *bowtie* propuesto para describir el diseño del metabolismo (Csete and Doyle, 2004; R. Tanaka et al., 2005; Tanaka, 2005). Además, es notorio el hecho que la distribución de grado (tanto de entrada como de salida) en el grafo de acoplamiento del metamodelo ajuste de forma significativa a una distribución de potencia. Si bien este patrón se ha descrito para el caso organismos individuales como *S. cerevisiae*, *E. coli* y *H. pylori* (Burgard et al., 2004), el hecho de que un patrón similar se detecte sobre un metamodelo podría sugerir la existencia de cierta universalidad en la estructura de los acoplamientos entre el universo de reacciones bioquímicas. Además de su interés *per se*, la información que brindan las relaciones de acoplamiento puede emplearse para facilitar la reconstrucción y el refinado de modelos metabólicos. Esto es posible debido a que de acuerdo con las propiedades del acoplamiento, al considerar las relaciones entre las reacciones en un

metamodelo estas tienen a conservarse o a volverse más estrictas al considerar una subred (*i.e.* un subconjunto de reacciones del metamodelo). Esto es válido siempre que el modelo de un organismo pueda considerarse una subred del metamodelo.

6.4.4 Impacto de la consistencia de un metamodelo en los resultados del *gap-filling*

En muchos estudios enfocados a la resolución del problema del *gap-filling*, gran parte de la atención se ha puesto en el desarrollo de algoritmos que predigan reacciones candidatas para completar los *gaps* (Brooks et al., 2012; Kumar et al., 2007; Latendresse, 2014; Reed et al., 2006b; Vlassis et al., 2014). Sin embargo, se ha dedicado menos interés al problema de la consistencia de las bases de datos metabólicas empleadas como referencia para realizar una reconstrucción. Como se ha mostrado a lo largo de este capítulo, la consistencia de la base de datos de referencia es necesaria para poder reconstruir modelos consistentes de forma automática. Empleando dos metamodelos diferentes, se realizó un *gap-filling* sobre el conjunto de datos, para evaluar como la completitud y consistencia del metamodelo afectan los resultados del completado. Al comparar cada par de modelos obtenidos, se encontró que en la mayoría de los casos (125 de 130), la versión completada con el metamodelo consistente contiene incluyen, en buena medida a la versión del modelo completada con el metamodelo inconsistente. Esto es debido a que las reacciones, que son consistentes en ambos metamodelos conducen a soluciones similares al aplicar el realizar *gap-filling* sobre un modelo. Sin embargo, la versión del metamodelo sin refinar (MM130.0) incluye un mayor número de reacciones meta-bloqueadas que no pueden ser resueltas en los modelos.

El número de reacciones meta-bloqueadas es considerablemente menor en MM130.1 y por eso los modelos completados que se obtienen con este metamodelo incluyen un mayor número de reacciones. En el caso de los 5 modelos cuya diferencia entra las versiones refinadas con los dos metamodelos es menor al 89% esta se vio que varía entre 65-80% lo que indica que una gran diferencia notable en los resultados del *gap-filling* obtenidos con cada metamodelo. Estos resultados muestran que en algunos casos, los resultados del auto-completado automático se ven muy afectados por el metamodelo empleado como referencia. Estos cinco modelos presentaron la mayor divergencia entre sus dos versiones refinadas se corresponden con los modelos de menor tamaño. Dado que estos organismos corresponden a parásitos y simbioses obligados que se encuentran adaptados a la vida intracelular. Como consecuencia, sus metabolismos suelen encontrarse muy reducidos y fragmentados debido a la presencia de eventos de complementación metabólica, por lo que la reconstrucción metabólica de este tipo de organismos suele ser más susceptible a que los algoritmos de *gap-filling* automático predigan artefactos.

Otro aspecto analizado fue la eficacia del *gap-filling*, *i.e.* la relación entre el número total reacciones desbloqueadas y el total de reacciones huérfanas añadidas. En muchos casos los resultados indican una eficacia muy baja, y este resultado es independiente del metamodelo empleado para realizar el *gap-filling*. Esto indica que muchas de las reacciones bloqueadas pueden representar artefactos en lugar de rutas metabólicas incompletas. En tales casos, las reacciones deberían ser eliminadas del modelo en lugar de completar una ruta que el organismo no codifica. Por ejemplo, la reconstrucción inicial de *M. genitalium* G-37 (Opt243273.1) contiene 292 reacciones de las cuales 118 se encuentra bloqueadas. Después de realizar el *gap-filling*, las versiones auto-completadas con ambos metamodelos incluyen más de 500 reacciones. Por su parte,

la reconstrucción manual *M. genitalium* G-37 iPS189 contiene un total de 265 reacciones, de las cuales 90 no tienen gen asociado (Suthers et al., 2009), lo que muestra que el *gap-filling* automático de dicho modelo sobrestima de forma muy significativa el número de reacciones presentes en el metabolismo del patógeno. Estas discrepancias ponen de manifiesto que muchas de las reacciones inconsistentes deberían ser eliminadas en lugar de completar los *gaps*. Un ejemplo donde se ve claramente este fenómeno surge al analizar la consistencia de la ruta de síntesis de la histidina. En este caso se identificaron 20 modelos donde la ruta se encuentra bloqueada y se ve que algunos de los organismos codifican las actividades de la ruta. En otros casos es más plausible suponer que la ruta no está presente. De esto se deduce que la revisión de un metamodelo es una condición necesaria, pero no suficiente, para poder realizar refinado de modelos en forma automática.

6.4.5 Extensión de metamodelo mediante la adición de nuevos modelos

La extensión del metamodelo MM130.1 mediante la inclusión de 36 nuevos GEMM no considerados durante su reconstrucción inicial mostró, por un lado, la existencia de un número relevante de reacciones involucradas en el metabolismo de algunas bacterias, que no estaban presentes en MM130.0. Esto pone de manifiesto la necesidad de actualizar con cierta periodicidad, un metamodelo, o una base de datos metabólica empleada para reconstruir modelos metabólicos a escala genómica. Por otra parte, la nueva información bioquímica, en términos de nuevas reacciones, que se incorpora al metamodelo al añadir 36 nuevos modelos es relativamente pequeña si se tiene en cuenta que esta representa menos del 8% del total de reacciones incluidas en MM130.1, lo que indica que el conjunto de reacciones presentes en estos modelos está contenida en el metamodelo y solo una pequeña fracción de estas representa nueva información bioquímica. Sin embargo, debido a la naturaleza combinatorial de las innovaciones metabólicas, un número reducido de nuevas reacciones añadidas a un metamodelo puede aumentar en mayor medida el número de metabolismos que se pueden inferir a partir del mismo.

6.4.6 Consideraciones finales

Si bien ha habido un avance considerable en la reconstrucción automática de GEMMs, los modelos generados de forma automática aún contienen un alto porcentaje de inconsistencias estructurales, como muestran los resultados que se han presentado en las secciones previas. Además, la detección de inconsistencias recurrentes manifestadas en forma de reacciones metabólicas bloqueadas indica que, en muchos casos, los *gaps* pueden encontrarse presentes en el metamodelo o base de datos metabólica empleada para realizar las reconstrucciones. No obstante, la revisión de bases de datos biológicas en general es una tarea que insume mucho tiempo. Además, en el caso particular de las bases de datos metabólicas, esta tarea presenta la dificultad adicional que es lidiar con la complejidad que representan las redes metabólicas. En este sentido, los resultados del análisis presentado en este capítulo muestran como el CBM y sus extensiones para el análisis de consistencia, pueden ser empleados para analizar y refinar bases de datos metabólicas. Como consecuencia, el análisis de consistencia presentado en este capítulo puede servir a modo de protocolo para poder detectar y resolver los *gaps* e inconsistencias, así como eliminar artefactos e información redundante, en bases de datos metabólicas.

Capítulo 7

7 Discusión general

A lo largo de los distintos capítulos de la Tesis se ha hecho una discusión particular de los problemas planteados y resueltos, haciéndose énfasis en los aspectos concretos desarrollados en cada apartado. En el presente capítulo se presenta una discusión general, centrada fundamentalmente en dos puntos: i) los aspectos metodológicos desarrollados a lo largo de la Tesis, poniendo los resultados obtenidos en el contexto del refinamiento de los modelos metabólicos; y ii) sobre el problema de la emergencia y establecimientos de la complementación metabólica, en el contexto de la evolución de consorcios bacterianos.

Construcción y refinado de modelos metabólicos

Como se ha visto a lo largo de esta Tesis, la consistencia de un GEMM depende, en gran medida, de la calidad y precisión de la anotación del genoma correspondiente, así como de la consistencia de la base de datos metabólica o metamodelo empleados durante la reconstrucción (Latendresse et al., 2012; Pah et al., 2013; Vitkin and Shlomi, 2012). Los errores en la anotación de genes metabólicos conducen a la presencia de artefactos en la formulación de un modelo, los cuales pueden ser detectados mediante un análisis de consistencia, basado en CBM (Orth and Palsson, 2010). La aplicación del análisis de consistencia sobre diversos GEMMs, junto a la revisión de las anotaciones correspondientes realizados en este trabajo, han permitido identificar dos tipos o clases de errores de anotación: (1) la ausencia de determinadas anotaciones funcionales, *e.g.* un gen no identificado o de función desconocida; y (2) la asignación incorrecta de una actividad enzimática a un cierto gen. La presencia de errores de tipo (1) y (2) conducirá a la ausencia de ciertas reacciones y a la inclusión de reacciones incorrectas (*i.e.* artefactos), respectivamente. En un GEMM ambos tipos de error se manifestaran como metabolitos *gap* y reacciones bloqueadas. En el caso de errores tipo (1) los *gaps* deberán de ser completados. Sin embargo, los errores de tipo (2) implican que el GEMM incluye una o más reacciones que no tienen lugar en el metabolismo del organismo, y por tanto dichas reacciones deberían ser eliminadas del modelo (Ponce-de-León et al., 2015). El problema reside en cómo diferenciar aquellas reacciones bloqueadas que son consecuencia de un error tipo (1), de las reacciones añadidas de forma errónea y que corresponden a errores de tipo (2). Como se ha visto en este trabajo, la distinción entre ambos tipos de errores dista mucho de ser trivial. En efecto, si no se consideran ambas situaciones con precauciones, un *gap-filling*

automático podría sobrestimar el complemento metabólico de un organismo introduciendo reacciones para resolver *gaps* que son consecuencia de errores de tipo (2). En tales casos, las reacciones inconsistentes deberían ser eliminadas en lugar de completar *gaps* ficticios.

El problema se vuelve aún más complejo al considerar organismos que exhiben complementación metabólica como es el caso de los endosimbiontes (Calle-Espinosa et al., 2016). Esto se debe a que si una ruta metabólica se encuentra distribuida entre dos organismos, al considerar el metabolismo de un único miembro de la asociación se tiene una visión parcial del sistema (Ponce-de-Leon et al., 2013). En efecto como se vio en los capítulos 4 y 5, muchas rutas metabólicas reconstruidas en modelos de endosimbiontes desembocan en *deadends*, que pueden ser rutas complementas, y los *deadends* metabolitos que son intercambiados. Sin embargo, la mayoría de los casos estos compuestos carecen de transportadores específicos y por lo tanto son intercambiados por difusión. Al igual que los errores de tipo (2) la presencia de *deadends* aparentes puede conducir a un completado incorrecto de *gaps*, cuando la solución consistiría en añadir los transportes por difusión. Sin embargo, los metamodelos no suelen incluir transportes por difusión para metabolitos que son intermediarios de rutas metabólicas, ya que la mayoría de las membranas no son permeables a estos compuestos. Un primer paso para resolver este problema, podría consistir en estimar la capacidad de difundir a través de una bicapa lipídica, de todos aquellos metabolitos *gap* de tipo RNP o RNC (*root deadends*) detectados. Empleando esta información se debería expandir el metamodelo añadiendo transportes para todos los *deadends* que potencialmente puedan difundir a través de una membrana celular y finalmente realizar un *gap-filling*.

Una característica que comparten la mayoría de los algoritmos de *gap-filling* automáticos basados en el CBM es que resuelven el problema de forma global. El problema consiste en encontrar un modelo consistente que implique la menor cantidad de cambios respecto a la versión original. Es decir, que la solución óptima que estos buscan es aquella en la que haya que introducir un número mínimo de cambios en el modelo (*e.g.* añadir nuevas reacciones, relajar la irreversibilidad de ciertas reacciones, etc.) para que el modelo sea consistente (Kumar et al., 2007; Vlassis et al., 2014). Sin embargo, tales algoritmos no incluyen de forma explícita la posibilidad de eliminar reacciones que se consideren artefactos. El cálculo de los módulos inconexos (ver apartado 3.6) brinda la posibilidad de resolver las inconsistencias de un modelo de forma iterativa. Por ejemplo, después de identificar el conjunto de UMs en un modelos se podría evaluar si un módulo inconexo en particular debe ser completado o removido del modelo, considerando los genes codificantes asociados a las reacciones involucradas en el UM, las propiedades físicoquímicas de los metabolitos *gap* de tipo RNP o RNC y el número de reacciones “huérfanas” que se deberían añadir para re-conectar el UM. En perspectiva, sería posible automatizar el proceso de análisis y resolución de los UMs, para que con esta información, un experto puede decidir, *a posteriori*, cuáles son los UM que tienen un correlato biológico y deben ser curados, y cuáles representan alguna clase de artefacto y deben ser eliminados del modelo.

Las principales ventajas que el análisis de los UMs presenta frente a otras estrategias de resolución de inconsistencias (Kumar et al., 2007; Latendresse, 2014; Reed et al., 2006b) es que es un cálculo que no requiere el ajuste de parámetros y el costo computacional es virtualmente nulo. Por otra parte, la aproximación de los UM puede también emplearse para interpretar otro tipo de problemas, como por ejemplo asistir en la interpretación de los resultados de experimentos de

knockout. En dicho caso, el cálculo y análisis de UMs se podría emplear para identificar qué módulos del metabolismo se desconectan como consecuencia del *knockout* de una o más enzimas. En el contexto del estudio de endosimbiontes, esta misma aproximación podría también emplearse para predecir eventos de complementación metabólica.

Evolución de la complementación metabólica

La evolución de la complementación metabólica en consorcios endosimbiontes y de las interacciones nutricionales en general se encuentra moldeada por múltiples factores, muchos de los cuales actuando en forma simultánea (Tan et al., 2015). En el caso de la evolución de endosimbiontes, un factor clave es el proceso de reducción genómica irreversible, o trinquete de Muller, que surge como consecuencia de la vida intracelular obligada y los cuellos de botella a los que están sujetas las poblacionales de endosimbiontes (Mira and Moran, 2002; Moran, 1996). Este proceso tiene como consecuencia que los genes cuya pérdida no represente un efecto deletéreo para el sistema se fijarán a nivel de la población por el efecto de la deriva génica. Además, la vida intracelular obligada limita la posibilidad de adquirir nuevos genes por transferencia horizontal. Los casos en que ocurre, los genes son transferidos del simbionte al hospedador (Husnik et al., 2013; Nikoh et al., 2010; Nikoh and Nakabachi, 2009). En contraste, un hospedador puede ser infectado por nuevas bacterias que devengan en endosimbiontes secundarios. En los casos donde múltiples endosimbiontes coexisten, es posible que existan rutas metabólicas redundantes entre los genomas de los endosimbiontes. Esto podría dar lugar a que genes, antes esenciales, pasen a ser prescindibles ante la presencia de un endosimbionte secundario. Estas situaciones pueden dar lugar a la posibilidad de emergencia de complementación metabólica. Un escenario posible es que la complementación no represente una ventaja selectiva, sino que sea una consecuencia neutral del propio proceso de reducción genómica.

Alternativamente, el establecimiento de una complementación metabólica podría representar una estrategia que presente ventajas desde un punto de vista selectivo. La división de tareas puede reducir el coste que conlleva la síntesis de compuestos esenciales como cofactores y aminoácidos y de las enzimas necesarias para su síntesis. Esta reducción sería benéfica, tanto para los endosimbiontes como para el hospedador. Los experimentos realizados con el GEMM del consorcio mostraron que la complementación metabólica puede exhibir rendimientos *cuasi*-óptimos con un número mínimo de genes. Además, el estudio cinético permitió ver que en ciertas condiciones la complementación es la estrategia que permite minimizar la concentración total de enzimas. Por lo tanto, aquellos hospedadores cuyos simbiontes sean capaces de satisfacer sus demandas nutricionales en forma más eficiente, exhibirán un incremento en sus tasas de crecimiento (Shachrai et al., 2010). Estos resultados podrían estar relacionados con los experimentos realizados con comunidades de *cross-feeders* sintético que intercambian distintos aminoácidos (Hosoda et al., 2011; Kerner et al., 2012; Mee et al., 2014). Para ciertas comunidades sintéticas, los experimentos muestran que los organismos cooperadores crecen de forma estable y exhiben tasa de crecimiento mayores que la de las cepa silvestre (Germerodt et al., 2016; Pande et al., 2014). Estos resultados experimentales muestran como la división de tareas permite un aumento de la eficiencia tanto a individual como poblacional. En efecto, el coste de sobre-producir un único aminoácido e intercambiarlo por otro es significativamente menor al coste de producir ambos aminoácidos para uno (Mee et al., 2014; Pande et al., 2014). Sin embargo, para que este tipo de

interacciones sean benéficas la estructuración espacial es un factor clave (Germerodt et al., 2016; Pande et al., 2016; Pfeiffer and Schuster, 2005). El caso de los consorcios endosimbiontes es un caso particular de comunidad microbiana con una estructura espacial claramente definida. La complementación metabólica podría considerarse como un caso particular de *cross-feeding*. Si bien *sensu stricto*, *cross-feeding* es un término que se usa para referir a las interacciones nutricionales en las que se intercambian productos “finales”, tales como aminoácidos o cofactores (Seth and Taga, 2014). Sin embargo, el concepto podría extenderse al caso de la complementación metabólica ya que implica un intercambio de compuestos que posibilita el crecimiento de ambos organismos.

El fenómeno de la complementación metabólica puede también estar relacionado a la hipótesis de la reina negra (BQH, *Black Queen Hypothesis*) (Morris et al., 2012). La BQH sostiene que siempre que una función metabólica sea redundante entre los miembros de una comunidad, y que el producto final de dicho proceso se permeable y por lo tanto parte se escape fuera del organismo que lo produce, el resto de los miembros de la comunidad podrán aprovechar el “bien público”. De esta manera, la BQH predice que la mayor parte de los organismos tenderán a perder los genes que codifican dicha función, hasta que solo un único miembro brinde el servicio a toda la comunidad. La explicación se basa en que la pérdida de la función confiere una ventaja selectiva al organismo, en términos del ahorro que significa emplear un bien común en lugar de producirlo. En forma similar, la interpretación de los incrementos en las tasas de crecimiento de los miembros observadas en los experimentos con comunidades sintéticas de *cross-feeders*, radica en la misma idea de que la división de tareas permite un aumento de la eficiencia individual. En este sentido, la emergencia de la complementación metabólica podría considerarse una consecuencia de la BQH. La comprensión de los principios y mecanismos que rigen la evolución de los ecosistemas microbianos puede ser de gran ayuda en el diseño de comunidades sintéticas de interés potencial para la industria (Großkopf and Soyer, 2014; Hosoda and Yomo, 2011). Las aproximaciones desarrolladas en esta Tesis se podrían extender para modelar este tipo de comunidades. Sin embargo, la consideración de mecanismo de regulación alostérica es aún un problema difícil de abordar en GEMM. Aunque es posible que esta aproximación permita eventualmente el diseño de comunidades sintéticas con aplicaciones en la industria (Hosoda and Yomo, 2011).

Capítulo 8

8 Conclusiones

Este trabajo se ha centrado en el desarrollo de metodologías para la reconstrucción y análisis de modelos metabólicos reconstruidos a escala genómica, y su aplicación estudio de modelos metabólicos de bacterias endosimbiontes, con énfasis en el fenómeno de la complementación metabólica. De acuerdo con este planteamiento a continuación se exponen las principales conclusiones del estudio realizado y que configuran la presente Tesis doctoral:

1. La definición de metabolito *gap*, así como su extensión al caso de los *pseudo-gaps*, combinada con la definición previa de reacción bloqueada, ha permitido introducir el concepto de módulo inconexo así como un método para su cálculo. La aplicación de estas aproximaciones a la revisión del modelo de *Blattabacterium* Bge iCG238, ha permitido identificar y resolver todas las inconsistencias detectadas, dando como resultado una versión refinada del modelo denominada iMP240.
2. Más allá del gran avance en la reconstrucción automática de GEMMs, los modelos generados aún contienen un alto porcentaje de inconsistencias estructurales. El análisis a gran escala realizado, ha permitido detectar la presencia de inconsistencias recurrentes que a su vez indican que tales *gaps* se encuentran presentes en el metamodelo o base datos metabólica empleada durante la reconstrucción de los modelos. La reconstrucción y revisado del metamodelo denominado MM166, ha permitido mostrar que el CBM puede ser empleado en la revisión y refinado de bases de datos metabólicas.
3. La reconstrucción metabólica de *B. aphidicola* BCc y *S. symbiotica* SCc, endosimbiontes del áfido del cedro, dio como resultado los respectivos modelos metabólicos denominados, iBCc98 y iSCc236. El análisis de dichos modelos ha permitido la revisión y corrección de las anotaciones genómicas correspondientes, así como una mejora en la predicción de las condiciones de crecimiento y las capacidades metabólicas de cada endosimbionte. Esto muestra la importancia del uso de modelos en el estudio de capacidades metabólicas de microorganismo no cultivables.

4. Combinando *iBCc98* e *iSCc236* se reconstruyó un modelo compartimentado del consorcio. El modelo denominado *iBSCc* reproduce los eventos de complementación descritos entre las bacterias, así como un nuevo caso no descrito previamente. Además, el análisis *in-silico* realizado sobre la biosíntesis distribuida del triptófano, muestra que el diseño de la complementación en dicha ruta, exhibe un rendimiento de producción cuasi-óptimo, empleando un número mínimo de genes, lo que sugiere que la fijación de dicho evento pudo haber sido consecuencia de un proceso selectivo.
5. El desarrollo y de un modelo simple para el análisis cinético de la complementación, muestra que: i) en presencia de inhibición enzimática, la complementación metabólica es la estrategia óptima que minimiza la concentración enzima necesaria para satisfacer la demanda del producto; y ii) que las propiedades fisicoquímicas de los intermediarios de la ruta pueden determinar, o al menos restringir, el intermediario intercambiado en una ruta complementada.

Modelado Basado en Restricciones

Apéndice A

A.1 Introducción al CBM

A.1.1 Definiciones generales

Este apéndice es una breve introducción al modelado basado en restricciones. A continuación se introducirán algunos términos relacionados con la estequiometría y la estructura de las redes bioquímicas que serán empleados a lo largo de la sección (Klamt and Stelling, 2006). Una reacción bioquímica se puede caracterizar mediante las siguientes propiedades: i) la estequiometría; ii) la dirección en que ocurre la reacción; iii) la enzima que cataliza la reacción; y iv) la cinética de la reacción. En el análisis estructural, sólo se tienen en cuenta las tres primeras propiedades, de forma que los aspectos dinámicos no se consideran (Palsson, 2006). La estructura de un red metabólica se puede definir en términos de cómo las sustancias se encuentran conectadas unas con otras a través de los flujos de las reacciones (Mahadevan and Palsson, 2005). De esta manera la estructura de la red viene dada por la estequiometría de las propias reacciones. En este contexto una red metabólica estará definida por un conjunto de metabolitos y un conjunto de reacciones, que se nombrarán como I y J , respectivamente y se llamará $J_{Irrev} \subseteq J$ al subconjunto de reacciones irreversibles.

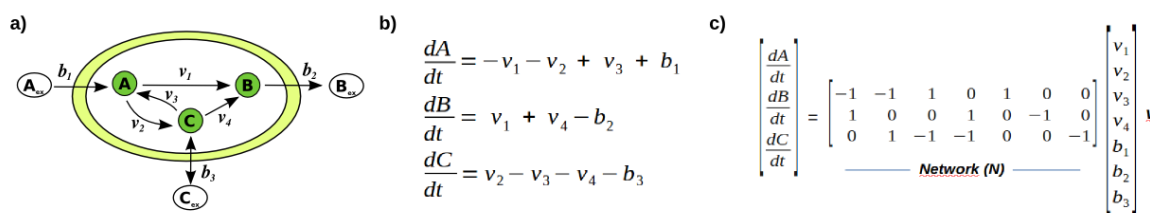


Figura A.1. Representaciones alternativas un sistema metabólico. En a) se representa un sistema compuesto por tres metabolitos A, B y C internos que participan en diversas reacciones; el diagrama también incluye los correspondientes metabolitos externos, A_e , B_e , y C_e . En b) se representan las ecuaciones de balance de los distintos metabolitos. En c) se representa el sistema de ecuaciones diferenciales en formato matricial.

La [figura A.1](#) muestra el diagrama de una red metabólica de ejemplo, junto a las variables correspondientes que describen el estado del sistema. Esta red, incluye tres metabolitos (internos), tres metabolitos (externos) y siete reacciones (todas irreversibles excepto b_3). De acuerdo con lo

anterior, una característica de cualquier sistema metabólico es la existencia de una frontera o contorno que define la interfase de intercambios con el ambiente exterior. A partir de esta idea de frontera, surgen las nociones de metabolitos *internos* y *externos*. Por un lado, los metabolitos *internos* son aquellas especies que son producidas y consumidas por los procesos internos del sistema y se consideran de forma explícita en la definición de la red. Por otra parte, los metabolitos externos representan fuentes o sumideros externos al sistema y por tanto no son incluidos de forma explícita en la descripción del sistema (Heinrich and Schuster, 1996). La descripción matemática de una red metabólica puede realizarse a través de las ecuaciones de balance de cada metabolito (Heinrich and Schuster, 1998). Estas describen la relación entre los cambios de concentración de cada metabolito y de los flujos de las reacciones que los producen y consumen. De esta manera, dada una red metabólica con I metabolitos y J reacciones, las ecuaciones de balance correspondientes a cada metabolito se pueden escribir de la siguiente forma:

$$\sum_{j \in J} n_{ij} \cdot v_j = b_i \quad \forall i \in I \quad (\text{Ec. A.1})$$

Donde v_j es el flujo a través de la reacción j ; n_{ij} es el coeficiente estequiométrico del metabolito i en la reacción j ; y b_i es la tasa de cambio de la concentración del metabolito i con respecto del tiempo, que como se verá más adelante, corresponden a los flujos de intercambio (apartado A.1.3). Como se muestra en la figura A.1 el sistema de ecuaciones A.1 se puede representar en forma más compacta empleando notación matricial:

$$N \cdot v = b \quad (\text{Ec. A.2})$$

Donde N es la matriz estequiométrica de $|I|$ filas (metabolitos) \times $|J|$ columnas (reacciones); v es un vector de dimensión $1 \times |J|$, donde cada entrada v_j es el flujo a través de la reacción j ; y b es el vector de las derivadas de las concentraciones de los metabolitos respecto del tiempo.

A.1.2 Imposición de restricción

Al igual que cualquier sistema físico, los sistemas biológicos se encuentran sujetos a distintos tipos de restricciones que incluyen, en primer lugar las leyes de conservación de materia y energía. Dichas restricciones pueden emplearse para reducir el espacio de posibles estados del sistema (Price et al., 2004). En el caso del metabolismo la restricción de la conservación de la masa está implícita en las ecuaciones de balance de cada metabolito. Dado que el objetivo del análisis estequiométrico es caracterizar las propiedades estructurales de una red metabólica, se supone que el sistema opera en estado estacionario, de manera que las concentraciones de los distintos metabolitos no cambian con el tiempo. Al imponer dicha condición sobre las ecuaciones de balance (Ec. A.2) se obtiene un sistema homogéneo de ecuaciones lineales:

$$N \cdot v = 0 \quad (\text{Ec. A.3})$$

De esta manera, cualquier vector v compatible con la ecuación A.3 es una distribución de flujos estacionaria. Sin embargo, en las condiciones fisiológicas de una célula muchas reacciones bioquímicas pueden considerarse procesos irreversibles. Si se sabe qué reacciones de la red son irreversibles es posible emplear esa información para añadir nuevas restricciones al sistema que

1 Por convención el signo de n_{ij} es negativo (positivo) si el metabolito i es consumido (producido) por la reacción j .

permitan eliminar estados que no son compatibles con la termodinámica de alguno de los procesos involucrados. La irreversibilidad de las reacciones puede formularse de forma sencilla introduciendo restricciones de signo sobre el valor del flujo de la siguiente forma:

$$v_j \geq 0, \quad \forall j \in J_{irrev} \quad (\text{Ec. A.4})$$

En donde $J_{irrev} \subseteq J$ es el conjunto de reacciones irreversibles, y la restricción garantiza que las reacciones irreversibles en el sentido compatible con la termodinámica. Al considerar de la restricción ecuación A.3) junto con A.4 el espacio de los posibles estados metabólicos queda restringido a las soluciones del sistema lineal y homogéneo de igualdades y desigualdades, definidas por la matriz estequiométrica N y el conjunto de reacciones irreversibles J_{irrev} (Gagneur and Klamt, 2004). Dicho espacio de soluciones asociado a una red, es comúnmente referido como el cono o el espacio de flujos definido mediante la siguiente expresión:

$$F = \{v \in \mathcal{R}^{|J|} : N \cdot v = 0, \quad v_j \geq 0 \quad \forall j \in J_{irrev}\} \quad (\text{Ec. A.5})$$

El espacio de flujos F contiene todos posibles estados en que puede funcionar asociado la red metabólica asociada. En lo que sigue, se denominará como distribución de flujos a cualquier vector \vec{v} que pertenezca a F . El término “modelado basado en restricciones” (Constraint-based Modelling, CBM) proviene del hecho que la definición del espacio de flujos, que se puede pensar como el “genotipo” de la red, se define mediante la imposición de restricciones que reducen o limitan los posibles estados funcionales, o fenotipo, que podría adoptar un organismo. Dentro del marco metodológico que implica el CBM, pueden incluirse aquellas aproximaciones desarrolladas para describir la estructura del cono de flujos F y que son referidas con el nombre genérico de análisis de rutas metabólicas o *pathway analysis* (Klamt and Stelling, 2003). El objetivo del análisis de rutas metabólicas es identificar un conjunto de distribuciones de flujo que permitan generar todo el espacio de flujos asociados una red (Schilling et al.). Una de dichas aproximaciones es el análisis de modos de flujo elementales (Schuster et al., 2000). En dicha aproximación se introdujo el concepto de modo elemental así como un algoritmo para calcularlos. Previo a introducir la definición de modo elemental se define el concepto modo de flujo:

$$M = \{\lambda v : v \in F, \quad \lambda > 0\} \quad (\text{Ec. A.6})$$

donde M es el conjunto de distribuciones de flujo escaladas por un escalar positivo. Para un cierto modo de flujos v denotaremos $R(v) = \{j : v_j \neq 0\}$, denominado soporte de v , al conjunto de índices de las reacciones que poseen un flujo distinto de cero. De esta manera $R(v)$ puede interpretarse como el conjunto de enzimas que operan en v . De acuerdo con lo anterior, un modo de flujo $e \in F$ se dice elemental si y solo si se cumple que ningún otro modo de flujo $e' \in F$ posee un conjunto de reacciones activas $R(e')$ tal que $R(e') \subset R(e)$ (Gagneur and Klamt, 2004). La definición implica que si un modo es elemental, entonces si se elimina alguna enzima, este no podrá operar en estado estacionario. La condición de elementalidad se puede expresar de la siguiente forma:

Definición A.1: dada una red metabólica con matriz N y J_{irrev} reacciones irreversibles, un modo de flujo $e \in F$ es elemental si y solo si se cumple que

$$\forall \vec{e}' \in F: R(e') \subseteq R(e) \text{ solo una de las siguientes condiciones es cierta: } \begin{cases} e'=0 \\ e' \simeq e \\ e' \simeq -e \end{cases}$$

esto es, o que \vec{e}' es el vector nulo o que \vec{e}' y \vec{e} son el mismo modo de flujo o \vec{e} es un modo de flujo reversible operando en un sentido y \vec{e}' es el modo operando en sentido reverso.

Tabla A.1. Conjunto de modos elementales asociado a la red de la figura A.1. Cada fila corresponde a un modo de flujo elemental y cada columna a un flujo de la red. Las entradas corresponden al valor del flujo de la reacción en el modo. Los flujos se encuentran separado en reacciones internas y flujos de intercambio (para un definición de flujo de intercambio ver apartado A.1.3).

	Reacciones interna				Flujos de intercambio		
	v_1	v_2	v_3	v_4	b_1	b_2	b_3
EM1	0	0	0	1	0	1	-1
EM2	1	0	0	0	1	1	0
EM3	0	1	0	1	1	1	0
EM4	0	1	0	0	1	0	1
EM5	1	0	1	0	0	1	-1
EM6	0	1	1	0	0	0	0

La red de ejemplo de la figura A.1 contiene un total de seis modos de flujo elementales. El conjunto de modos se encuentra representado en formato matricial en la tabla A.1. El sistema incluye un total de cinco modos de flujo elemental que permiten la producción neta de alguno de los metabolitos, a partir del consumo de otro (e.g. $A \rightarrow B$), pero además el sistema incluye un modo cíclico, EM6. Dicho modo, si bien no viola ninguna de las restricciones impuestas por las ecuaciones A.3 y A.4 y por tanto es un modo de flujo, no tiene intercambio con el ambiente. Como consecuencia dicho modo no podría exhibir flujos diferentes de cero en el estado estacionario o de lo contrario violaría las leyes de la termodinámica (ver sección 2.3). A partir de esta observación se pueden distinguir dos clases de modos elementales: i) los modos en los cuales hay intercambio con el ambiente, y ii) los modos cíclicos que si bien son soluciones del espacio de flujos carecen de significado físico debido a que violan el segundo principio de la termodinámica².

A.1.3 Flujos de intercambio con el ambiente

El espacio de flujo definido en CBM puede ser modificado mediante la introducción otras clases de restricciones que permitan reducir el espacio de flujos o los estados funcionales asociados a una red metabólica. Por ejemplo, se han propuesto extensiones que permiten introducir restricciones regulatorias derivadas de la redes de regulación génica, también se han propuesto para introducir restricciones de tipo cinéticas. En particular, es de interés poder introducir algún tipo de restricción que permita definir las condiciones de crecimiento del organismo. Como se menciona al inicio de esta sección, al considerar un sistema metabólico surgen dos clases de metabolitos, los metabolitos

2 Estas situaciones suelen confundirse con los denominados ciclos fútiles. Pero estos últimos, a diferencia con los modos cíclicos, se encuentran acoplados a un flujo que va de una fuente a un sumidero (e.g. la hidrólisis del ATP en ADP) y como consecuencia son termodinámicamente factibles.

internos y externos, en general los metabolitos externos no se suelen considerar al realizar el análisis de una red metabólica. Sin embargo, al trabajar con modelos metabólicos reconstruidos a escala genómica, es de gran utilidad considerar en el análisis a los metabolitos externos ya que, como se verá a continuación, estos pueden ser empleados para introducir restricciones que definan el ambiente o medio de crecimiento de un organismo. De esta manera los metabolitos extracelulares (que al considerar un modelo a escala genómica son los metabolitos externos) se pueden representar de forma explícita si se realizan ciertas modificaciones en el modelo.

En primer lugar se deben añadir a la matriz estequiométrica las filas correspondientes a los metabolitos del medio extracelular que pueden ser importados o exportados por la célula. Denotaremos como $I' = I \cup I_{Ex}$ al conjunto expandido de metabolitos que incluye tanto a los metabolitos internos como a los metabolitos extra-celulares. Al considerar de forma explícita las ecuaciones de balance de los metabolitos extra-celulares se debe relajar la restricción de estado estacionario sobre estos metabolitos. Esto puede hacerse empleando las variables auxiliares b_i ya que representan el balance sobre cada metabolito (Ec. A.1 y Ec. A.2). Para simplificar la notación, las variables b_i correspondientes a los metabolitos extracelulares se pueden ser tratar como flujos, lo que da como resultado la siguiente matriz estequiométrica extendida:

$$N' = \begin{pmatrix} N & 0 \\ N_{Ex} & -Id \end{pmatrix} \cdot \begin{pmatrix} v \\ b \end{pmatrix} = 0 \quad (\text{Ec. A.7})$$

donde N' es la matriz estequiométrica expandida, $v' = (v, b)^T$ es el vector de flujos expandido que incluye a los flujos de intercambio (*i.e.* $J' = J \cup J_{Ex}$); $-Id$ es un matriz de identidad³; y N_{Ex} contiene los coeficientes estequiométricos de las reacciones en que participan los metabolitos extracelulares (*e.g.* procesos de transporte). Las variables $b_j \forall j \in J_{Ex}$ se denominan flujos de intercambio y si bien son tratadas como flujos, son variables auxiliares empleadas para representar el balance neto de cada metabolito extracelular. De esta forma, el signo de un flujo de intercambio b_i indica si el metabolito i está siendo producido ($b_i > 0$) o consumido ($b_i < 0$) por el sistema. Finalmente, con el objetivo de representar distintas condiciones de crecimiento, se pueden introducir cotas sobre los flujos de intercambio, que definan qué metabolitos podrán ser consumidos y cuáles podrán ser producidos por el sistema:

$$\alpha_i \leq b_i \leq \beta_i \quad (\text{Ec. A.8})$$

Donde α_i y β_i son parámetros que permiten definir el estatus de un flujo de intercambio. Por ejemplo, si quiere considerar al metabolito i como un sumidero, α_i será definido como cero y β_i como más infinito. De acuerdo con lo anterior, el estatus de fuente/sumidero de cada metabolito externo i queda definido por los valores de sus cotas (Ec. A.7) de la siguiente forma:

- $0 \leq b_i \leq +\infty$ (Sumidero)
- $-\infty \leq b_i \leq 0$ (Fuente)
- $-\infty \leq b_i \leq +\infty$ (Fuente/Sumidero)

3 El coeficiente estequiométrico de un metabolito i asociado a un flujo de intercambio b_i es por convención -1. De esta forma para valores positivos de b_i se considera que el metabolito es producido por el sistema mientras que si es negativo el metabolito es consumido por el sistema.

En términos prácticos, para definir el estatus de fuente o sumidero, es suficiente con definir los valores de los parámetros α_i y β_i lo suficientemente grandes (o pequeños). En forma adicional, es posible emplear medidas experimentales de tasas netas de consumo/excreción para ajustar las cotas de los flujos de intercambio con valores experimentales. A lo largo de este trabajo, siempre que se hable de la matriz estequiométrica en el contexto de modelos metabólicos reconstruidos a escala genómica, se estará haciendo referencia a la versión extendida de la matriz (Ec. A.6), la cual incluye a los metabolitos extracelulares y los correspondientes flujos de intercambio. Además, el uso de los flujos de intercambio permite discriminar de forma clara los modos de flujo cíclicos ya que son aquellos modos en los que todos los flujos de intercambio adoptan el valor cero.

A.1.4 La ecuación de Biomasa

Los modelos metabólicos a escala genómica permiten calcular diversos estados fenotípicos, en particular estados que predicen la formación de biomasa, o el crecimiento del organismo (Feist and Palsson, 2008). Para predecir el crecimiento de un organismo, primero es necesario definir “crecimiento” a nivel del modelo. Una aproximación ampliamente extendida es el uso de una ecuación de biomasa, la cual se utiliza como función objetivo en diversos métodos y extensiones del modelado basado en restricciones (Lewis et al., 2012; Price et al., 2004). La ecuación de biomasa es una reacción ficticia que describe los requerimientos nutricionales y energéticos de una célula para su crecimiento (ver figura A.2). La formulación de una ecuación de biomasa requiere conocer la composición de una célula en términos de sus componentes moleculares, así como sus proporciones relativas. De esta forma el nivel de detalle de la ecuación de biomasa formulada para un cierto organismo dependerá, en gran medida, de la información experimental disponible sobre la composición de la célula así como los requerimientos energéticos necesarios para generar la biomasa a partir de los diversos precursores.

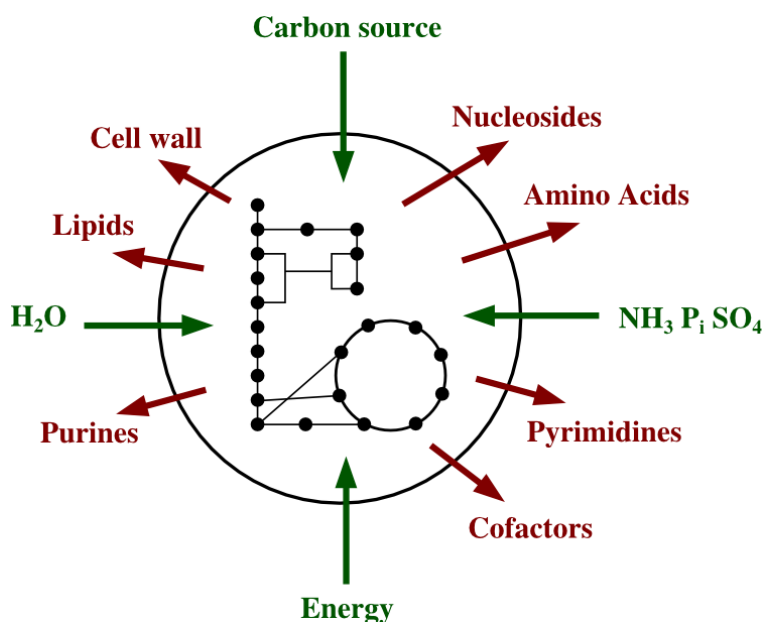
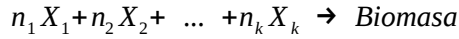


Figura A.2. Figura esquemática que representa como la célula transforma nutrientes en componentes de la biomasa. El esquema representa una red metabólica y los intercambios con el ambiente que tienen lugar durante el crecimiento del organismo. En verde se indican los nutrientes que son consumidos por el organismo (flechas hacia el interior de la célula) y en rojo los componentes de la biomasa que se sintetizan durante el crecimiento (flechas hacia el exterior del sistema).

El proceso de formulación comienza mediante la definición del contenido macromolecular de la célula (*i.e.* la fracción en peso de las proteínas, RNA, lípidos, etc.) y luego los metabolitos con de los que se compone cada clase de macromolécula (*e.g.* aminoácidos, lípidos, nucleótidos). De esta manera se puede formular una ecuación de biomasa en términos de las cantidades de cada metabolito requeridas para producir biomasa. La ecuación se incluye como una reacción más del modelo y tiene la siguiente forma general:



Donde X_i son los distintos componentes de la biomasa (*e.g.* aminoácidos, lípidos, nucleótidos) y n_i es el coeficiente estequiométrico o la proporción de dicho compuesto en 1g de peso seco de célula (Varma and Palsson, 1994). Posiblemente, la primera cuantificación a nivel molecular de los componentes de la biomasa de un organismo haya sido realizada para *E. coli* K12 por Neidhardt y Curtiss (1987). Los autores describen en una tabla las proporciones de los distintos componentes de las macro moléculas. Esta misma información fue la empleada en los trabajos pioneros de Varma y Palsson para introducir una ecuación de biomasa en un modelo estequiométrico (Varma and Palsson, 1993). En dichos trabajo, los autores emplean programación lineal para calcular distribuciones de flujos que optimicen la formación de biomasa dando origen al Análisis de Balance de Flujos (FBA), que es introducido en el [apartado A.2.1](#).

A.2 Métodos generales

A.2.1 Análisis de Balance de Flujos

El Análisis de Balance de Flujos (Flux Balance Analysis, FBA) es una aproximación computacional empleada en el análisis de modelos metabólicos reconstruidos a escala genómica. Esta aproximación combina la descripción del espacio de flujos F definido por la ecuación A.5, con técnicas de optimización (Orth et al., 2010; Savinell and Palsson, 1992; Varma and Palsson, 1994). Utilizando una ecuación de biomasa (ver [apartado A.1.3](#)), el FBA permite predecir la distribución de flujos v que maximiza la tasa de crecimiento de un organismo. El FBA se plantea como problema de programación lineal canónico de forma que el cálculo de la distribución de flujos óptima se puede realizar empleando el método simplex:

$$\begin{aligned} Z &= \text{Max } v_{\text{Biomass}} \\ \text{s.t.} \\ N \cdot \vec{v} &= 0 \\ \alpha_j &\leq v_j \leq \beta_j \quad \forall j \in J \\ \alpha_j &= 0 \quad \forall j \in J_{\text{Irrev}} \end{aligned} \tag{LP A.1}$$

Donde v_{Biomass} representa la tasa de producción de biomasa, o crecimiento del organismo. Los parámetros α y β representan las cotas cotas máximas y mínimas, respectivamente, que son impuestas sobre cada reacción, y J_{Irrev} corresponde al conjunto de reacciones irreversibles. Esta aproximación ha permitido predecir tasas de crecimiento (Edwards and Palsson, 2000), tasas de consumo de nutrientes (Edwards et al., 2001), genes esenciales, entre otras muchas aplicaciones (Feist and Palsson, 2008; McCloskey et al., 2013; Oberhardt et al., 2009).

A.2.2 Análisis de Variabilidad de Flujos

El Análisis de Variabilidad de Flujos (Flux Variability Analysis, FVA) es una extensión del FBA desarrollada para estudiar la presencia de óptimos alternativos⁴ que puede exhibir el modelo de un cierto organismo (Mahadevan and Schilling, 2003). La idea es medir la “holgura” que tiene cada el flujo de cada reacciones cuando el sistema opera en estado óptimo o *cuasi*-óptimo. El algoritmo puede dividirse en dos fases. En la primera fase se emplea FBA para calcular la tasa de crecimiento máxima: $v_{biomasa}^*$. En la segunda etapa, el valor óptimo $v_{biomasa}^*$ se emplea para formular la siguiente la restricción $v_{biomasa} \geq \lambda v_{biomasa}^*$, $\lambda \in [0-1]$, donde λ es un parámetro que permite variar la distancia respecto al óptimo. A continuación, el rango de variabilidad de los flujos se calcula de forma iterativa resolviendo un conjunto de problemas lineales donde en cada iteración el flujo a través de cada reacción de la red es maximizado y minimizado y donde el valor objetivo del problema original es impuesto como restricción.

$$\forall j \in J$$

$$Z = \text{Min/Max } v_j$$

s. t .

(LPA.2)

$$N \cdot \vec{v} = 0$$

$$\alpha_j \leq v_j \leq \beta_j \quad \forall j \in J$$

$$\alpha_j = 0 \quad \forall j \in J_{Irrev}$$

$$v_{biomasa} \geq \lambda v_{biomasa}^*, \quad \lambda \in [0-1]$$

donde $v_{biomasa}^*$ es la tasa de crecimiento máxima predicha por FBA y λ es el parámetro de holgura con el que se controla la distancia respecto al óptimo. El resultado para un cierto λ es un conjunto de cotas mínimas y máximas asociadas a cada reacción.

A.2.3 Análisis de Acoplamiento de Flujos

El *Flux Coupling Analysis* (FCA), al igual que el FBA y el FVA, es otra extensión del CBM (introducido en la sección previa), desarrollado para identificar acoplamientos estequiométricos entre las reacciones de una red (Burgard et al., 2004). La idea del acoplamiento entre las reacciones de una red se introdujo originalmente mediante el nombre “subconjuntos enzimáticos”, que son subconjuntos de reacciones cuyos flujos son linealmente dependientes en cualquier distribución de flujos estacionario (Pfeiffer et al., 1999). La detección de dichos subconjuntos puede realizarse mediante el análisis del espacio nulo de la matriz estequiométrica (Gagneur and Klamt, 2004; Pfeiffer et al., 1999). Calculando una base del espacio nulo de la matriz estequiométrica N y se representan los vectores de la base como columnas de la matriz K , dado que las filas de K representan a las reacciones, los subconjuntos de filas linealmente dependientes indican que la relación entre los flujos de dichas reacciones será la misma en cualquier estado estacionario. Como consecuencia tales reacciones se comportan como una unidad funcional ya que conociendo el flujo

4 Dado un modelo y una función objetivo, existe la posibilidad de que exista más de una solución óptima, que se corresponde a distribuciones de flujo diferentes, pero que permite alcanzar el mismo valor máximo en la tasa de crecimiento, y como consecuencia tales soluciones son denominadas óptimos alternativos (también se habla de que la solución es degenerada). La presentación de soluciones alternativas depende de la estructura del problema y la función objetivo.

de una se conoce el flujo de las demás, y el bloqueo de una reacción del subconjunto es suficiente para bloquear las otras reacciones. El ejemplo más sencillo de un subconjunto es una cadena lineal de reacciones donde el flujo estacionario será el mismo de cada reacción debido al balance de masas. Como consecuencia, si el sistema se encuentra en estado estacionario, bastará con conocer el valor de una velocidad para inferir el valor de las demás. Como se verá a continuación, con la introducción del FCA la idea de acoplamiento fue extendida de forma que se consideran otras clases de acoplamiento, *e.g.* el acoplamiento direccional. En este contexto, el concepto de subconjunto enzimático es denominado conjunto de acoplamiento completo (FCS, *full coupling set*) término empleado a lo largo de este trabajo para a las reacciones cuyos flujos son linealmente dependientes (Burgard et al., 2004).

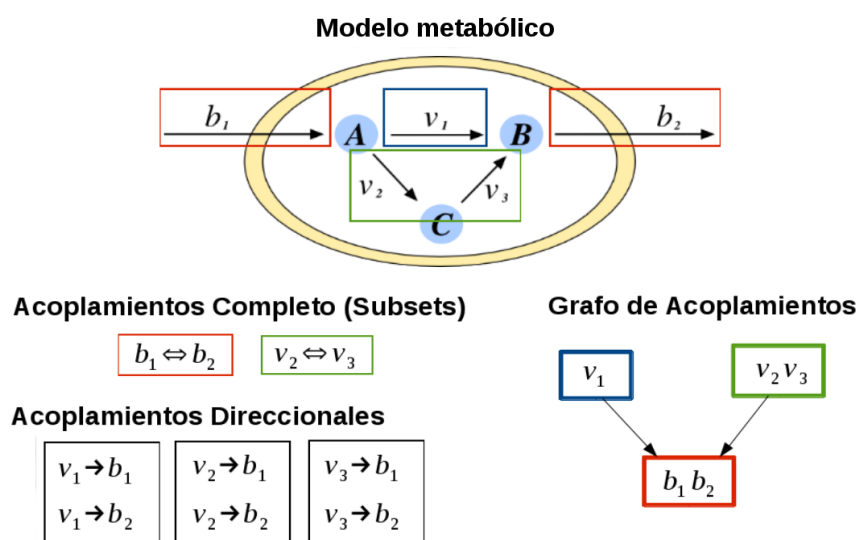


Figura A.3. Acoplamientos del flujo en una red metabólica de ejemplo. La figura representa una pequeña red metabólica compuesta por tres metabolitos y cinco reacciones. La figura incluye la representación de los conjuntos de acoplamiento completo, los acoplamientos direccionales y el grafo de acoplamiento que surge de dichas relaciones. Los rectángulos de colores indican los FCS.

El FCA permite establecer las relaciones de acoplamiento entre pares de reacciones de una red. El algoritmo propuesto en la publicación original, denominado *Flux Coupling Finder* (FCF) emplea programación lineal fraccional para identificar las razón máxima y mínima entre un par de flujos (*i.e.* $\max v_1/v_2$ y $\min v_1/v_2$), para cada posible par de reacciones (ver figura A.3). Una vez calculadas, la comparación de las relaciones entre flujos permite determinar la relación de acoplamiento para cualquier par de flujos v_1 y v_2 de la red. De acuerdo con las propiedades matemáticas del análisis, existen las siguientes clases de acoplamientos:

1. **Acoplamiento direccional** ($v_1 \rightarrow v_2$) se da da siempre que un flujo distinto a cero en v_1 necesariamente implica que el flujo de v_2 es distinto que cero, pero no recíprocamente.
2. **Acoplamiento parcial** ($v_1 \leftrightarrow v_2$) ocurre si un flujo diferente de cero en v_1 implica un flujo diferente de cero en v_2 y recíprocamente.
3. **Acoplamiento completo** ($v_1 \leftrightarrow v_2$) ocurre siempre que el flujo diferente de cero en v_1 implique un flujo diferente de cero en v_2 y *vice versa*, pero además que la relación entre ambos flujos sea constante.

Si un par de reacciones no pertenece a ninguna de las tres clases de acoplamiento antes mencionadas, se clasifican como no-acopladas. Las reacciones que están que se implican mutuamente se agrupan en conjuntos de reacciones acopladas o en general FCS. Además, como fue propuesto por Burgard et al., (2004) la información de los acoplamientos puede representar en forma de grafo lo que da como resultado la red de acoplamientos entre las reacciones del sistema. La figura A.3 representa una versión simplificada de la red introducida en la figura A.1. El FCA es una técnica computacional muy demandante ya que implica la resolución de un número de programas lineales del orden del cuadrado del número reacciones de la red. Sin embargo, desde la publicación del FCF, se han desarrollado al menos dos variantes de dicho algoritmo que permiten reducir, en forma sensible, el tiempo del cálculo. Estas variantes se corresponden a los algoritmo FFCA y el F2C2 (David et al., 2011; Larhlimi et al., 2012).

A.2.4 Predicción de medios mínimos de crecimiento

El FVA (ver apartado A.2.2) puede emplearse para identificar el conjunto de metabolitos necesarios para que un modelo prediga un flujo de producción de biomasa mayor que cero, *i.e.* para que el modelo prediga crecimiento. Esto se realiza imponiendo una cota inferior en el flujo de producción de biomasa, y luego se realiza el FVA sobre el conjunto de flujos de intercambio $j \in J_{EX}$. Aquellos flujos de intercambio cuya cota máxima sea menor que cero (*i.e.* que son consumidos por el sistema) implican que en cualquier modo de flujo que implique crecimiento el metabolito asociado al intercambio deberá ser consumido por el sistema. De forma alternativa, se puede iterar sobre los flujos de intercambio $j \in J_{EX}$ y en cada iteración se acota a cero el valor del flujo j y se maximiza la producción de biomasa; si el óptimo es cero, el sistema requiere del metabolito transportado para poder crecer. Si bien la aproximación descrita permite identificar los compuestos que necesariamente deben estar presentes en cualquier medio de cultivo para que el modelo tenga un crecimiento mayor que cero, el conjunto puede no ser suficiente. Por ejemplo, si el organismo puede emplear en forma alternativa dos fuentes de carbono, ninguna será detectada como necesaria ya que si se quita una aún puede recurrir a la otra. Sin embargo, si se quitan ambas en forma simultánea el modelo exhibirá un flujo de crecimiento igual a cero. Si bien el método basado en FVA puede ser extendido para pares y triadas de compuestos, la combinatoria puede hacer que el cálculo exhaustivo se vuelva inviable. En este contexto, el problema de identificar un medio de crecimiento mínimo puede plantearse como un problema de programación entera-mixta (*Mixed Integer Linear Programming*, MILP) donde se asocia una variable binaria a cada flujo de intercambio mediante de una restricción de forma que si la variable binaria toma el valor cero el valor del intercambio queda acotado a cero y en caso contrario el valor quedará limitado por las propias cotas impuestas sobre la variable (Burgard et al., 2001; Klitgord and Segrè, 2010b; Ponce de León et al., 2008; Suthers et al., 2009). De esta forma, si se impone una cota mínima el flujo de crecimiento y se minimiza la suma de las variables binarias, se obtendrá una distribución de flujos v que exhibe crecimiento y que importa del exterior un número de metabolitos mínimo.

El algoritmo requiere la incorporación de un conjunto de variables binarias $y_j \in \{0,1\}$, una por cada flujo de intercambio. Las restricciones $\alpha_j \cdot y_j \leq v_j$ se activan cuando la variable $y_j = 0$ y esto conduce a que la acota inferior del flujo de intercambio sea cero y el sistema no puede consumir

dicho metabolito⁵. Además, para garantizar que el medio mínimo predicho permite al sistema producir biomasa, se debe incluir un cota inferior sobre el flujo de producción de biomasa similar a la empleada en el FVA (*i.e.* $v_{biomasa} \geq \epsilon$). El parámetro ϵ debe fijarse en un valor lo suficientemente pequeño de forma que garantice que la solución sea mínima (desde un punto de vista biológico). En la práctica basta con fijar dicho parámetro en un valor del orden de 1×10^{-3} . Finalmente, la función objetivo a optimizar es la minimización del sumatorio sobre todas las variable binarias, que implica minimizar el número de metabolitos externos importados por sistema. El problema se puede formular de la siguiente manera:

$$\begin{aligned}
 \text{Min: } & \sum_{j \in J_{EX}} y_j \\
 \text{s.t. } & N \cdot \vec{v} = 0 \\
 & \alpha_j \leq v_j \leq \beta_j \quad \forall j \in J_{Int} \\
 & \alpha_j \cdot y_j \leq v_j \quad \forall j \in J_{Exchange} \\
 & v_{biomasa} \geq \lambda \cdot v_{biomasa}^* \\
 & \lambda \in [0-1] \\
 & y_j \in \{0,1\} \quad \forall j \in J_{Exchange}
 \end{aligned} \tag{MILP A.1}$$

Este procedimiento puede emplearse de forma iterativa en forma similar al algoritmo *k-shortest* (de Figueiredo et al., 2009). Además por construcción, los medios mínimos se corresponden con el concepto de conversiones elementales propuesto por Urbanczik y Wagner (2005), posteriormente generalizado mediante la introducción de los patrones elementales de flujo (Kaleta et al., 2009).

A.2.5 Experimentos de *knockout in-silico*

La robustez o fragilidad de una red metabólica con respecto a la pérdida o fallo en el funcionamiento de las distintas enzimas que conforman la red puede ser estimada empleando modelos estequiométricos reconstruidos a escala genómica y técnicas de optimización (Edwards and Palsson; Segrè et al., 2002; Shlomi et al., 2005). La simulación de un *knockout* es llevada adelante de la siguiente forma: en primer lugar se emplea la tabla de asociaciones gen-proteína-reacción para inferir qué reacciones se verían afectadas en la ausencia del gen. Si la pérdida del gen implica que una o más reacciones no pueden tener lugar, se fijan en cero los flujos respectivos. Finalmente, se evalúa la tasa de crecimiento del organismo empleando un método de optimización: si este valor no supera un cierto umbral, se considera que el *knockout* es letal, *i.e.* que el gen es esencial para el crecimiento del organismo; en caso contrario el gen se considera prescindible en dicha condición.

Existen diversos métodos de optimización para predecir el fenotipo que el organismo adoptará como respuesta a la perturbación génica. Los métodos más extendidos incluyen el análisis de balance de flujos (FBA) (Edwards and Palsson, 2000), la minimización del ajuste metabólico (MOMA, Minimization Of the Metabolic Adjustment) (Segrè et al., 2002) y la minimización de la regulación on/off (ROOM, Regulatory ON/OFF Minimization) (Shlomi et al., 2005). En el caso de FBA, simplemente se supone que después de la perturbación el organismo ajustará su metabolismo

5 Esto es debido a que dado que por definición los flujos de intercambio toman valores negativos cuando el metabolito correspondiente es consumido por el sistema (ver apartado 8.1.3)

a la distribución que maximiza la tasa de crecimiento (ver apartado A.2.1). De esta forma se simula el *knockout* y se maximiza la producción de biomasa y se examina el valor de dicho flujo, como se describe en el párrafo anterior.

Por otra parte, tanto MOMA como ROOM suponen que el organismo buscará mantener el estado metabólico más parecido al estado silvestre (*i.e.* el estado del sistema previo a la perturbación). Para realizar la predicción de dicho estado calculan la distribución de flujos que que minimiza una cierta “distancia” al estado de referencia (*i.e.* la solución óptima predicha por FBA, antes de introducir la perturbación). La diferencia entre estos dos métodos se basa en la definición de la distancia empleada: mientras que MOMA emplea una distancia euclidiana, ROOM define una métrica similar a lo que sería una distancia de Hamming, aunque no sea una distancia estrictamente hablando. Los experimentos de *knockout in-silico* realizado en este trabajo fueron calculados empleando MOMA y FBA. A continuación se describe la formulación de MOMA como problema de programación cuadrática estándar con restricciones lineales:

$$\begin{aligned} \text{Min } Z &= \frac{1}{2} \mathbf{v}^T \mathbf{Q} \mathbf{v} - \mathbf{w}^T \cdot \mathbf{v} \\ \text{s. t.} & \\ \mathbf{N} \cdot \vec{v} &= 0 \\ \beta_j &\leq v_j \leq \alpha_j \quad \forall j \in \mathbf{J} \end{aligned} \tag{QPA.1}$$

donde \mathbf{w}^T es la distribución de flujos de referencia correspondiente a la cepa salvaje, que por lo general se estima empleando FBA, y \mathbf{Q} es la matriz identidad. La implementación del método empleada en este trabajo forma parte de la biblioteca de modelado basado en restricciones para lenguaje Python COBRApy (ver apéndice C para más detalles).

Material Suplementario

Apéndice B

M.S. - Capítulo 3

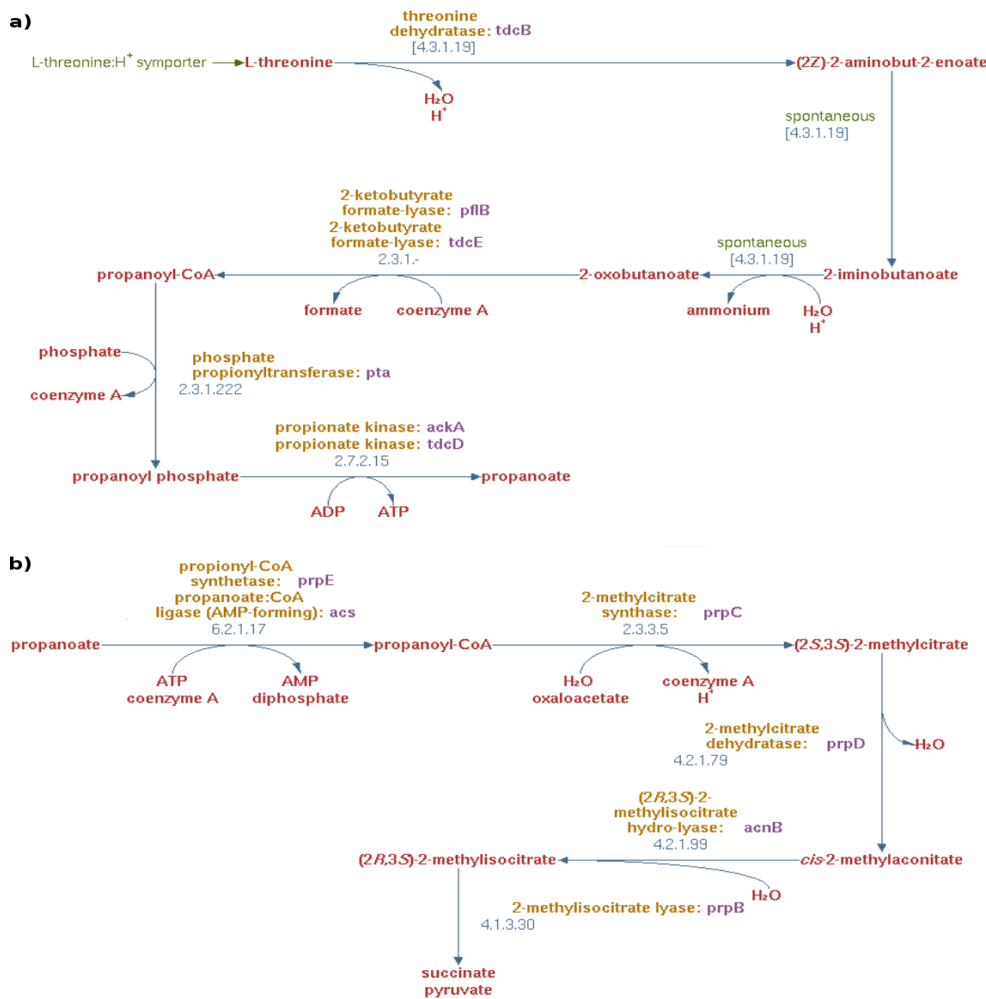


Figure S.3.1. Diagramas que representan vías metabólicas definidas en MetaCyc (Caspi et al., 2014). En a) se presenta la vía anaeróbica de degradación de la L-treonina; en b) se presenta el ciclo I del ácido 2-metilcitríco que se corresponde con la vía de degradación de propinato mayormente distribuida.

M.S. - Capítulo 4

Tabla S.4.1. Modelo a escala genómica de *Blattabacterium* Bge iMP240. La hoja de cálculo en formato XLS incluye la descripción del modelo iMP240. Incluye tres tablas: una que corresponde a los metabolitos, otra para las reacciones y una tercera para los flujos de intercambio. Descargar tabla:

https://github.com/migp11/Thesis_Supp_material/raw/master/cap4/Tabla_S_4_1.xls

Tabla S.4.2 Tabla comparativa entre los modelos iCG238 y iMP240. La hoja de cálculo en formato XLS describe las mayores diferencias entre las dos versiones del modelo. Contiene tres tablas: la primera incluye los genes que han sido agregados así como aquellos que se han quitado, como consecuencia de revisión; la segunda presenta las reacciones que se han quitado; y la tercera las nuevas reacciones que se han añadido al modelo. Descargar tabla:

https://github.com/migp11/Thesis_Supp_material/raw/master/cap4/Tabla_S_4_2.xls

M.S. - Capítulo 5

Tabla S.5.1. Modelo a escala genómica de *Serratia symbiotica* SCc iSCc236. La hoja de cálculo en formato XLS incluye la descripción del modelo iSCc236. Contiene tres tablas: una que corresponde a las reacciones; la segunda incluye los flujos de intercambio; y en la tercera está la información de los metabolitos. Descargar tabla:

https://github.com/migp11/Thesis_Supp_material/raw/master/cap5/Tabla_S51_iSCc226.xlsx

Tabla S.5.2. Modelo a escala genómica de *Buchnera aphidicola* BCc iBCc98. El fichero corresponde a la representación del modelo iBCc98 en formato XLS, e incluye las siguientes cuatro tablas: i) el listado de las reacciones; ii) el listado de los flujos de intercambio; iii) el listado de los metabolitos; iv) el listado de las reacciones y genes metabólicos que no se incluyeron en el modelo. Descargar tabla:

https://github.com/migp11/Thesis_Supp_material/raw/master/cap5/Tabla_S52_iBCc98.xlsx

Tabla S.5.3. Modelo del consorcio endosimbionte iBSCc. El fichero en formato XLS incluye tres tablas: una que corresponde a las reacciones; la segunda incluye los flujos de intercambio; en la tercera está la información de los metabolitos; y en la cuarta Las reacciones y genes metabólicos que no se incluyeron en el modelo. Descargar tabla:

https://github.com/migp11/Thesis_Supp_material/raw/master/cap5/Tabla_S53_iBSCc.xlsx

Tabla S.5.4. Datos experimentales de la cinética del crecimiento del áfido *S. graminum* y su endosimbionte *B. aphidicola* BSg. Los datos presentados en la tabla fueron obtenidos por Baumann y Baumman (1994). E incluyen las siguientes medidas: i) N° de días; N° áfidos contados; N° de muestras; peso medio de los áfidos (μg); concentración de proteína y ADN ($\mu\text{g}/\text{áfido}$); N° de copias de rRNA 16S *Buchnera* / pg de ADN de áfido; N° de copias de rRNA 16S (10^6) *Buchnera* / áfido. Descargar tabla:

https://github.com/migp11/Thesis_Supp_material/raw/master/cap5/Tabla_S54_Baumann.xlsx

Tabla S.5.5. Composición en aminoácidos del proteoma de los áfidos: *Aphis fabae* y de *A. pisum*. Los resultados para el caso de *A. fabae* corresponden a los resultados de los experimentos realizados por (Douglas et al., 2001), mientras que los datos de *A. pisum* corresponden a (Russell et al., 2014). Descargar tabla:

https://github.com/migp11/Thesis_Supp_material/raw/master/cap5/Tabla_S55_aa_aphid.xlsx

Tabla S.5.6. Resultados de las simulaciones con *iBSCc* sobre la organización óptima de la complementación en la ruta del triptófano. La tabla incluye los patrones de flujo correspondientes a las soluciones óptimas para los distintos diseños de la complementación para el caso de la optimización de la meta biomasa y para el caso de la optimización del triptófano. Además para ambos casos se presentan los resultados de la simulación relajando y sin relajar la restricción sobre los protones. Descargar tabla:

https://github.com/migp11/Thesis_Supp_material/raw/master/cap5/Tabla_S56_trp_exp.xlsx

Tabla S.5.7. Parámetros fisicoquímicos y estimadores de permeabilidad de la membrana celular a los distintos metabolitos del modelo de consorcio *iBSCc*. La tabla incluye identificadores de los compuestos para distintas bases de datos así como el corriente SMILES. Los parámetros y estimadores incluidos están descritos en el apartado 4.2.5. Descargar tabla:

https://github.com/migp11/Thesis_Supp_material/raw/master/cap5/Tabla_S57_cheminfo.xlsx

M.S. - Capítulo 6

Figura S.6.1. Curación de los módulos inconexos de MM130.0. El diagrama representa el conjunto de módulos inconexos detectados en MM130.0, junto con los artefactos detectados y el conjunto de reacciones añadidas durante el proceso de curación. Los nodos blancos corresponden a reacciones bloqueadas que fueron resueltas; los nodos verdes corresponden a las reacciones añadidas; y los nodos rojos indican las reacciones que fueron eliminadas del metamodelo (meta-bloqueadas). Descargar figura:

https://github.com/migp11/Thesis_Supp_material/raw/master/cap6/Figura_S_6_1.svg

Tabla S.6.1. Descripción del conjunto de datos y los resultados principales del análisis de consistencia. La tabla contiene tres secciones. En A) se describen los 130 modelos y los resultados del análisis de consistencia. En B) se describe el mapeo entre las versiones empleadas en este trabajo y las empleadas en la publicación original. En C) se describen los 36 modelos adicionales empleado para extender el metamodelo. Descargar tabla:

https://github.com/migp11/Thesis_Supp_material/raw/master/cap6/Tabla_S_6_1.xlsx

Tabla S.6.2. Clasificación de las reacciones de acuerdo a la frecuencia con que se encontraron bloqueadas. La tabla contiene cuatro secciones. En A) se lista el conjunto de 2195 reacciones que están incluidas en al menos uno de los 130 modelos. Cada fila corresponde a una reacciones con sus respectivos atributos. Las secciones B), C), y D) representan las reacciones (agrupadas por subsistema) clasificadas como “nunca bloqueadas”, “bloqueadas en algunos casos” y “siempre bloqueadas”, respectivamente. Descargar tabla:

https://github.com/migp11/Thesis_Supp_material/raw/master/cap6/Tabla_S_6_2.xlsx

Tabla S.6.3. Descripción del metamodelo MM130 y los resultados de la revisión. La tabla contiene cinco secciones. En A) se incluye el conjunto completo de reacciones y su descripción. Dos columnas fueron añadidas para indicar si la reacciones fue añadida o removida durante la curación. La última columna indica el ID del módulo inconexo que contiene la reacción. En forma similar, las secciones B) y C) describen el conjunto de flujos de intercambio y metabolitos, respectivamente. En D) se listan las 157 ecuaciones de biomasa junto a los modelos que tienen dicha formulación. En E) se listan los nuevos compuesto detectados como componentes de la biomasa. En F) y G) se describen las 226 reacciones y 113 metabolitos, respectivamente, identificadas en los 36 GEMM empleados para extender el metamodelo. Descargar tabla:

https://github.com/migp11/Thesis_Supp_material/raw/master/cap6/Tabla_S_6_3.xlsx

Tabla S.6.4. Resultados y análisis comparativo de los 130 y 36 GEMMs curados empleando de forma alternativa MM130.0, MM1301 y MM166.1. Los resultado del *gap-filling* realizado empleando alternativamente MM130.0 y MM130.1 se encuentran en las secciones A) y B), respectivamente. En la sección C) se muestran los resultados de la comparación entre cada para de modelos curados empleando alternativamente MM130.0 y MM130.1. En la sección D) y E) se presentan los resultados del *gap-filling*, empleando MM166.1, sobre los 130 modelos iniciales y los 36 nuevos, respectivamente. Descargar tabla:

https://github.com/migp11/Thesis_Supp_material/raw/master/cap6/Tabla_S_6_4.xlsx

Fichero S.6.1. Metamodelos en formato SBML. El fichero comprimido en formato ZIP incluye las dos versiones del metamodelo curado, MM130.1 y MM166.1, en formato SBML. Descargar fichero:

https://github.com/migp11/Thesis_Supp_material/raw/master/cap6/Fichero_S_6_1.zip

Herramientas Computacionales

Apéndice C

Herramientas de software para CBM

Los métodos originales basados en análisis estequiométrico empleados para analizar los modelos fueron desarrollados sobre COBRApy (<https://github.com/opencobra/cobrapy>) (Ebrahim et al., 2013), una biblioteca desarrollada en Python que incluye clases y funcionalidades desarrolladas para el análisis de modelos metabólicos reconstruidos a escala genómica. En el caso de métodos más estándar, como el FBA, FVA y MOMA, se emplearon las implementaciones incluida en COBRApy. El método utilizados en el capítulo 2 (e.g. *fastcore*, *oneprune*), fueron re-implementados en lenguaje. El código desarrollado se encuentra disponible en la siguiente URL:

<https://github.com/migp11/consistency-analysis/>.

Los programas incluidos permiten realizar el análisis de consistencia de un modelo, calculando las reacciones bloqueadas, los metabolitos *gap* y el conjunto de módulos inconexos. El FCA fue realizado empleando la implementación denominada FFCA disponible para el Matlab (David et al., 2011).

Otras herramientas de software

Los problemas de optimización tanto en el ámbito de la programación lineal como en la programación entera mixta fueron resueltos empleando Gurobi Solver (Gurobi Optimization, 2012). Para los algoritmos basados en grafos se empleó el módulo de Python denominado NetworkX (Hagberg et al., 2008). Las figuras que incluyen grafos fueron dibujadas empleando yEd Graph Editor (yWorks, 2013). y las demás gráficos utilizando la biblioteca de Python Matplotlib (Hunter, 2007). Los problemas de optimización no lineal planteados en la apartado 4.4.4 del capítulo 4 fueron implementados y resueltos empleando la función *fmincon* de Matlab 2015a para Linux. Debido a la no-linealidad del problema, para cada conjunto de parámetros, la minimización de la función objetivo se realizó múltiples veces, empleando “semillas” aleatorias para garantizar que la solución identificada sea óptima.

Apéndice D

Listado de publicaciones

Esta Tesis ha dado lugar a las siguientes publicaciones:

Ponce-de-Leon, M., Montero, F., Peretó, J., 2013. *Solving gap metabolites and blocked reactions in genome-scale models: application to the metabolic network of *Blattabacterium cuenoti**. BMC Syst. Biol. 7, 114. doi:10.1186/1752-0509-7-114

Ponce-de-León, M., Calle-Espinosa, J., Pereto, J., Montero, F., 2015. *Consistency Analysis of Genome-Scale Models of Bacterial Metabolism: A Metamodel Approach*. PLoS One 10, e0143626. doi:10.1371/journal.pone.0143626

Calle-Espinosa, J., **Ponce-de-Leon, M.**, Santos-Garcia, D., Silva, F.J., Montero, F., Peretó, J., 2016. *Nature lessons: The whitefly bacterial endosymbiont is a minimal amino acid factory with unusual energetics*. J. Theor. Biol. 407, 303–317. doi:10.1016/j.jtbi.2016.07.024

*Mori, M., ***Ponce-de-León, M.**, Peretó, J., Montero, F., 2016. *Metabolic Complementation in Bacterial Communities: Necessary Conditions and Optimality*. Front. Microbiol. 7, 1553. doi:10.3389/fmicb.2016.01553

Ponce-de-León, M., Tamarit, D., Calle-Espinosa, J., Pereto, J., Latorre A., Montero, F., 2016. *Metabolic reconstruction and anlysis of an endosymbiotic consortiums: the case of the cedar aphid* (en preparación)

(*) Contribución equivalente

Publicaciones del autor no incluidas en esta Tesis

- *Greif, G., ***Ponce de Leon, M.**, Lamolle, G., Rodriguez, M., Piñeyro, D., Tavares-Marques, L.M., Reyna-Bello, A., Robello, C., Alvarez Valin, F., 2013. *Transcriptome analysis of the bloodstream stage from the parasite Trypanosoma vivax*. BMC Genomics 14, 149. doi:10.1186/1471-2164-14-149
- Ponce de Leon, M.**, de Miranda, A., Alvarez-Valin, F., Carels, N. 2014. *The Purine Bias of Coding Sequences is Determined by Physicochemical Constraints on Proteins*. Bioinform. Biol. Insights 93. doi:10.4137/BBI.S13161
- Carels, N., Carels, N., **Ponce de Leon, M.**, 2015. *An Interpretation of the Ancestral Codon from Miller's Amino Acids and Nucleotide Correlations in Modern Coding Sequences*. Bioinform. Biol. Insights 37. doi:10.4137/BBI.S24021

(*) Contribución equivalente

Bibliografía

- Acuña, V., Chierichetti, F., Lacroix, V., Marchetti-Spaccamela, A., Sagot, M.-F., Stougie, L., 2009. Modes and cuts in metabolic networks: complexity and algorithms. *Biosystems*. 95, 51–60.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–10.
- Anantharaman, V., Aravind, L., 2003. Application of comparative genomics in the identification and analysis of novel families of membrane-associated receptors in bacteria. *BMC Genomics* 4, 34.
- Arita, M., 2012. From metabolic reactions to networks and pathways. *Methods Mol. Biol.* 804, 93–106.
- Arita, M., 2004. The metabolic world of *Escherichia coli* is not small. *Proc. Natl. Acad. Sci. U. S. A.* 101, 1543–7.
- Barabási, A.-L., Oltvai, Z.N., 2004. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–13.
- Barton, M.D., Delneri, D., Oliver, S.G., Rattray, M., Bergman, C.M., 2010. Evolutionary Systems Biology of Amino Acid Biosynthetic Cost in Yeast. *PLoS One* 5, e11935.
- Barve, A., Hosseini, S.-R., Martin, O.C., Wagner, A., 2014. Historical contingency and the gradual evolution of metabolic properties in central carbon and genome-scale metabolisms. *BMC Syst. Biol.* 8, 48.
- Barve, A., Wagner, A., 2013. A latent capacity for evolutionary innovation through exaptation in metabolic systems. *Nature* 500, 203–206.
- Baumann, L., Baumann, P., 1994. Growth Kinetics of the Endosymbiont *Buchnera aphidicola* in the Aphid *Schizaphis graminum*. *Appl. Environ. Microbiol.* 60, 3440–3.
- Baumann, L., Baumann, P., Moran, N.A., Sandström, J., Thao, M.L., 1999. Genetic Characterization of Plasmids Containing Genes Encoding Enzymes of Leucine Biosynthesis in Endosymbionts (*Buchnera*) of Aphids. *J. Mol. Evol.* 48, 77–85.
- Baumann, P., Baumann, L., Clark, M.A., 1996. Levels of *Buchnera aphidicola* chaperonin GroEL during growth of the aphid *Schizaphis graminum*. *Curr. Microbiol.* 32, 279–285.
- Baumann, P., Moran, N.A., Baumann, L., 2006. *The Prokaryotes*. Springer New York, New York, NY.
- Beard, D.A., Babson, E., Curtis, E., Qian, H., 2004. Thermodynamic constraints for biochemical networks. *J. Theor. Biol.* 228, 327–33.
- Beard, D.A., Liang, S., Qian, H., 2002. Energy balance for analysis of complex metabolic networks. *Biophys. J.* 83, 79–86.
- Belda, E., Silva, F.J., Peretó, J., Moya, A., 2012. Metabolic networks of *Sodalis glossinidius*: a systems biology approach to reductive evolution. *PLoS One* 7, e30652.
- Bellés, X., Martín, D., Piulachs, M.-D., 2005. The mevalonate pathway and the synthesis of juvenile hormone in insects. *Annu. Rev. Entomol.* 50, 181–99.
- Brock, M., Maerker, C., Sch??tz, A., V??lker, U., Buckel, W., 2002. Oxidation of propionate to pyruvate in *Escherichia coli*: Involvement of methylcitrate dehydratase and aconitase. *Eur. J. Biochem.* 269, 6184–6194.

- Brohée, S., Faust, K., Lima-Mendez, G., Vanderstocken, G., van Helden, J., 2008. Network Analysis Tools: from biological networks to clusters and pathways. *Nat. Protoc.* 3, 1616–29.
- Brooks, J.P., Burns, W.P., Fong, S.S., Gowen, C.M., Roberts, S.B., 2012. Gap detection for genome-scale constraint-based models. *Adv. Bioinformatics* 2012, 323472.
- Burgard, A.P., Nikolaev, E. V, Schilling, C.H., Maranas, C.D., 2004. Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res.* 14, 301–12.
- Burgard, A.P., Vaidyaraman, S., Maranas, C.D., 2001. Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments. *Biotechnol. Prog.* 17, 791–7.
- Burke, G.R., Moran, N.A., 2011. Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. *Genome Biol. Evol.* 3, 195–208.
- Caligiuri, M.G., Bauerle, R., 1991. Identification of amino acid residues involved in feedback regulation of the anthranilate synthase complex from *Salmonella typhimurium*. Evidence for an amino-terminal regulatory site. *J. Biol. Chem.* 266, 8328–35.
- Calle-Espinosa, J., Ponce-de-Leon, M., Santos-Garcia, D., Silva, F.J., Montero, F., Peretó, J., 2016. Nature lessons: The whitefly bacterial endosymbiont is a minimal amino acid factory with unusual energetics. *J. Theor. Biol.* 407, 303–317.
- Campbell, M.A., Van Leuven, J.T., Meister, R.C., Carey, K.M., Simon, C., McCutcheon, J.P., 2015. Genome expansion via lineage splitting and genome reduction in the cicada endosymbiont *Hodgkinia*. *Proc. Natl. Acad. Sci.* 112, 10192–10199.
- Caspi, R., Altman, T., Billington, ... P., Karp, P.D., 2014. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* 42, D459-71.
- Chandrasekaran, S., Price, N.D., 2010. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U. S. A.* 107, 17845–50.
- Charles, H., Balmand, S., Lamelas, A., Cottret, L., Pérez-Brocal, V., Burdin, B., Latorre, A., Febvay, G., Colella, S., Calevro, F., Rahbé, Y., 2011. A genomic reappraisal of symbiotic function in the aphid/*Buchnera* symbiosis: reduced transporter sets and variable membrane organisations. *PLoS One* 6, e29096.
- Cheng, T., Zhao, Y., Li, X., Lin, F., Xu, Y., Zhang, X., Li, Y., Wang, R., Lai, L., 2007. Computation of octanol-water partition coefficients by guiding an additive model with knowledge. *J. Chem. Inf. Model.* 47, 2140–2148.
- Chistoserdova, L., Kalyuzhnaya, M.G., Lidstrom, M.E., 2009. The expanding world of methylotrophic metabolism. *Annu. Rev. Microbiol.* 63, 477–99.
- Chiu, H.-C., Levy, R., Borenstein, E., 2014. Emergent biosynthetic capacity in simple microbial communities. *PLoS Comput. Biol.* 10, e1003695.
- Cohen, O., Ashkenazy, H., Burstein, D., Pupko, T., 2012. Uncovering the co-evolutionary network among prokaryotic genes. *Bioinformatics* 28, i389–i394.
- Consortium, T.I.A.G., 2010. Genome Sequence of the Pea Aphid *Acyrtosiphon pisum*. *PLoS Biol.* 8, e1000313.
- Cornish-Bowden, A., 2013. *Fundamentals of Enzyme Kinetics*. John Wiley & Sons.
- Crane, B.R., Siegel, L.M., Getzoff, E.D., 1997. Structures of the siroheme- and Fe4S4-containing active center of sulfite reductase in different states of oxidation: heme activation via reduction-gated exogenous

- ligand exchange. *Biochemistry* 36, 12101–19.
- Crawford, I.P., 1989. Evolution of a biosynthetic pathway: the tryptophan paradigm. *Annu. Rev. Microbiol.* 43, 567–600.
- Croes, D., Couche, F., Wodak, S.J., van Helden, J., 2006. Inferring meaningful pathways in weighted metabolic networks. *J. Mol. Biol.* 356, 222–36.
- Csete, M., Doyle, J., 2004. Bow ties, metabolism and disease. *Trends Biotechnol.* 22, 446–50.
- Dale, C., Moran, N. a, 2006. Molecular interactions between bacterial symbionts and their hosts. *Cell* 126, 453–65.
- Dale, J.M., Popescu, L., Karp, P.D., 2010. Machine learning methods for metabolic pathway prediction. *BMC Bioinformatics* 11, 15.
- David, L., Marashi, S.-A., Larhlimi, A., Mieth, B., Bockmayr, A., 2011. FFCA: a feasibility-based method for flux coupling analysis of metabolic networks. *BMC Bioinformatics* 12, 236.
- de Figueiredo, L.F., Podhorski, A., Rubio, A., Kaleta, C., Beasley, J.E., Schuster, S., Planes, F.J., 2009. Computing the shortest elementary flux modes in genome-scale metabolic networks. *Bioinformatics* 25, 3158–65.
- Dekel, E., Alon, U., 2005. Optimality and evolutionary tuning of the expression level of a protein. *Nature* 436, 588–92.
- Douglas, A.E., 1992. Requirement of pea aphids (*Acyrtosiphon pisum*) for their symbiotic bacteria. *Entomol. Exp. Appl.* 65, 195–198.
- Douglas, A.E., Darby, A.C., Birkle, L.M., Walters, K.F.A., 2002. The ecological significance of symbiotic micro-organisms in animals: perspectives from the microbiota of aphids, in: *Genes in the Environment*. Blackwell Scientific, Oxford, pp. 306–325.
- Douglas, A.E., Minto, L.B., Wilkinson, T.L., 2001. Quantifying nutrient production by the microbial symbionts in an aphid. *J. Exp. Biol.* 204, 349–58.
- Dreyfuss, J.M., Zucker, J.D., Hood, H.M., Ocasio, L.R., Sachs, M.S., Galagan, J.E., 2013. Reconstruction and validation of a genome-scale metabolic model for the filamentous fungus *Neurospora crassa* using FARM. *PLoS Comput. Biol.* 9, e1003126.
- Dunn, M.F., 2012. Allosteric regulation of substrate channeling and catalysis in the tryptophan synthase holoenzyme complex. *Arch. Biochem. Biophys.* 519, 154–66.
- Ebrahim, A., Lerman, J.A., Palsson, B.O., Hyduke, D.R., Ebrahim A, Lerman JA, Palsson BO, H.D., 2013. COBRAPy: CONstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* 7, 74.
- Edwards, J.S., Ibarra, R.U., Palsson, B.O., 2001. In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat. Biotechnol.* 19, 125–30.
- Edwards, J.S., Palsson, B.O., 2000. The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. U. S. A.* 97, 5528–33.
- Edwards, J.S., Palsson, B.O., Robustness analysis of the *Escherichia coli* metabolic network. *Biotechnol. Prog.* 16, 927–39.
- Evans, G.A., 2000. Designer science and the “omic” revolution. *Nat. Biotechnol.* 18, 127.
- Famili, I., Palsson, B.O., 2003. The convex basis of the left null space of the stoichiometric matrix leads to the definition of metabolically meaningful pools. *Biophys. J.* 85, 16–26.

- Faust, K., Raes, J., 2012. Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* 10, 538–50.
- Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V., Palsson, B.Ø., 2007a. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* 3, 121.
- Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V., Palsson, B.Ø., 2007b. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 {MG1655} that accounts for 1260 {ORFs} and thermodynamic information. *Mol. Syst. Biol.* 3, 121.
- Feist, A.M., Herrgård, M.J., Thiele, I., Reed, J.L., Palsson, B.Ø., 2009. Reconstruction of biochemical networks in microorganisms. *Nat. Rev. Microbiol.* 7, 129–43.
- Feist, A.M., Palsson, B.O., 2010. The biomass objective function. *Curr. Opin. Microbiol.* 13, 344–9.
- Feist, A.M., Palsson, B.Ø., 2008. The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nat. Biotechnol.* 26, 659–67.
- Ferry, J.G., 2010. How to Make a Living by Exhaling Methane. *Annu. Rev. Microbiol.* 64, 453–473.
- Flamholz, A., Noor, E., Bar-Even, A., Liebermeister, W., Milo, R., 2013. Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc. Natl. Acad. Sci. U. S. A.* 110, 10039–44.
- Flamholz, A., Noor, E., Bar-Even, A., Milo, R., 2012. eQuilibrator--the biochemical thermodynamics calculator. *Nucleic Acids Res.* 40, D770-5.
- Fleming, R.M.T., Maes, C.M., Saunders, M.A., Ye, Y., Palsson, B.Ø., 2012. A variational principle for computing nonequilibrium fluxes and potentials in genome-scale biochemical networks. *J. Theor. Biol.* 292, 71–7.
- Fleming, R.M.T., Thiele, I., Provan, G., Nasheuer, H.P., 2010. Integrated stoichiometric, thermodynamic and kinetic modelling of steady state metabolism. *J. Theor. Biol.* 264, 683–92.
- Francke, C., Siezen, R.J., Teusink, B., 2005. Reconstructing the metabolic network of a bacterium from its genome. *Trends Microbiol.* 13, 550–8.
- Freilich, S., Zarecki, R., Eilam, O., Segal, E.S., Henry, C.S., Kupiec, M., Gophna, U., Sharan, R., Ruppin, E., 2011. Competitive and cooperative metabolic interactions in bacterial communities. *Nat. Commun.* 2, 589.
- Frishman, D., 2007. Protein annotation at genomic scale: the current status. *Chem. Rev.* 107, 3448–66.
- Gagneur, J., Klamt, S., 2004. Computation of elementary modes: a unifying framework and the new binary approach. *BMC Bioinformatics* 5, 175.
- Galperin, M.Y., Koonin, E. V., 2004. “Conserved hypothetical” proteins: prioritization of targets for experimental study. *Nucleic Acids Res.* 32, 5452–63.
- Germerodt, S., Bohl, K., Lück, A., Pande, S., Schröter, A., Kaleta, C., Schuster, S., Kost, C., 2016. Pervasive Selection for Cooperative Cross-Feeding in Bacterial Communities. *PLOS Comput. Biol.* 12, e1004986.
- Gil, R., Peretó, J., 2015. Small genomes and the difficulty to define minimal translation and metabolic machineries. *Front. Ecol. Evol.* 3, 123.
- Gilks, W.R., Audit, B., De Angelis, D., Tsoka, S., Ouzounis, C.A., 2002. Modeling the percolation of annotation errors in a database of protein sequences. *Bioinformatics* 18, 1641–1649.
- Goelzer, a, Fromion, V., 2011. Bacterial growth rate reflects a bottleneck in resource allocation. *Biochim. Biophys. Acta* 1810, 978–88.

- González-Domenech, C.M., Belda, E., Patiño-Navarrete, R., Moya, A., Peretó, J., Latorre, A., 2012. Metabolic stasis in an ancient symbiosis: genome-scale metabolic networks from two *Blattabacterium cuenoti* strains, primary endosymbionts of cockroaches. *BMC Microbiol.* 12 Suppl 1, S5.
- Gosalbes, M.J., Lamelas, A., Moya, A., Latorre, A., 2008. The striking case of tryptophan provision in the cedar aphid *Cinara cedri*. *J. Bacteriol.* 190, 6026–9.
- Graber, J.R., Breznak, J.A., 2005. Folate cross-feeding supports symbiotic homoacetogenic spirochetes. *Appl. Environ. Microbiol.* 71, 1883–9.
- Green, M.L., Karp, P.D., 2004. A Bayesian method for identifying missing enzymes in predicted metabolic pathway databases. *BMC Bioinformatics* 5, 76.
- Großkopf, T., Soyer, O.S., 2016. Microbial diversity arising from thermodynamic constraints. *ISME J.* 1–9.
- Großkopf, T., Soyer, O.S., 2014. Synthetic microbial communities. *Curr. Opin. Microbiol.* 18, 72–77.
- Gudmundsson, S., Thiele, I., 2010. Computationally efficient flux variability analysis. *BMC Bioinformatics* 11, 489.
- Gurobi Optimization, I., 2012. Houston, Texas: Gurobi Optimization, Inc.
- Hagberg, A.A., Schult, D.A., Swart, P.J., 2008. Exploring network structure, dynamics, and function using NetworkX, in: *Proceedings of the 7th Python in Science Conference (SciPy2008)*. Pasadena, CA USA, CA USA, pp. 11–15.
- Hansen, A.K., Moran, N.A., 2011. Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. *Proc. Natl. Acad. Sci. U. S. A.* 108, 2849–54.
- Hanson, A.D., Pribat, A., Waller, J.C., de Crecy-Lagard, V., de Crécy-Lagard, V., 2010. “Unknown” proteins and “orphan” enzymes: the missing half of the engineering parts list--and how to find it. *Biochem. J.* 425, 1–11.
- Hardin, G., 1960. The competitive exclusion principle. *Science* 131, 1292–7.
- Hartwell, L.H., Hopfield, J.J., Leibler, S., Murray, A.W., 1999. From molecular to modular cell biology. *Nature* 402, C47-52.
- Heinrich, R., Schuster, S., 1998. The modelling of metabolic systems. Structure, control and optimality. *Biosystems.* 47, 61–77.
- Heinrich, R., Schuster, S., 1996. *The Regulation of Cellular Systems*. Springer.
- Henry, C.S., Broadbelt, L.J., Hatzimanikatis, V., 2007. Thermodynamics-based metabolic flux analysis. *Biophys. J.* 92, 1792–1805.
- Henry, C.S., DeJongh, M., Best, A. a, Frybarger, P.M., Linsay, B., Stevens, R.L., 2010. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat. Biotechnol.* 28, 977–82.
- Horswill, A.R., Dudding, A.R., Escalante-Semerena, J.C., 2001. Studies of propionate toxicity in *Salmonella enterica* identify 2-methylcitrate as a potent inhibitor of cell growth. *J. Biol. Chem.* 276, 19094–101.
- Hosoda, K., Suzuki, S., Yamauchi, Y., Shiroguchi, Y., Kashiwagi, A., Ono, N., Mori, K., Yomo, T., 2011. Cooperative adaptation to establishment of a synthetic bacterial mutualism. *PLoS One* 6.
- Hosoda, K., Yomo, T., 2011. Designing symbiosis. *Bioeng. Bugs* 2, 338–341.
- Hosseini, S.-R., Barve, A., Wagner, A., 2015. Exhaustive Analysis of a Genotype Space Comprising 10(15) Central Carbon Metabolisms Reveals an Organization Conducive to Metabolic Innovation. *PLoS Comput.*

- Biol. 11, e1004329.
- Humphreys, N.J., Douglas, A.E., 1997. Partitioning of symbiotic bacteria between generations of an insect: A quantitative study of a *Buchnera* sp. in the pea aphid (*Acyrtosiphon pisum*) reared at different temperatures. *Appl. Environ. Microbiol.* 63, 3294–3296.
- Hunter, J.D., 2007. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95.
- Husnik, F., Nikoh, N., Koga, R., Ross, L., Duncan, R.P., Fujie, M., Tanaka, M., Satoh, N., Bachtrog, D., Wilson, A.C.C., Von Dohlen, C.D., Fukatsu, T., McCutcheon, J.P., 2013. XHorizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell* 153, 1567–1578.
- Ish-Am, O., Kristensen, D.M., Ruppin, E., 2015. Evolutionary Conservation of Bacterial Essential Metabolic Genes across All Bacterial Culture Media. *PLoS One* 10, e0123785.
- Jankowski, M.D., Henry, C.S., Broadbelt, L.J., Hatzimanikatis, V., 2008. Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophys. J.* 95, 1487–99.
- Jensen, P. a, Lutz, K. a, Papin, J. a, 2011. TIGER: Toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks. *BMC Syst. Biol.* 5, 147.
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., Barabási, A.L., 2000. The large-scale organization of metabolic networks. *Nature* 407, 651–4.
- Jerby, L., Shlomi, T., Ruppin, E., 2010. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Mol. Syst. Biol.* 6, 401.
- Jiang, Z.-F.F., Xia, F., Johnson, K.W.,, Ghanim, M., 2012. Genome sequences of the primary endosymbiont “*Candidatus portiera aleyrodidarum*” in the whitefly *Bemisia tabaci* B and Q biotypes. *J. Bacteriol.*
- Jiménez, N., González-Candelas, F., Silva, F.J., 2000. Prephenate dehydratase from the aphid endosymbiont (*Buchnera*) displays changes in the regulatory domain that suggest its desensitization to inhibition by phenylalanine. *J. Bacteriol.* 182, 2967–9.
- Johansson, P., Hederstedt, L., 1999. Organization of genes for tetrapyrrole biosynthesis in gram--positive bacteria. *Microbiology* 145 (Pt 3, 529–38.
- Juan, D., Pazos, F., Valencia, A., 2008. High-confidence prediction of global interactomes based on genome-wide coevolutionary networks. *Proc. Natl. Acad. Sci. U. S. A.* 105, 934–9.
- Kacser, H., Burns, J.A., 1973. The control of flux. *Symp. Soc. Exp. Biol.* 27, 65–104.
- Kafri, M., Metzler-Raz, E., Jona, G., Barkai, N., 2015. The Cost of Protein Production. *Cell Rep.* 14, 22–31.
- Kaleta, C., de Figueiredo, L.F., Schuster, S., 2009. Can the whole be less than the sum of its parts? Pathway analysis in genome-scale metabolic networks using elementary flux patterns. *Genome Res.* 19, 1872–83.
- Kaleta, C., Schäuble, S., Rinas, U., Schuster, S., 2013. Metabolic costs of amino acid and protein production in *Escherichia coli*. *Biotechnol. J.* 8, 1105–14.
- Kanehisa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30.
- Karp, P.D., 1998. Metabolic databases. *Trends Biochem. Sci.* 23, 114–116.
- Karp, P.D., Keseler, I.M., Shearer, ..., C., Ingraham, J., 2007. Multidimensional annotation of the *Escherichia coli* K-12 genome. *Nucleic Acids Res.* 35, 7577–90.

- Karp, P.D., Latendresse, M., Paley, S.M., Ong, M.K.Q., Billington, R., Kothari, A., Weaver, D., Lee, T., Subhraveti, P., Spaulding, A., Fulcher, C., Keseler, I.M., Caspi, R., 2015. Pathway Tools version 19.0: Integrated Software for Pathway/Genome Informatics and Systems Biology. *Brief. Bioinform.* bbv079.
- Kerner, A., Park, J., Williams, A., Lin, X.N., 2012. A programmable *Escherichia coli* consortium via tunable symbiosis. *PLoS One* 7, e34032.
- Kim, S., Thiessen, P.A., Bolton, E.E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B.A., Wang, J., Yu, B., Zhang, J., Bryant, S.H., 2015. PubChem Substance and Compound databases. *Nucleic Acids Res.* 44, D1202-13.
- Kirschner, K., Lane, A.N., Strasser, A.W., 1991. Reciprocal communication between the lyase and synthase active sites of the tryptophan synthase henzyme complex. *Biochemistry* 30, 472–8.
- Klamt, S., Stelling, J., 2006. Stoichiometric and constraint-based modelling. *Syst. Model. Cell. Biol. From Concepts to Nuts Bolts* 73–96.
- Klein, C.C., Alves, J.M.P., Serrano, M.G., Buck, G.A., Vasconcelos, A.T.R., Sagot, M.-F., Teixeira, M.M.G., Camargo, E.P., Motta, M.C.M., 2013. Biosynthesis of vitamins and cofactors in bacterium-harbouring trypanosomatids depends on the symbiotic association as revealed by genomic analyses. *PLoS One* 8, e79786.
- Klitgord, N., Segrè, D., 2010a. The importance of compartmentalization in metabolic flux models: yeast as an ecosystem of organelles. *Genome Inform.* 22, 41–55.
- Klitgord, N., Segrè, D., 2010b. Environments that induce synthetic microbial ecosystems. *PLoS Comput. Biol.* 6, e1001002.
- Kumar, V.S., Dasika, M.S., Maranas, C.D., 2007. Optimization based automated curation of metabolic reconstructions. *BMC Bioinformatics* 8, 212.
- Kumar, V.S., Maranas, C.D., 2009. GrowMatch: an automated method for reconciling in silico/in vivo growth predictions. *PLoS Comput. Biol.* 5, e1000308.
- Kuzuyama, T., 2002. Mevalonate and nonmevalonate pathways for the biosynthesis of isoprene units. *Biosci. Biotechnol. Biochem.* 66, 1619–27.
- Kwak, J.-H., Hong, K.-W., Lee, S.-H., Hong, J.-H., Lee, S.-Y., 1999. Identification of Amino Acid Residues Involved in Feedback Inhibition of the Anthranilate Synthase in *Escherichia coli*. *BMB Rep.* 32, 20–24.
- Lai, C.Y., Baumann, L., Baumann, P., 1994. Amplification of trpEG: adaptation of *Buchnera aphidicola* to an endosymbiotic association with aphids. *Proc. Natl. Acad. Sci. U. S. A.* 91, 3819–23.
- Lamelas, A., Gosalbes, M.J., Manzano-Marín, A., Peretó, J., Moya, A., Latorre, A., 2011a. *Serratia symbiotica* from the aphid *Cinara cedri*: a missing link from facultative to obligate insect endosymbiont. *PLoS Genet.* 7, e1002357.
- Lamelas, A., Gosalbes, M.J., Moya, A., Latorre, A., 2011b. New clues about the evolutionary history of metabolic losses in bacterial endosymbionts, provided by the genome of *Buchnera aphidicola* from the aphid *Cinara tujafilina*. *Appl. Environ. Microbiol.* 77, 4446–54.
- Larhlimi, A., David, L., Selbig, J., Bockmayr, A., 2012. F2C2: a fast tool for the computation of flux coupling in genome-scale metabolic networks. *BMC Bioinformatics* 13, 57.
- Latendresse, M., 2014. Efficiently gap-filling reaction networks. *BMC Bioinformatics* 15, 225.
- Latendresse, M., Kruppenacker, M., Trupp, M., Karp, P.D., 2012. Construction and completion of flux balance models from pathway databases. *Bioinformatics* 28, 388–96.

- Lee, N.-R., Lakshmanan, M., Aggarwal, S., Song, J.-W., Karimi, I.A., Lee, D.-Y., Park, J.-B., 2014. Genome-scale metabolic network reconstruction and in silico flux analysis of the thermophilic bacterium *Thermus thermophilus* HB27. *Microb. Cell Fact.* 13, 61.
- Lee, P., Hsu, A., Ha, H., Clarke, C., 1997. A C-methyltransferase involved in both ubiquinone and menaquinone biosynthesis: isolation and identification of the *Escherichia coli* ubiE gene. *J. Bacteriol.* 179, 1748–1754.
- Lee, S., Phalakornkule, C., Domach, M.M., Grossmann, I.E., 2000. Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. *Comput. Chem. Eng.* 24, 711–716.
- Lewis, N.E., Hixson, K.K., Conrad, T.M., ..., Palsson, B.Ø., 2010. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol. Syst. Biol.* 6, 390.
- Lewis, N.E., Nagarajan, H., Palsson, B.O., 2012. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat. Rev. Microbiol.* 10, 291–305.
- Li, C., Donizelli, M., Rodriguez, N., Dharuri, H., Endler, L., Chelliah, V., Li, L., He, E., Henry, A., Stefan, M.I., Snoep, J.L., Hucka, M., Le Novère, N., Laibe, C., 2010. BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Syst. Biol.* 4, 92.
- Li, L., Stoeckert, C.J., Roos, D.S., 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–89.
- Lin, S., Cronan, J.E., 2011. Closing in on complete pathways of biotin biosynthesis. *Mol. Biosyst.* 7, 1811–21.
- Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J., 2001. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 46, 3–26.
- López-Sánchez, M.J., Neef, A., Peretó, J., Patiño-Navarrete, R., Pignatelli, M., Latorre, A., Moya, A., 2009. Evolutionary convergence and nitrogen metabolism in *Blattabacterium* strain Bge, primary endosymbiont of the cockroach *Blattella germanica*. *PLoS Genet.* 5, e1000721.
- Luginbuhl, G.H., Hofler, J.G., Decedue, C.J., Burns, R.O., 1974. Biodegradative L-threonine deaminase of *Salmonella typhimurium*. *J. Bacteriol.* 120, 559–61.
- M. Ullah, H. Schmidt, K.-H.C. and O.W., 2006. Deterministic modelling and stochastic simulation of biochemical pathways using MATLAB. *M. Syst. Biol. (Stevenage)*. 153, 53–60.
- Ma, H.-W., Zeng, A.-P., 2003. The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics* 19, 1423–30.
- Macdonald, S.J., Lin, G.G., Russell, C.W., Thomas, G.H., Douglas, A.E., 2012. The central role of the host cell in symbiotic nitrogen metabolism. *Proc. Biol. Sci.* 279, 2965–73.
- MacDonald, S.J., Thomas, G.H., Douglas, A.E., 2011. Genetic and metabolic determinants of nutritional phenotype in an insect-bacterial symbiosis. *Mol. Ecol.* 20, 2073–84.
- Mackie, A., Keseler, I.M., Nolan, L., Karp, P.D., Paulsen, I.T., 2013. Dead end metabolites--defining the known unknowns of the *E. coli* metabolic network. *PLoS One* 8, e75210.
- Mahadevan, R., Palsson, B.O., 2005. Properties of metabolic networks: structure versus function. *Biophys. J.* 88, L07-9.
- Mahadevan, R., Schilling, C.H., 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* 5, 264–76.

- Manzano-Marín, A., Latorre, A., 2014. Settling down: the genome of *Serratia symbiotica* from the aphid *Cinara tujafilina* zooms in on the process of accommodation to a cooperative intracellular life. *Genome Biol. Evol.* 6, 1683–1698.
- Manzano-Marín, A., Simon, J.-C., Latorre, A., 2016. Reinventing the Wheel and Making It Round Again: Evolutionary Convergence in Buchnera – *Serratia* Symbiotic Consortia between the Distantly Related Lachninae Aphids *Tuberolachnus salignus* and *Cinara cedri*. *Genome Biol. Evol.* 8, 1440–1458.
- Marashi, S.-A., Bockmayr, A., 2011. Flux coupling analysis of metabolic networks is sensitive to missing reactions. *Biosystems.* 103, 57–66.
- Marashi, S.-A., David, L., Bockmayr, A., 2012. On flux coupling analysis of metabolic subsystems. *J. Theor. Biol.* 302, 62–9.
- Martínez-Cano, D.J., Reyes-Prieto, M., Martínez-Romero, E., Partida-Martínez, L.P., Latorre, A., Moya, A., Delaye, L., 2015. Evolution of small prokaryotic genomes. *Front. Microbiol.* 5, 742.
- Martino, P. Di, Fursy, R., Bret, L., Sundararaju, B., Phillips, R.S., 2003. Indole can act as an extracellular signal to regulate biofilm formation of *Escherichia coli* and other indole-producing bacteria. *Can. J. Microbiol.* 49, 443–9.
- McCloskey, D., Palsson, B.Ø., Feist, A.M., 2013. Basic and applied uses of genome-scale metabolic network reconstructions of *Escherichia coli*. *Mol. Syst. Biol.* 9, 661.
- McCutcheon, J.P., Moran, N.A., 2007. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proc. Natl. Acad. Sci.* 104, 19392–19397.
- McCutcheon, J.P., von Dohlen, C.D., 2011. An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Curr. Biol.* 21, 1366–72.
- Mee, M.T., Collins, J.J., Church, G.M., Wang, H.H., 2014. Syntrophic exchange in synthetic microbial communities. *Proc. Natl. Acad. Sci. U. S. A.* 111, E2149–56.
- Meléndez-Hevia, E., De Paz-Lugo, P., Cornish-Bowden, A., Cárdenas, M.L., 2009. A weak link in metabolism: the metabolic capacity for glycine biosynthesis does not satisfy the need for collagen synthesis. *J. Biosci.* 34, 853–72.
- Melendez-Hevia, E., Paz-Lugo, P. De, 2008. Branch-point stoichiometry can generate weak links in metabolism: the case of glycine biosynthesis. *J. Biosci.* 33, 771–80.
- Michalkova, V., Benoit, J.B., Weiss, B.L., Attardo, G.M., Aksoy, S., 2014. Vitamin B6 generated by obligate symbionts is critical for maintaining proline homeostasis and fecundity in tsetse flies. *Appl. Environ. Microbiol.* 80, 5844–53.
- Miles, E.W., 2013. The tryptophan synthase $\alpha_2\beta_2$ complex: a model for substrate channeling, allosteric communication, and pyridoxal phosphate catalysis. *J. Biol. Chem.* 288, 10084–91.
- Mira, A., Moran, N.A., 2002. Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. *Microb. Ecol.* 44, 137–43.
- Monk, J., Nogales, J., Palsson, B.O., 2014. Optimizing genome-scale network reconstructions. *Nat. Biotechnol.* 32, 447–52.
- Montagud, A., Zelezniak, A., Navarro, E., de Córdoba, P.F., Urchueguía, J.F., Patil, K.R., 2011. Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium *Synechocystis* sp. PCC6803. *Biotechnol. J.* 6, 330–42.
- Moran, N.A., 1996. Accelerated Evolution and Muller's Ratchet in Endosymbiotic Bacteria. *Proc. Natl. Acad.*

- Sci. 93, 2873–2878.
- Moran, N.A., Bennett, G.M., 2014. The Tiniest Tiny Genomes. *Annu. Rev. Microbiol.* 68, 195–215.
- Mori, M., Hwa, T., Martin, O.C., Martino, A. De, Marinari, E., 2016. Constrained Allocation Flux Balance Analysis. *PLOS Comput. Biol.* 12, 371.
- Morowitz, H., Smith, E., 2007. Energy flow and the organization of life. *Complexity*.
- Morowitz, H.J., Kostelnik, J.D., Yang, J., Cody, G.D., 2000. The origin of intermediary metabolism. *Proc. Natl. Acad. Sci. U. S. A.* 97, 7704–8.
- Morris, J., Lenski, R.E., Zinser, E.R., Black, T., Hypothesis, Q., Loss, A.G., This, S., Feeds, R.S.S., Journal, A.S.M., 2012. The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss.
- Moya, A., Peretó, J., Gil, R., Latorre, A., 2008. Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat. Rev. Genet.* 9, 218–29.
- Müller, A.C., Bockmayr, A., 2013a. Fast thermodynamically constrained flux variability analysis. *Bioinformatics* 29, 903–9.
- Müller, A.C., Bockmayr, A., 2013b. Flux modules in metabolic networks. *J. Math. Biol.* 1–29.
- Nagrath, D., Avila-Elchiver, M., Berthiaume, F., Tilles, A.W., Messac, A., Yarmush, M.L., 2007. Integrated energy and flux balance based multiobjective framework for large-scale metabolic networks. *Ann. Biomed. Eng.* 35, 863–85.
- Neidhardt, F.C. (Frederick C., Curtiss, R., 1987. *Escherichia coli and Salmonella : cellular and molecular biology*. ASM Press, p. 2822.
- Niehaus, T.D., Thamm, A.M.K., de Crécy-Lagard, V., Hanson, A.D., 2015. Proteins of Unknown Biochemical Function: A Persistent Problem and a Roadmap to Help Overcome It. *Plant Physiol.* 169, 1436–1442.
- Nikoh, N., Hosokawa, T., Moriyama, M., Oshima, K., Hattori, M., Fukatsu, T., 2014. Evolutionary origin of insect-Wolbachia nutritional mutualism. *Proc. Natl. Acad. Sci.* 111, 10257–10262.
- Nikoh, N., McCutcheon, J.P., Kudo, T., Miyagishima, S.Y., Moran, N.A., Nakabachi, A., 2010. Bacterial genes in the aphid genome: Absence of functional gene transfer from Buchnera to its host. *PLoS Genet.* 6, e1000827.
- Nikoh, N., Nakabachi, A., 2009. Aphids acquired symbiotic genes via lateral gene transfer. *BMC Biol.* 7, 12.
- Nikolaev, E. V, Burgard, A.P., Maranas, C.D., 2005. Elucidation and structural analysis of conserved pools for genome-scale metabolic reconstructions. *Biophys. J.* 88, 37–49.
- Noor, E., Bar-Even, A., Flamholz, A., Reznik, E., Liebermeister, W., Milo, R., 2014. Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLoS Comput. Biol.* 10, e1003483.
- Noor, E., Eden, E., Milo, R., Alon, U., 2010. Central Carbon Metabolism as a Minimal Biochemical Walk between Precursors for Biomass and Energy. *Mol. Cell* 39, 809–820.
- Notebaart, R.A., Teusink, B., Siezen, R.J., Papp, B., 2008. Co-regulation of metabolic genes is better explained by flux coupling than by network distance. *PLoS Comput. Biol.* 4, e26.
- Notebaart, R.A., van Enkevort, F.H.J., Francke, C., Siezen, R.J., Teusink, B., 2006. Accelerating the reconstruction of genome-scale metabolic networks. *BMC Bioinformatics* 7, 296.
- Notredame, C., Higgins, D.G., Heringa, J., 2000. T-Coffee: A novel method for fast and accurate multiple

- sequence alignment. *J. Mol. Biol.* 302, 205–17.
- Oberhardt, M.A., Pálsson, B.Ø., Papin, J.A., 2009. Applications of genome-scale metabolic reconstructions. *Mol. Syst. Biol.* 5, 320.
- Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M., Pálsson, B.Ø., 2011. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Mol. Syst. Biol.* 7, 535.
- Orth, J.D., Pálsson, B.Ø., 2010. Systematizing the generation of missing metabolic knowledge. *Biotechnol. Bioeng.* 107, 403–412.
- Orth, J.D., Thiele, I., Pálsson, B.Ø., 2010. What is flux balance analysis? *Nat. Biotechnol.* 28, 245–8.
- Osterman, A., Overbeek, R., 2003. Missing genes in metabolic pathways: a comparative genomics approach. *Curr. Opin. Chem. Biol.* 7, 238–51.
- Overbeek, R., Begley, T., Butler, ..., O., Vonstein, V., 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* 33, 5691–5702.
- Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F., Stevens, R., 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 42, D206-14.
- Pah, A.R., Guimerà, R., Mustoe, A.M., Amaral, L.A.N., 2013. Use of a global metabolic network to curate organismal metabolic networks. *Sci. Rep.* 3, 1695.
- Pál, C., Papp, B., Lercher, M.J., 2005. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat. Genet.* 37, 1372–5.
- Pál, C., Papp, B., Lercher, M.J., Csermely, P., Oliver, S.G., Hurst, L.D., 2006. Chance and necessity in the evolution of minimal metabolic networks. *Nature* 440, 667–70.
- Pálsson, B.O., 2006. *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press.
- Pálsson, B.O., 1997. Bioinformatics: What lies beyond bioinformatics? *Nat. Biotechnol.* 15, 3–4.
- Pande, S., Kaftan, F., Lang, S., Svatoš, A., Germerodt, S., Kost, C., 2016. Privatization of cooperative benefits stabilizes mutualistic cross-feeding interactions in spatially structured environments. *ISME J.* 10, 1413–23.
- Pande, S., Merker, H., Bohl, K., Reichelt, M., Schuster, S., de Figueiredo, L.F., Kaleta, C., Kost, C., 2014. Fitness and stability of obligate cross-feeding interactions that emerge upon gene loss in bacteria. *ISME J.* 8, 953–62.
- Parter, M., Kashtan, N., Alon, U., 2007. Environmental variability and modularity of bacterial metabolic networks. *BMC Evol. Biol.* 7, 169.
- Patiño-Navarrete, R., Moya, A., Latorre, A., Peretó, J., 2013. Comparative genomics of *Blattabacterium cuenoti*: the frozen legacy of an ancient endosymbiont genome. *Genome Biol. Evol.* 5, 351–61.
- Patiño-Navarrete, R., Piulachs, M.-D., Belles, X., Moya, A., Latorre, A., Peretó, J., 2014. The cockroach *Blattella germanica* obtains nitrogen from uric acid through a metabolic pathway shared with its bacterial endosymbiont. *Biol. Lett.* 10.
- Pellegrini, M., Marcotte, E.M., Thompson, M.J., Eisenberg, D., Yeates, T.O., 1999. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl. Acad. Sci. U. S. A.* 96, 4285–8.
- Peregrín-Alvarez, J.M., Sanford, C., Parkinson, J., 2009. The conservation and evolutionary modularity of

- metabolism. *Genome Biol.* 10, R63.
- Pérez-Brocal, V., Gil, R., Ramos, S., Lamelas, A., Postigo, M., Michelena, J.M., Silva, F.J., Moya, A., Latorre, A., 2006. A small microbial genome: the end of a long symbiotic relationship? *Science* (80-). 314, 312–313.
- Pfeiffer, T., Bonhoeffer, S., 2004. Evolution of cross-feeding in microbial populations. *Am. Nat.* 163, E126–35.
- Pfeiffer, T., Sánchez-Valdenebro, I., Nuño, J.C., Montero, F., Schuster, S., 1999. METATOOL: for studying metabolic networks. *Bioinformatics* 15, 251–7.
- Pfeiffer, T., Schuster, S., 2005. Game-theoretical approaches to studying the evolution of biochemical systems. *Trends Biochem. Sci.* 30, 20–5.
- Pham-The, H., González-Álvarez, I., Bermejo, M., Garrigues, T., Le-Thi-Thu, H., Cabrera-Pérez, M.Á., 2013. The Use of Rule-Based and QSPR Approaches in ADME Profiling: A Case Study on Caco-2 Permeability. *Mol. Inform.* 32, 459–479.
- Piñero-Fernandez, S., Chimere, C., Keyser, U.F., Summers, D.K., 2011. Indole transport across *Escherichia coli* membranes. *J. Bacteriol.* 193, 1793–8.
- Pinney, J.W., Shirley, M.W., McConkey, G.A., Westhead, D.R., 2005. metaSHARK: software for automated metabolic network prediction from DNA sequence and its application to the genomes of *Plasmodium falciparum* and *Eimeria tenella*. *Nucleic Acids Res.* 33, 1399–409.
- Poliakov, A., Russell, C.W., Ponnala, L., Hoops, H.J., Sun, Q., Douglas, A.E., van Wijk, K.J., 2011. Large-scale label-free quantitative proteomics of the pea aphid-*Buchnera* symbiosis. *Mol. Cell. Proteomics* 10, M110.007039.
- Ponce-de-León, M., Calle-Espinosa, J., Pereto, J., Montero, F., 2015. Consistency Analysis of Genome-Scale Models of Bacterial Metabolism: A Metamodel Approach. *PLoS One* 10, e0143626.
- Ponce-de-León, M., Montero, F., Peretó, J., 2013. Solving gap metabolites and blocked reactions in genome-scale models: application to the metabolic network of *Blattabacterium cuenoti*. *BMC Syst. Biol.* 7, 114.
- Ponce de León, M., Cancela, H., Acerenza, L., 2008. A strategy to calculate the patterns of nutrient consumption by microorganisms applying a two-level optimisation principle to reconstructed metabolic networks. *J. Biol. Phys.* 34, 73–90.
- Poolman, M.G., Sebu, C., Pidcock, M.K., Fell, D.A., 2007. Modular decomposition of metabolic systems via null-space analysis. *J. Theor. Biol.* 249, 691–705.
- Price, D.R.G., Feng, H., Baker, J.D., Bavan, S., Luetje, C.W., Wilson, A.C.C., 2014. Aphid amino acid transporter regulates glutamine supply to intracellular bacterial symbionts. *Proc. Natl. Acad. Sci. U. S. A.* 111, 320–5.
- Price, D.R.G., Wilson, A.C.C., 2014. A substrate ambiguous enzyme facilitates genome reduction in an intracellular symbiont. *BMC Biol.* 12, 110.
- Price, N.D., Famili, I., Beard, D.A., Palsson, B.Ø., 2002. Extreme pathways and Kirchhoff's second law. *Biophys. J.* 83, 2879–82.
- Price, N.D., Reed, J.L., Palsson, B.Ø., 2004. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.* 2, 886–97.
- Priya, V.K., Sarkar, S., Sinha, S., 2014. Evolution of tryptophan biosynthetic pathway in microbial genomes: a comparative genetic study. *Syst. Synth. Biol.* 8, 59–72.

- Rao, Q., Rollat-Farnier, P.-A., Zhu, D.-T., Santos-Garcia, D., Silva, F.J., Moya, A., Latorre, A., Klein, C.C., Vavre, F., Sagot, M.-F., Liu, S.-S., Mouton, L., Wang, X.-W., 2015. Genome reduction and potential metabolic complementation of the dual endosymbionts in the whitefly *Bemisia tabaci*. *BMC Genomics* 16, 226.
- Reed, J.L., Famili, I., Thiele, I., Palsson, B.O., 2006a. Towards multidimensional genome annotation. *Nat. Rev. Genet.* 7, 130–41.
- Reed, J.L., Patel, T.R., Chen, K.H., Joyce, A.R., Applebee, M.K., Herring, C.D., Bui, O.T., Knight, E.M., Fong, S.S., Palsson, B.O., 2006b. Systems approach to refining genome annotation. *Proc. Natl. Acad. Sci. U. S. A.* 103, 17480–17484.
- Reed, J.L., Vo, T.D., Schilling, C.H., Palsson, B.O., 2003. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904) (GSM/GPR). *Genome Biol.* 4, R54.
- Rivera-Perez, C., Nouzova, M., Lamboglia, I., Noriega, F.G., 2014. Metabolic analysis reveals changes in the mevalonate and juvenile hormone synthesis pathways linked to the mosquito reproductive physiology. *Insect Biochem. Mol. Biol.* 51, 1–9.
- Romano, A.H., Conway, T., 1996. Evolution of carbohydrate metabolic pathways. *Res. Microbiol.* 147, 448–55.
- Romero, P., Wagg, J., Green, M.L., Kaiser, D., Krummenacker, M., Karp, P.D., 2005. Computational prediction of human metabolic pathways from the complete human genome. *Genome Biol.* 6, R2.
- Rouhbakhsh, D., Clark, M.A., Baumann, L., Moran, N.A., Baumann, P., 1997. Evolution of the tryptophan biosynthetic pathway in *Buchnera* (aphid endosymbionts): studies of plasmid-associated *trpEG* within the genus *Uroleucon*. *Mol. Phylogenet. Evol.* 8, 167–76.
- Rouhbakhsh, D., Lai, C.Y., von Dohlen, C.D., Clark, M.A., Baumann, L., Baumann, P., Moran, N.A., Voegtlin, D.J., 1996. The tryptophan biosynthetic pathway of aphid endosymbionts (*Buchnera*): genetics and evolution of plasmid-associated anthranilate synthase (*trpEG*) within the aphididae. *J. Mol. Evol.* 42, 414–21.
- Ruppin, E., Papin, J.A., de Figueiredo, L.F., Schuster, S., 2010. Metabolic reconstruction, constraint-based analysis and game theory to probe genome-scale metabolic networks. *Curr. Opin. Biotechnol.* 21, 502–10.
- Russell, C.W., Bouvaine, S., Newell, P.D., Douglas, A.E., 2013. Shared metabolic pathways in a coevolved insect-bacterial symbiosis. *Appl. Environ. Microbiol.* 79, 6117–23.
- Russell, C.W., Poliakov, A., Haribal, M., Jander, G., van Wijk, K.J., Douglas, A.E., 2014. Matching the supply of bacterial nutrients to the nutritional demand of the animal host. *Proc. Biol. Sci.* 281, 20141163.
- Sabree, Z.L., Kambhampati, S., Moran, N.A., 2009. Nitrogen recycling and nutritional provisioning by *Blattabacterium*, the cockroach endosymbiont. *Proc. Natl. Acad. Sci. U. S. A.* 106, 19521–6.
- Samal, A., Matias Rodrigues, J.F., Jost, J., Martin, O.C., Wagner, A., 2010. Genotype networks in metabolic reaction spaces. *BMC Syst. Biol.* 4, 30.
- Santos, F., Boele, J., Teusink, B., 2011. A practical guide to genome-scale metabolic models and their analysis. *Methods Enzymol.* 500, 509–532.
- Sauro, H.M., Ingalls, B., 2004. Conservation analysis in biochemical networks: computational issues for software writers. *Biophys. Chem.* 109, 1–15.
- Savinell, J.M., Palsson, B.O., 1992. Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism. *J. Theor. Biol.* 154, 421–54.

- Schellenberger, J., Lewis, N.E., Palsson, B.Ø., 2011a. Elimination of thermodynamically infeasible loops in steady-state metabolic models. *Biophys. J.* 100, 544–53.
- Schellenberger, J., Park, J.O., Conrad, T.M., Palsson, B.Ø., 2010. BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics* 11, 213.
- 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.Ø., 2011b. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc.* 6, 1290–307.
- Schilling, C.H., Letscher, D., Palsson, B.O., 2000. 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., Schuster, S., Palsson, B.O., Heinrich, R., Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol. Prog.* 15, 296–303.
- Schomburg, I., Chang, A., Placzek, S., Söhngen, C., Rother, M., Lang, M., Muretto, C., Ulas, S., Stelzer, M., Grote, A., Scheer, M., Schomburg, D., 2013. BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Res.* 41, D764-72.
- Schuster, S., Fell, D.A., Dandekar, T., 2000. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.* 18, 326–332.
- Segrè, D., Vitkup, D., Church, G.M., 2002. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. U. S. A.* 99, 15112–7.
- Seth, E.C., Taga, M.E., 2014. Nutrient cross-feeding in the microbial world. *Front. Microbiol.* 5, 350.
- Shachrai, I., Zaslaver, A., Alon, U., Dekel, E., 2010. Cost of unneeded proteins in *E. coli* is reduced after several generations in exponential growth. *Mol. Cell* 38, 758–67.
- Shlomi, T., Berkman, O., Ruppin, E., 2005. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc. Natl. Acad. Sci. U. S. A.* 102, 7695–700.
- Simonet, P., Duport, G., Gaget, K., Weiss-Gayet, M., Colella, S., Febvay, G., Charles, H., Viñuelas, J., Heddi, A., Calevro, F., 2016. Direct flow cytometry measurements reveal a fine-tuning of symbiotic cell dynamics according to the host developmental needs in aphid symbiosis. *Sci. Rep.* 6, 19967.
- Sloan, D.B., Moran, N.A., 2012a. Endosymbiotic bacteria as a source of carotenoids in whiteflies. *Biol. Lett.* 8, 986–989.
- Sloan, D.B., Moran, N.A., 2012b. Genome reduction and co-evolution between the primary and secondary bacterial symbionts of psyllids. *Mol. Biol. Evol.* 29, 3781–92.
- Smallbone, K., Simeonidis, E., Swainston, N., Mendes, P., 2010. Towards a genome-scale kinetic model of cellular metabolism. *BMC Syst. Biol.* 4, 6.
- Sowa, S.M., Keeley, L.L., 1996. Free amino acids in the hemolymph of the cockroach, *Blaberus discoidalis*. *Comp. Biochem. Physiol. A. Physiol.* 113, 131–4.
- Spraggon, G., Kim, C., Nguyen-Huu, X., Yee, M.C., Yanofsky, C., Mills, S.E., 2001. The structures of anthranilate synthase of *Serratia marcescens* crystallized in the presence of (i) its substrates, chorismate and glutamine, and a product, glutamate, and (ii) its end-product inhibitor, L-tryptophan. *Proc. Natl. Acad. Sci. U. S. A.* 98, 6021–6.
- Stein, L., 2001. Genome annotation: from sequence to biology. *Nat. Rev. Genet.* 2, 493–503.

- Stephanopoulos, G., Vallino, J.J., 1991. Network rigidity and metabolic engineering in metabolite overproduction. *Science* 252, 1675–81.
- Stolyar, S., Van Dien, S., Hillesland, K.L., Pinel, N., Lie, T.J., Leigh, J. a, Stahl, D. a, 2007. Metabolic modeling of a mutualistic microbial community. *Mol. Syst. Biol.* 3, 92.
- Suthers, P.F., Dasika, M.S., Kumar, V.S., Denisov, G., Glass, J.I., Maranas, C.D., 2009. A Genome-Scale Metabolic Reconstruction of *Mycoplasma genitalium*, iPS189. *PLoS Comput. Biol.* 5.
- Taffs, R., Aston, J.E., Brileya, K., Jay, Z., Klatt, C.G., McGlynn, S., Mallette, N., Montross, S., Gerlach, R., Inskeep, W.P., Ward, D.M., Carlson, R.P., 2009. In silico approaches to study mass and energy flows in microbial consortia: a syntrophic case study. *BMC Syst. Biol.* 3, 114.
- Tan, J., Zuniga, C., Zengler, K., 2015. Unraveling interactions in microbial communities - from co-cultures to microbiomes. *J. Microbiol.* 53, 295–305.
- Tanaka, R., 2005. Scale-Rich Metabolic Networks. *Phys. Rev. Lett.* 94, 168101.
- Tanaka, R., Csete, M., Doyle, J., 2005. Highly optimised global organisation of metabolic networks. *Syst. Biol. (Stevenage)*. 152, 179–84.
- Tanaka, T., Tateno, Y., Gojobori, T., 2005. Evolution of vitamin B6 (pyridoxine) metabolism by gain and loss of genes. *Mol. Biol. Evol.* 22, 243–50.
- Tang, X.F., Ezaki, S., Atomi, H., Imanaka, T., 2001. Anthranilate synthase without an LLES motif from a hyperthermophilic archaeon is inhibited by tryptophan. *Biochem. Biophys. Res. Commun.* 281, 858–65.
- Tetko, I. V, Poda, G.I., 2004. Application of ALOGPS 2.1 to predict log D distribution coefficient for Pfizer proprietary compounds. *J. Med. Chem.* 47, 5601–4.
- Textor, S., Wendisch, V.F., De Graaf, A.A., Müller, U., Linder, M.I., Linder, D., Buckel, W., 1997. Propionate oxidation in *Escherichia coli*: evidence for operation of a methylcitrate cycle in bacteria. *Arch. Microbiol.* 168, 428–36.
- Thiele, I., Palsson, B.Ø., 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* 5, 93–121.
- Thiele, I., Swainston, N., Fleming, R.M.T., ..., Palsson, B.Ø., 2013. A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.* 31, 419–25.
- Thiele, I., Vlassis, N., Fleming, R.M.T., 2014. FASTGAPFILL : Efficient gap filling in metabolic networks. *Bioinformatics* btu321-.
- Thomas, G.H., Zucker, J., Macdonald, S.J., Sorokin, A., Goryanin, I., Douglas, A.E., 2009. A fragile metabolic network adapted for cooperation in the symbiotic bacterium *Buchnera aphidicola*. *{BMC} Syst. Biol.* 3, 24.
- UniProt-Consortium, 2014. UniProt: a hub for protein information. *Nucleic Acids Res.* 43, D204–12.
- Urbanczik, R., Wagner, C., 2005. Functional stoichiometric analysis of metabolic networks. *Bioinformatics* 21, 4176–4180.
- Valencia, A., 2005. Automatic annotation of protein function. *Curr. Opin. Struct. Biol.* 15, 267–74.
- Vallino, J.J., 2003. Modeling microbial consortiums as distributed metabolic networks. *Biol. Bull.* 204, 174–9.
- Van Leuven, J.T., Meister, R.C., Simon, C., McCutcheon, J.P., 2014. Sympatric speciation in a bacterial endosymbiont results in two genomes with the functionality of one. *Cell* 158, 1270–80.

- Varma, A., Palsson, B.O., 1994. Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. *Nat. Biotechnol.* 12, 994–998.
- Varma, A., Palsson, B.O., 1993. Metabolic capabilities of *Escherichia coli*: I. synthesis of biosynthetic precursors and cofactors. *J. Theor. Biol.* 165, 477–502.
- Veber, D.F., Johnson, S.R., Cheng, H.-Y., Smith, B.R., Ward, K.W., Kopple, K.D., 2002. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* 45, 2615–23.
- Vellozo, A.F., Véron, A.S., Baa-Puyoulet, P., Huerta-Cepas, J., Cottret, L., Febvay, G., Calevro, F., Rahbé, Y., Douglas, A.E., Gabaldón, T., Sagot, M.-F., Charles, H., Colella, S., 2011. CycADS: an annotation database system to ease the development and update of BioCyc databases. *Database (Oxford)*. 2011, bar008.
- Větrovský, T., Baldrian, P., 2013. The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLoS One* 8, e57923.
- Vigneron, A., Masson, F., Vallier, A., Balmand, S., Rey, M., Vincent-Monégat, C., Aksoy, E., Aubailly-Giraud, E., Zaidman-Remy, A., Heddi, A., 2014. Insects recycle endosymbionts when the benefit is over. *Curr. Biol.* 24, 2267–2273.
- Vitkin, E., Shlomi, T., 2012. MIRAGE: a functional genomics-based approach for metabolic network model reconstruction and its application to cyanobacteria networks. *Genome Biol.* 13, R111.
- Vlassis, N., Pacheco, M.P., Sauter, T., 2014. Fast reconstruction of compact context-specific metabolic network models. *PLoS Comput. Biol.* 10, e1003424.
- Wagner, A., 2005. Distributed robustness versus redundancy as causes of mutational robustness. *Bioessays* 27, 176–88.
- Wagner, A., Fell, D.A., 2001. The small world inside large metabolic networks. *Proc. Biol. Sci.* 268, 1803–10.
- Wagner, C., Urbanczik, R., 2005. The geometry of the flux cone of a metabolic network. *Biophys. J.* 89, 3837–45.
- Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J., 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–91.
- Weaver, D.S., Keseler, I.M., Mackie, A., Paulsen, I.T., Karp, P.D., 2014. A genome-scale metabolic flux model of *Escherichia coli* K-12 derived from the EcoCyc database. *BMC Syst. Biol.* 8, 79.
- White, D., Drummond, J.T., Fuqua, C., 2012. *The Physiology and Biochemistry of Prokaryotes*. Oxford University Press.
- Whitehead, L.F., Douglas, A.E., 1993. Populations of Symbiotic Bacteria in the Parthenogenetic Pea Aphid (*Acyrtosiphon pisum*) Symbiosis. *Proc. R. Soc. B Biol. Sci.* 254, 29–32.
- Wilcox, J.L., Dunbar, H.E., Wolfinger, R.D., Moran, N.A., 2003. Consequences of reductive evolution for gene expression in an obligate endosymbiont. *Mol. Microbiol.* 48, 1491–500.
- Wilson, a C.C., Ashton, P.D., Calevro, F., Charles, H., Colella, S., Febvay, G., Jander, G., Kushlan, P.F., Macdonald, S.J., Schwartz, J.F., Thomas, G.H., Douglas, a E., 2010. Genomic insight into the amino acid relations of the pea aphid, *Acyrtosiphon pisum*, with its symbiotic bacterium *Buchnera aphidicola*. *Insect Mol. Biol.* 19 Suppl 2, 249–58.
- Wixon, J., Kell, D., 2000. The Kyoto encyclopedia of genes and genomes—KEGG. *Yeast* 17, 48–55.

- Wright, J., Wagner, A., 2008. Exhaustive identification of steady state cycles in large stoichiometric networks. *BMC Syst. Biol.* 2, 61.
- Wu, D., Daugherty, S.C., Van Aken, S.E., Pai, G.H., Watkins, K.L., Khouri, H., Tallon, L.J., Zaborsky, J.M., Dunbar, H.E., Tran, P.L., Moran, N.A., Eisen, J.A., 2006. Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biol.* 4, e188.
- Yamada, T., Waller, A.S., Raes, J., Zelezniak, A., Perchat, N., Perret, A., Salanoubat, M., Patil, K.R., Weissenbach, J., Bork, P., 2012. Prediction and identification of sequences coding for orphan enzymes using genomic and metagenomic neighbours. *Mol. Syst. Biol.* 8, 581.
- Yang, F., Qian, H., Beard, D.A., 2005. Ab initio prediction of thermodynamically feasible reaction directions from biochemical network stoichiometry. *Metab. Eng.* 7, 251–9.
- Yanofsky, C., 1981. Attenuation in the control of expression of bacterial operons. *Nature* 289, 751–8.
- Yi, P., Han, Z., Li, X., Olson, E.N., 2006. The mevalonate pathway controls heart formation in *Drosophila* by isoprenylation of Ggamma1. *Science* 313, 1301–3.
- Yus, E., Maier, T., Michalodimitrakis, K., van Noort, V., Yamada, T., Chen, W.-H., Wodke, J. a H., Güell, M., Martínez, S., Bourgeois, R., Kühner, S., Raineri, E., Letunic, I., Kalinina, O. V, Rode, M., Herrmann, R., Gutiérrez-Gallego, R., Russell, R.B., Gavin, A.-C., Bork, P., Serrano, L., 2009. Impact of genome reduction on bacterial metabolism and its regulation. *Science* 326, 1263–8.
- yWorks, 2013. yEd Graph Editor.
- Zelezniak, A., Andrejev, S., Ponomarova, O., Mende, D.R., Bork, P., Patil, K.R., 2015. Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proc. Natl. Acad. Sci.* 112, 201421834.
- Zhou, W., Nakhleh, L., 2011. Properties of metabolic graphs: biological organization or representation artifacts? *BMC Bioinformatics* 12, 132.
- Zhuang, K., Vemuri, G.N., Mahadevan, R., 2011. Economics of membrane occupancy and respiration. *Mol. Syst. Biol.* 7, 500.
- Zientz, E., Dandekar, T., Gross, R., 2004. Metabolic interdependence of obligate intracellular bacteria and their insect hosts. *Microbiol. Mol. Biol. Rev.* 68, 745–70.
- Zomorodi, A.R., Maranas, C.D., 2012. OptCom: A multi-level optimization framework for the metabolic modeling and analysis of microbial communities. *PLoS Comput. Biol.* 8, e1002363.