UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
Maestría y Doctorado en Ciencias Bioquímicas

"Construcción de un modelo a escala genómica del metabolismo de *Pseudomonas chlororaphis* DSM 50083 y evaluación de sus capacidades metabólicas"

TESIS
QUE PARA OPTAR POR EL GRADO DE:
Doctor en Ciencias

PRESENTA:
Fabián Moreno Avitia

TUTOR PRINCIPAL:
Dr. José Adelfo Escalante Lozada
Instituto de Biotecnología UNAM

MIEMBROS DEL COMITÉ TUTOR
Dra. Gloria Soberón
Instituto de Investigaciones Biomédicas UNAM

Dr. José Utrilla Carreri
Centro de Ciencias Genómicas, UNAM

Cuernavaca, Morelos. Septiembre, 2021

UNIVERSIDAD NACIONAL AUTÓNOMA

DE MÉXICO

INSTITUTO DE BIOTECNOLOGÍA

Construcción de un modelo a escala genómica del metabolismo de *Pseudomonas chlororaphis* DSM 50083 y evaluación de sus capacidades metabólicas

T E S I S

QUE PARA OBTENER EL GRADO DE:

DOCTOR EN CIENCIAS

P R E S E N T A:

M en C. Fabián Moreno Avitia

DIRECTOR DE TESIS:

Dr. José Adelfo Escalante Lozada

Cuernavaca, Morelos, Septiembre de 2021.

COMITÉ TUTOR:

Dr. José Adelfo Escalante Lozada

Dr. José Utrilla Carreri

Dra. Gloria Soberón

JURADO DE EXAMEN:

Dra. Elda Guadalupe Espín Ocampo

Dra. Rosa María Gutiérrez Ríos

Dr. Alejandro García Rubio

Dr. Luis Servín González

Dr. Mauricio Trujillo Roldán

AGRADECIMIENTOS

Gracias al Dr. José Adelfo Escalante Lozada por muchos motivos. Por aceptarme como alumno de maestría. Por darme la oportunidad de trabajar en su grupo de investigación y por el apoyo constante. Agradezco que me haya aceptado como estudiante de doctorado y darme la confianza de iniciar en un área de investigación completamente nueva para mí. Me dio la libertad de proponer cosas nuevas, establecer colaboraciones, equivocarme y aprender de ello. No puedo enlistar todo lo que aprendí durante el tiempo que trabajé bajo su asesoría. Gracias por darme el ejemplo de lo que es ser un excelente investigador. Después de 8 años, le agradezco haberme ayudado a lograr una de mis más grandes metas.

Al Dr. José Utrilla, gracias por sus muchas asesorías y consejos a lo largo de todo mi doctorado. Gran parte de este trabajo se debe a su participación.

A la Dra. Gloria Soberón, gracias por aportar su gran conocimiento durante todas las reuniones sobre mi trabajo y por la asesoría en el trabajo de laboratorio.

A lo miembros del jurado, Dra. Elda Guadalupe Espín, Dra. Rosa María Gutiérrez, Dr. Alejandro García, Dr. Luis Servín y Dr. Mauricio Trujillo. Gracias por el tiempo y esfuerzo invertido para mejorar este trabajo.

Al Dr. Guillermo Gosset, por aportar siempre ideas para mejorar mi trabajo y que me haya permitido trabajar en sus proyectos de investigación. Lo considero un gran ejemplo a seguir como investigador y líder de grupo.

Al Dr. Alfredo Martínez, por brindarme asesoría en diferentes momentos durante mi trabajo de maestría y doctorado. Aprendí mucho de sus trabajos y su visión de la investigación, sin duda, es un referente para mí.

Al Dr. Francisco Bolívar, por dar siempre el ejemplo como investigador y como gran divulgador de la ciencia.

A la M.C. Georgina Hernández, por haberme ayudado constantemente a lo largo de mi estancia en este grupo. Gracias por haberme enseñado a trabajar con equipos cromatográficos de principio a fin.

Gracias a mis compañeros y amigos del laboratorio Man, Susy, Dulce, Alma, Eliseo, Carlos, Estefanía, Andrea, Caheri, y Fernando. La compañía de gente como ustedes siempre aligera el trabajo.

A César y Andrés, por su amistad, enseñanza y apoyo. Por ayudarme a entender dos visiones muy diferentes de ingeniería de vías metabólicas, que me ayudaron a desarrollar este trabajo.

A la señora Mercedes y Aurelia (Meche y Aure), gracias por toda su ayuda. Son dos de las personas más importantes en la historia de este grupo de investigación.

A Toño y Gloria, por facilitar todo mediante un excelente trabajo.

A mis amiga Esmeralda, gracias, con pocas personas me he entendido tanto.

A mis amigos Leonel y Maritere, gracias por hacerme compañía durante todo este tiempo, por la ayuda, por las comidas y por todo.

A Evelyn, mi novia. Gracias por tu compañía, apoyo y paciencia. Gracias a este tiempo contigo tengo una visión diferente de muchas cosas. Gracias por ayudarme a entender que equivocarse está bien y que puede ser divertido…para ti.

A mi familia, que además de agradecerles, les dedico este trabajo. A mi madre y a mi padre, Rosa María y Crispín, está demás decir que este trabajo no existiría sin su apoyo incondicional. A mis hermanos Zuleyka, Crispín y Farah, gracias por ser mis ejemplos a seguir. A mis sobrinos Johanna, Leonardo, Renata, ellos me ayudaron a esta siempre contento con sus michas llamadas y mensajes.

CONTENIDO

Índice de tablas

Índice de figuras

RESUMEN

Pseudomonas chlororaphis es una bacteria que se asocia a las raíces de diversas plantas, la cual presenta actividad antagónica contra diferentes organismos y propiedades promotoras del crecimiento vegetal. *P. chlororaphis* posee relevantes características desde el punto de vista biotecnológico, las cuales comparte con otras especies del mismo género, pero con un fenotipo no patogénico. Parte de papel antagónico contra diversos organismos se debe a la producción de fenazinas y pirrolnitrina. Para incrementar la base de conocimiento disponible sobre este organismo, en este trabajo se construyó el primer modelo a escala genómica de *P. chlororaphis* DSM 50083, el cual contiene 1267 genes y 2289 reacciones metabólicas. Este modelo fue validado experimentalmente y como caso de estudio, fue analizado para generar estrategias nuevas que favorezcan la producción de fenazina-1-carboxiamida (PCN), las cuales son extrapolables para otras fenazinas. Adicionalmente, el modelo describe el proceso de desnitrificación y la capacidad para consumir sacarosa, trehalosa, manosa y galactosa de *P. chlororaphis.* La optimización de la distribución de los flujos metabólicos *in silico* sugirieron a compuestos con cadenas hidrocarbonadas medianas como los compuestos con mayores rendimientos de PCN. Buscando explotar el potencial del modelo, se realizaron simulaciones utilizando glucosa y glicerol como fuentes de carbono para incrementar la producción de PCN. Los resultados mostraron que una óptima producción de PCN requiere de un incremento del flujo metabólico a través del ciclo de los ácidos tricarboxílicos y la síntesis de glutamina. Finalmente, se encontró que la producción de PCN está estrechamente ligada a la producción de H_2O_2, el cual puede estar generando un estrés oxidativo en cepas productoras de fenazinas. Esto indica que la implementación de las estrategias aquí descritas para mejorar la respuesta a estrés oxidativo podría mejorar el desempeño de los procesos de producción de PCN utilizando a *P. chlororaphis.*

SUMMARY

P. chlororaphis is a bacterium that can be isolated from diverse plant roots, which has been reported as an antagonist for diverse organism and as a plant growth promoter. *P. chlororaphis* possess relevant features from a biotechnological approach, which are shared with other *Pseudomonas,* but in a nonpathogenic phenotype. The antagonistic role against other organism is due to the production of a wide variety of compounds, where we can find pyrrolnitrin and phenazines. To increase the knowledge base of this bacterium, in this work, the first genome scale model of *P. chlororaphis* DSM 50083 was constructed, containing 1267 genes and 2289 metabolic reactions. This model was validated experimentally and applied to analyze the generation of new strategies to improve the productions of phenazine-1-carboxamide (PCN) that can be extrapolated to other phenazines. Additionally, the model describes the denitrification and the metabolism of sucrose, trehalose, mannose, and galactose by *P. chlororaphis.* The *in silico* flux distribution optimization suggested that medium length carbon chain compounds produce the highest PCN yield. To explore the model potential, we simulated possible modifications to increase the production of PCN using glucose and glycerol as carbon sources. Results indicated that an optimum PCN production needs an increase in the metabolic flux through TCA and glutamine synthesis. Finally, we found that the PCN productions is tightly related to the hydrogen peroxide production, which can be generating an oxidative stress in phenazine overproducer strain. This, the implementation of the strategies described here could lead to improve the performance of *P. chlororaphis* in PCN production processes.

INTRODUCCIÓN

En Ingeniería de Vías Metabólicas (IVM) se busca incrementar la producción de uno o más compuestos o modificar una propiedad celular a través de la eliminación, adición o modulación de la expresión de genes de interés mediante el uso de herramientas de ingeniería genética (Stephanopoulos 1999). En esta área de investigación se hace uso de diferentes microorganismos para la producción un gran número de compuestos de interés industrial, como aminoácidos, compuestos aromáticos, polímeros, biocombustibles, antibióticos, péptidos, proteínas, etcétera (Sanchez-Garcia et al. 2016; Yuan et al. 2019; Liu et al. 2021). Entre los organismos usados para la producción de dichos compuestos se encuentran bacterias, hongos, levaduras, plantas, células de insecto y mamíferos (Becker et al. 2011; Ghosh et al. 2015; Kavšček et al. 2015; Sanchez-Garcia et al. 2016).

Para modificar el metabolismo de un organismo es necesario poseer un vasto conocimiento de éste, es decir, conocer su metabolismo, regulación y herramientas disponibles para su modificación genética. Por esta razón, diversos grupos de investigación han optado por la utilización de organismos altamente estudiados como *Escherichia coli, Bacillus* spp., *Saccharomyces cerevisiae, Pichia pastoris* o células de ovario de hámster chino (CHO) para la producción diferentes compuestos (Dong et al. 2014; Ghosh et al. 2015; Kavšček et al. 2015; Sanchez-Garcia et al. 2016; Adamczyk et al. 2017; Gupta et al. 2017; Peña et al. 2018; Xiang et al. 2020). Muchos compuestos producidos industrialmente son tóxicos para algunos de los organismos antes mencionados cuando estos se acumulan, inclusive si se tratan de metabolitos endógenos (Li et al. 2020).

Existen diversos microorganismos no modelo que han desarrollado capacidades metabólicas muy particulares que pueden ser explotadas en biotecnología (Löbs et al. 2017; Li et al. 2020). En las últimas décadas, se han aislado microorganismos capaces de soportar grandes concentraciones de solventes, ya sea por su capacidad de metabolizarlos o bien, por la composición de su membrana citoplasmática (Kusumawardhani et al. 2018). Entre estos microorganismos se encuentran *Bacillus* OS-1906 y DS-994, *Arthrobacter* ST-1 y *Pseudomonas* spp. (Sardessai et al. 2002). El género *Pseudomonas* cuentan con una gran variedad de especies tolerantes a solventes reportadas (Sardessai et al. 2004). Un ejemplo de

tolerancia a altas concentraciones de solventes es *Pseudomonas putida,* capaz de crecer activamente en presencia de hasta 50% V/V de tolueno (Ramos et al. 1995). Por esta característica y otras relacionadas al metabolismo de *P. putida,* ésta ha sido utilizada en los últimos años para la producción de diferentes compuestos, como terpenoides, derivados de aminoácidos, ramnolípidos y policétidos (Nogales et al. 2008; Puchałka et al. 2008; Palsson 2009; Borgos et al. 2013; Loeschcke et al. 2015). Esto hace pensar en el potencial del género *Pseudomonas* para su implementación como modelo estudio para la producción de diversos compuestos.

Características de *Pseudonomas* spp. y su potencial como modelo de producción de compuestos de interés biotecnológico

Diversas especies de *Pseudomonas* son patógenas para distintos organismos (Höfte et al. 2007; Ruffner et al. 2015; Azam et al. 2019). Durante el proceso de colonización del tejido del hospedero, las células de *Pseudomonas* experimentan un estrés oxidativo (provocado por la respuesta inmune inflamatoria), el cual contrarrestan mediante la actividad de su metabolismo central, produciendo equivalentes reductores (NADPH) (Nikel et al. 2014). Los equivalentes reducidos generados ayudan a reducir las especies reactivas de oxígeno, estrategia que sirve de protección durante el proceso de infección. Esto ocurre debido a que *Pseudomonas* spp. carecen de la enzima fosfofructocinasa 1 (Pfk1), dirigiendo el carbono a través de la vía de las pentosas fosfato (VPF) o de la Vía Entner-Duodoroff (VED) produciendo una alta cantidad de NADPH (Nikel et al. 2014). Estudios sugieren que este mecanismo es pieza clave para combatir el estrés oxidativo (Nikel et al. 2020; Cavinato et al. 2020). Desde el punto de vista de ingeniería de vías metabólica, la alta velocidad de regeneración del NADPH, permite la producción de enzimas pertenecientes a vías metabólicas que serían difícilmente soportadas por otros microorganismos (Nikel et al. 2014). Por ejemplo, estudios en *E. coli* están orientados a incrementar la relación de NADPH/NADP para favorecer la producción de ciclohexano, metilhidroxibutirato y propanol (Lee et al. 2013). Por lo tanto, un microorganismo que de forma natural produce niveles elevados de NADPH, resulta atractivo desde el punto de vista biotecnológico (Nikel et al. 2014; Nikel et al. 2020).

4

Otro factor involucrado con la tolerancia a compuestos tóxicos y al éxito de *Pseudomonas* spp. como patógeno es su inusual riqueza en transportadores (hasta 12% del genoma de *P. putida* contra 6.5 % de *E. coli,* Ecocyc, https://biocyc.org/group?id=:ALL-PROTEINS-6&orgid=ECOLI), monooxigenasas, dioxigenasas, oxidorreductasas, ferrodoxinas, citocromos, deshidrogenasas, proteínas relacionadas con el metabolismo de compuestos sulfurados, bombas de secreción y glutatión-*S*-transferasas normalmente asociadas a la protección contra el estrés o compuestos tóxicos (Nelson et al. 2002). Esta capacidad de responder a diferentes tipos de estrés puede ser explotada para fines industriales.

Las bacterias del género *Pseudomonas*, cuentan con capacidades metabólicas poco abundantes o inexistentes en microorganismos modelo como *E. coli* y *S. cerevisiae* (Jiménez et al. 2002; Loeschcke et al. 2015). Dentro de su repertorio, se encuentra la habilidad para degradar compuestos aromáticos como, benzoato, fenilacetato, tolueno, estireno, naftaleno, fenol, policlorobifenilos, dioxinas y nitrotoluenos, aunque los compuestos pueden variar entre especies (Jiménez et al. 2002) (Tabla 1). Además, se ha encontrado que diferentes especies de *Pseudomonas* son capaces de producir una amplia variedad de compuestos de forma natural, que resultan de interés biotecnológicos, entre ellos se encuentran antifúngicos, antibióticos, hormonas de plantas, biosurfactantes y diversos polímeros (Tabla 1) (Gross et al. 2009).

Su gran capacidad de adaptación está ligada al tamaño de su genoma (generalmente >6 Mb), el cual alberga, . Los genes codificantes para las enzimas de degradación de compuestos aromáticos forman parte de dos operones que se encuentran en un plásmido conocido como pTOL:WW0 (Sentchilo et al. 2000), con excepción de unas pocas especies en el que estos genes se encuentran en cromosoma (Sentchilo et al. 2000).

De acuerdo a análisis filogenéticos, el género *Pseudomonas* se puede dividir en 19 grandes grupos: *P. fluorescens, P. gessardii, P. fragi, P. jessenii, P. koreensis, P. mandelii, P. chlororaphis, P. corrugata, P. syringae, P. lutea, P. asplenii, P. putida, P. anguilliseptica, P. straminea, P. auruginosa, P. oleovarans, P. oryzihabitans, P. stutzeri* y *P. pertucinogena* (Gomila et al. 2015). A pesar de esta diversidad, hasta hace algunos años, gran parte de la bibliografía se concentraba en *P. aeruginosa, P. putida y P. fluorescens* (Palleroni et al. 2004).

Las bacterias del género *Pseudomonas* comparten las características antes mencionadas; sin embargo, las diferencias en su metabolismo provocan que posean distintos perfiles de metabolitos secundarios o diferentes rendimientos de éstos (Gross et al. 2009). En el área de ingeniería de vías metabólicas, *P. putida* KT2440 ha sido prácticamente la única explotada para fines de producción. Por esto, es importante estudiar a distintos miembros de este género y así conocer el potencial de estas bacterias como modelos de producción de polímeros, compuestos aromáticos, pigmentos y antifúngicos (Catone et al. 2014; Peng et al. 2018; Sabarinathan et al. 2018; Huccetogullari et al. 2019). En este trabajo, se plantea estudiar el metabolismo de *P. chlororaphis,* una especie no patógena que se distingue por su capacidad de producir acrilamida, pirrolnitrina, ácido 3-indol-acético y fenazinas, pero cuyo metabolismo no ha sido estudiado a profundidad (Ciskanik et al. 1995; Gunther et al. 2005; Pizarro-Tobías et al. 2008; Pierson et al. 2010; Dimkpa et al. 2012).

Tabla 1. Fuentes de carbono metabolizables y compuestos producidos naturalmente por *Pseudomonas spp.*

Fuentes de carbono metabolizables	Compuestos producidos
Glucosa, Xilosa, Glicerol, Tolueno, Xileno, Acetato, Quinato, Benzoato, Fenilacetato, Cafeato, Nicotinato	Polihroxialcanoatos Alginato Fenazinas Bacteriocinas Policétidos Ramnolípidos Terpenoides Compuestos aromáticos

(Jiménez et al. 2002; Loeschcke et al. 2015)

Pseudomonas chlororaphis

P. chlororaphis es una bacteria no patógena que habita en la rizosfera, frecuentemente asociada a las raíces de plantas de arroz, soya, colza y tomate (EFSA Panel on BIOHAZ 2009; Zhao et al. 2013; McCully et al. 2014; Town et al. 2016). Esta bacteria es reconocida

por su utilidad para el biocontrol de plagas como *Fusarium oxysporum* (pudrición basal), *Sclerotinia sclerotiorum* (moho blanco)*, Galleria mellonella* (polilla de cera)*, Spodoptera littoralis* (rosquilla negra)*, Heliothis virescens* (gusano del fruto)*, Plutella xylostella* (palomilla dorso de diamante) y *Bombyx mori* (gusano de seda) (figura 2) (Gunther et al. 2005; Wang et al. 2015)*.* La capacidad de esta bacteria para controlar distintos tipos de plagas está ligada a la producción de diversos compuestos como ramnolípidos, pirrolnitrina y fenazinas. Además, *P. chlororaphis* es capaz de promover el crecimiento de plantas mediante la producción de ácido 3-indol-acético (Gunther et al. 2005; Shen et al. 2013).
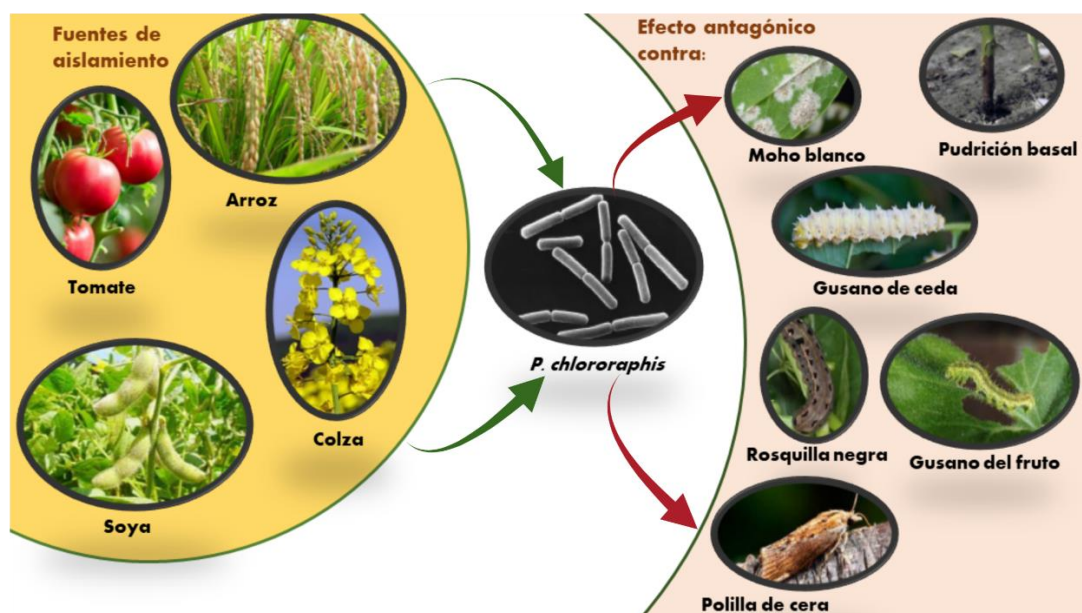
Figura 1. Fuentes de origen y organismos blanco del efecto antagónico de *P. chlororaphis.* Este microorganismo puede ser aislado de muestras de raíces de tomate, arroz, soya y colza. En este ambiente natural, *P. chlororaphis* desarrolla un papel como biocontrolador de plagas como moho blanco, pudrición basal, gusano de ceda, rosquilla negra, gusano de fruto, polilla de cera, entre otros.

Anteriormente se demostrado la diversidad de compuestos que distintas cepas de *P. chlororaphis* pueden utilizar como fuente de carbono, entre los que podemos encontrar: manitol, cetogluconato, fenilacetato, antranilato, diversos aminoácidos, inositol, manosa, glucosa y compuestos aromáticos (Nishimori et al. 2016).

P. chlororaphis es una bacteria desnitrificante, capaz de utilizar como último aceptor de electrones compuestos como NO_3^-, NO_2^-, NO y N_2O para producir N_2 (Nishimori et al. 2016). Esta característica puede ser explotada a nivel industrial para el tratamiento de aguas residuales (Schipper et al. 2010). Queda mucho por estudiar en torno al metabolismo de *P. chlororaphis* y al ampliar dicho conocimiento, será posible explotar el potencial de esta bacteria para producir industrialmente compuestos que ésta sintetiza de forma natural. En este trabajo se plantea el estudio la cepa *P. chlororaphis* DSM 50083 (también reportada como ATCC 9446, CCM 1975, ICPB 2392, IFO 3904, NBRC 3904, NCIB 9392, NRRL B-560, NCBI Taxonomy Browser) dado que ya existen reportes de su implementación en al menos una patente y al mismo tiempo no se ha estudiado a fondo este microorganismo (Huang et al. 1966, Sonoyama et al. 1975, Yokozeki et al. 1977). Por lo tanto, este trabajo contribuye a incrementar la base de conocimiento de esta bacteria a nivel de cepa y especie.

Biología de sistemas para el estudio de redes metabólicas

Anteriormente, el estudio del metabolismo consistía en estudiar las reacciones que se consideran parte de una misma vía metabólica y manipular a voluntad esta red como un elemento individual (Stephanopoulos 1999). Este enfoque ha demostrado ser útil, sin embargo, se pueden omitir ciertos aspectos importantes del metabolismo, como los intercambios de metabolitos entre diferentes vías a nivel celular. El metabolismo de un organismo está constituido por cientos de reacciones, las cuales no actúan de forma individual. Alteraciones en una vía metabólica podría tener repercusiones sobre otras vías con las que podría creerse no existe una interconexión.

El enfoque en biología de sistemas es integrativo y global, es decir, trata de simular las interacciones del mayor número de elementos posibles de un sistema a través de modelos matemáticos. La aplicación de este enfoque al estudio del metabolismo ha resultado en modelos donde se incluyen múltiples vías metabólicas interconectadas hasta lo que hoy en día se conoce como Modelos a Escala Genómica (MEGs) (Feist et al. 2006; Thiele et al. 2010; Monk et al. 2013). En estos modelos se busca construir una red metabólica con el mayor número de reacciones realizadas por las enzimas codificadas en el genoma de un

organismo. Los MEGs han sido de gran utilidad para entender el funcionamiento de diversos microorganismos y tejidos celulares (Famili et al. 2003; Feist et al. 2006; Oberhardt et al. 2008; King et al. 2015; Zuñiga et al. 2016; Magnúsdóttir et al. 2017; Özcan et al. 2019; Nogales et al. 2020).

Gracias al avance de las tecnologías *ómicas*, el desarrollo de la biología de sistemas se ha visto acelerado. En particular, la secuenciación masiva es cada vez más accesible, por lo que obtener el genoma completo de uno o más organismos es factible para el desarrollo de un proyecto. Para la construcción de la red metabólica de un organismo, es necesaria la secuenciación y anotación de su genoma. A partir de la anotación del genoma, se elabora un borrador de la red metabólica (el cual debe ser curado manualmente) (Nogales 2014). Este borrador, debe incluir todas las reacciones metabólicas conocidas para dicho organismo (Thiele et al. 2010).

El modelo completamente curado debe ser convertido en un modelo matemático para poder simular el metabolismo. La metodología de análisis es la denominada Modelado con Base en Restricciones (Constraint-Based Modeling) (Schellenberger et al. 2011; Agren et al. 2013). En este método, las reacciones son convertidas en una matriz estequiométrica donde cada reacción es representada por una columna y cada metabolito por una fila (Nogales 2014). Como su nombre lo indica, cada elemento de la matriz es el coeficiente estequiométrico de dicho metabolito dentro de la reacción en cuestión. Este modelo puede ser utilizado para diferentes análisis, pero el más común para evaluar el funcionamiento y certeza del modelo es el Análisis de Balance de Flujos (ABF) (Varma et al. 1994). En el ABF se asume un estado estacionario, donde la concentración de los metabolitos no cambia respecto al tiempo. Gracias a esta suposición, se puede utilizar esta matriz como un gran sistema de ecuaciones lineales y calcular los flujos mediante programación lineal, donde la función objetivo es seleccionada según el propósito del análisis (Varma et al. 1994). Las restricciones utilizadas para optimizar la función objetivo son la direccionalidad de las reacciones, límites de flujo superiores e inferiores para determinadas reacciones, velocidades de consumo de las fuentes de carbono y producción de determinado metabolito en cierta condición (Thiele et al. 2010).

Actualmente existen diferentes herramientas computacionales para llevar a cabo este análisis, dentro de las más utilizadas se encuentra COBRA (Constrain Based Reconstruction

Analysis) y RAVEN (Reconstruction Analysis and Visiualization of Metabolic Networks), ambos como herramientas de Matlab (Matlab toolbox) (Schellenberger et al. 2011; Agren et al. 2013). Una vez que el modelo es confiable, se puede utilizar como guía para contextualizar datos transcriptómicos y fluxómicos, como guía en ingeniería metabólica, probar hipótesis *in silico* y analizar propiedades de la red metabólica (Oberhardt et al. 2009).

Análisis de balance de flujos aplicado a ingeniería de vías metabólicas

El ABF aplicado a un modelo metabólico a gran escala de un organismo, es una herramienta poderosa en el área de IVM. Este análisis ha sido utilizado anteriormente para observar los efectos de alteraciones en el metabolismo central sobre el resto de las reacciones dentro de la red metabólica (King et al. 2015). A través del ABF es posible calcular el máximo rendimiento teórico (el cual toma en cuenta el mayor número de reacciones dentro de un organismo) y la distribución de los flujos dentro del metabolismo central para conseguir un rendimiento óptimo (Chubukov et al. 2018; Klamt et al. 2018).

Dado que el ABF se basa en maximizar una función objetivo, generalmente se asume la reacción de formación de biomasa como objetivo celular (Varma et al. 1994). Esto supone que el microorganismo busca la máxima generación de biomasa. Si existen discrepancias entre las reacciones que generan los metabolitos de la biomasa y su consumo, el organismo no será capaz crecer *in silico*, o simplemente no reproducirá los datos experimentales (Nogales 2014). Por lo que mediante el uso de ABF es posible proponer reacciones que por omisiones o errores en la anotación del genoma de un organismo no se encuentran dentro de la red metabólica. Este tipo de análisis suele conducir a correcciones en la anotación de genomas o inclusive, a encontrar reacciones que se creían ausentes en un organismo en particular (Guzmán et al. 2015). Por otro lado, buscar la validación de genes propuestos como esenciales ha llevado al hallazgo de isoenzimas no predichas por las herramientas de anotación de genomas (Guzmán et al. 2015). Al generarse una mutante *in vivo* de un gen esencial, el organismo en cuestión debería perder la capacidad de crecer; sin embargo, al mantener su capacidad para generar biomasa hace notar la presencia de enzimas que pueden

suplir a la enzima faltante. Éste tipo de análisis también han permitido conocer ejemplos de promiscuidad enzimática (Guzmán et al. 2015).

Actualmente existen herramientas basadas en el ABF para rediseñar de forma racional el metabolismo y así conseguir una mayor producción de un compuesto. Para esto, se han diseñado algoritmos que son capaces de sugerir la eliminación de reacciones de la red metabólica (Burgard et al. 2003; Patil et al. 2005). Además, estos algoritmos permiten sugerir la incorporación de reacciones heterólogas mediante el uso de bases de datos de reacciones metabólicas de diferentes organismos (Pharkya et al. 2004). De la misma forma, es posible predecir el fenotipo de mutantes y estimar que genes deben ser sobre expresados (Segre et al. 2002; Ranganathan et al. 2010). Muchas más herramientas están siendo desarrolladas incluyendo algoritmos cada vez más complejos.

ANTECEDENTES

Actualmente se cuenta con un gran número de MEGs de diferentes microorganismos, construidos para distintos fines, como la producción de aminoácidos, polímeros o precursores de estos, combustibles, biorremediación, entender interacción patógeno-hospedero, estudiar metabolismo de tejidos humanos y líneas cancerígenas (Famili et al. 2003; Nogales et al. 2008; Oberhardt et al. 2009; Orth et al. 2011; Borgos et al. 2013; Babaei et al. 2014; Swainston et al. 2016; Bartell et al. 2017; Özcan et al. 2019; Nogales et al. 2020).

El modelo más completo de una *Pseudomonas* es el modelo iJN1462 de *P. putida* KT2440, el cual cuenta con 2927 reacciones, 2153 metabolitos y 1462 genes (Nogales et al. 2020). Anteriormente, una versión previa de este modelo se usó para comparar la producción de PHAs con distintas fuentes de carbono (Nogales et al. 2008). El proceso de construcción de este modelo permitió encontrar vías de síntesis de PHAs y predecir el fenotipo de la cepa silvestre y de las mutantes generadas (Nogales et al. 2008).

La producción de alginato en *P. fluorescens* ha sido estudiada mediante ABF a escala genómica, comparando la cepa silvestre contra una mutante sobre productora de este polímero. Mediante el análisis del metabolismo se encontró que la condición ideal para la producción de alginato es usar cultivos limitados en nitrógeno y durante la fase estacionaria (Borgos et al. 2013).

Los MEGs no sólo son útiles para fines de producción, también han resultado importantes para entender el metabolismo de bacterias patógenas y diseñar estrategias para combatirlas (Raghunathan et al. 2009; Dunphy et al. 2018). Bajo esta idea, se construyó un MEG de *P. aeruginosa* PAO1 con 1030 genes y 883 metabolitos. Este modelo puede ser optimizado para simular el metabolismo de *P. aeruginosa* dentro del ambiente generado por sus hospederos (Puchałka et al. 2008).

Los distintos modelos de *Pseudomonas* han sido utilizados con fines de comparación de su metabolismo (Babaei et al. 2014). Este estudio buscaba evaluar mediante ABF la capacidad de los MEGs para simular las diferencias metabólicas de organismos filogenéticamente cercanos como *P. putida, P. aeruginosa* y *P. fluorescens*, puesto que sus redes metabólicas

resultan similares. Estas simulaciones fueron comparadas con evidencia experimental previamente reportada para estos organismos, encontrando que estos modelos sí son capaces de emular tales diferencias (Babaei et al. 2014).

PLANTEAMIENTO DEL PROBLEMA

El enfoque tradicional de IVM se centra en la recolección de información sobre el metabolismo del organismo de interés, concentrando esta búsqueda en la vía metabólica relacionada con el metabolito que se desea producir. Este enfoque ha dado buenos resultados para la producción de diversos compuestos; sin embargo, en ocasiones, las modificaciones genéticas implementadas pueden no garantizar una mejora del desempeño de los microorganismos (Ohnishi et al. 2002; Becker et al. 2011). Con el uso de MEGs y ABF se han podido crear diseños racionales de microorganismos, los cuales han permitido el incremento del rendimiento del compuesto de interés (Nogales et al. 2008; Becker et al. 2011).

Como se mencionó anteriormente, en este trabajo se planteó el estudio del metabolismo de *P. chlororaphis* DSM 50083 por su potencial como nuevo modelo de estudio y de producción industrial de diferentes compuestos. Este estudio se realizó con un enfoque integrativo de biología de sistemas, a través de la construcción de un MEG de este microorganismo. Con esto, se buscó analizar *in silico* el metabolismo de *P. chlororaphis* DSM 50083 para estudiar cuantitativamente sus capacidades de consumo de sustratos y producción de compuestos de interés. Además, se espera poder usar el MEG y ABF para conocer la distribución de los flujos dentro del metabolismo, encontrar genes propuestos como esenciales por el modelo y eliminar vacíos dentro de la red metabólica originados por errores de anotación. Utilizar el modelo como herramienta de diseño celular para IVM con algoritmos ya establecidos.

JUSTIFICACIÓN

Los microorganismos más utilizados dentro del laboratorio han sido seleccionados con base en el conocimiento que se tenía de ellos y a la disponibilidad de herramientas moleculares probadas en estos (Fatma et al. 2020). Es necesario extender el panorama de modelos de producción hacia aquellos microorganismos que naturalmente son capaces de producir compuestos que pueden ser de interés industrial. Por otro lado, también es necesario que dichos organismos sean capaces de consumir diversas fuentes de carbono eficientemente, a la vez que deben ser capaces de combatir el estrés ocasionado por la acumulación de metabolitos. Por las características de las bacterias del género *Pseudomonas*, creemos que *P. chlororaphis* DSM 50083 es una candidata como nuevo modelo de producción. Por último, este microorganismo es reconocido como una especie no patógena, avalado por la European Food Safety Authority, lo que incrementa enormemente su potencial sobre aquellos microrganismos que deben ser modificados para disminuir sus factores de virulencia (EFSA Panel on BIOHAZ 2009). La implementación de tecnologías ómicas nos permite recaudar una gran cantidad de información en poco tiempo y tener un entendimiento general del metabolismo y regulación de un organismo. Mediante un MEG predictivo que conjunte la información disponible de *P. chlororaphis* DSM 50083 será posible realizar diseños racionales para la producción de diversos compuestos.

HIPÓTESIS

La construcción de un modelo a escala genómica de *P. chlororaphis* DSM 50083 nos permitirá simular su fenotipo y los flujos de su red metabólica, conocer sus capacidades metabólicas, y demostrar el potencial de esta bacteria como modelo de producción de fenazinas.

OBJETIVO

Construir y validar experimentalmente un modelo basado en restricciones a escala genómica de *P. chlororaphis* DSM 50083, analizar *in silico* sus capacidades metabólicas y su potencial para la producción de fenazinas utilizando diferentes fuentes de carbono.

Objetivos particulares

- Secuenciar y anotar el genoma de *P. chlororaphis* DSM 50083.
- Recopilar información bibliográfica y generar datos experimentales de *P. chlororaphis*: fuentes de carbono metabolizables, velocidades específicas de crecimiento y velocidad de consumo de oxígeno.
- Construir la red metabólica de *P. chlororaphis* DSM 50083 con base en los genes anotados, la información recolectada y generada experimentalmente.
- Curar el modelo metabólico de *P. chlororaphis* DSM 50083 y validarlo mediante simulación y comparación con datos fenotípicos experimentales :
  - Consumo de diferentes fuentes de carbono.
  - Velocidad específica de crecimiento.
  - Desnitrificación.
- Explorar *in silico* el potencial de *P. chlororaphis* DSM 50083 para la producción de fenazina-1-carboxiamida.

MATERIALES Y MÉTODOS

En la figura 2 se muestra de forma general el esquema de trabajo que se utilizó para conseguir los objetivos planteados anteriormente.
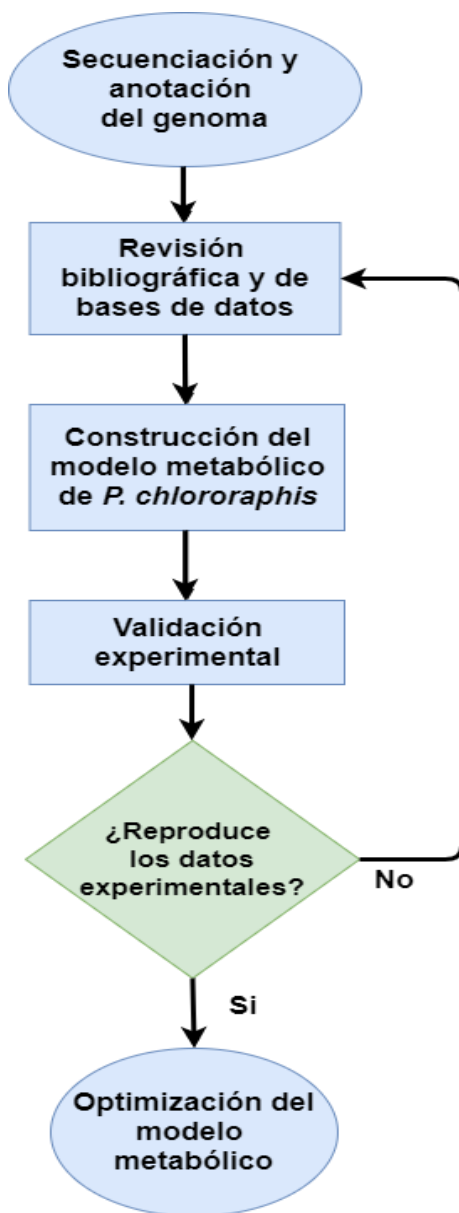
Figura 2. Esquema de las diferentes etapas del proceso de reconstrucción de un MEG. La construcción de modelos metabólicos consiste en una secuencia de tareas con un enfoque iterativo. La actualización del mismo se realiza constantemente hasta obtener una estructura que satisfaga los objetivos planteados.

Reactivación y extracción de ADN cromosomal de *P. chlororaphis* DSM 50083

La cepa bacteriana *P. chlororaphis* DSM 50083 fue adquirida a través de la colección de cultivos ATCC®. El liofilizado de *P. chlororaphis* DSM50083 se resuspendió en 2 mL de caldo nutritivo (BBL nutrient broth, Becton Dickinson), se plaquearon 200 µL en una caja con medio LB sólido sin antibiótico y se incubó a 30°C. Las colonias obtenidas presentaron un fenotipo circular, convexo, entero (de margen liso) y de un color amarillo. A partir de las colonias en medio LB sólido, se aislaron 7 de ellas e inocularon 7 tubos de ensayo con 3 mL de caldo nutritivo (BBL nutrient broth, Becton Dickinson). Se tomaron 0.8 mL de estos cultivos y se mezclaron con un volumen igual de glicerol al 80 % dentro de un criovial de 2 mL. Estos tubos fueron etiquetados con número del uno al siete, siendo el tubo N° 1 el utilizado durante todo el proyecto. Los gliceroles hechos a partir de estas colonias sí presentaron viabilidad y alta velocidad de crecimiento en medio rico (tiempo de duplicación menor a 30 min).

Para la extracción de ADN cromosomal, se inoculó un matraz de 500 mL con 50 mL de caldo nutritivo con 200 µL de un cultivo en medio LB con 16 horas incubación de *P. chlororaphis* DSM 50083. Se incubó a 30 °C por 12 h. Este cultivo se centrifugó a 6010 RCF (centrífuga Eppendorf 5424 R) por 5 min y se resuspendió el pellet en buffer Tris-EDTA pH 7 (trisamoniometano, Tris-HCl 10 mM; ácido etilendiaminotetraacético, EDTA 1mM). Se añadieron 250 µL de SDS (dodecil sulfato de sodio) al 10 %, se agregaron 5 µL de RNAsa (Ribonuclease A, Merk, número de catálogo R6513) y lisozima (Merk, número de catálogo 10837059001) con una punta de espátula, para después incubarse a 37 °C por 45 min. Terminado el tiempo de incubación, se añadieron 750 µL de una solución al 5% de bromuro de cetiltrimetilamonio (CTAB) y NaCl 0.7 M, se agitó la mezcla vigorosamente e incubó por 20 min a 65°C. Después, se añadió un volumen de cloroformo/isoamílico 24:1 y se centrifugó a 6000 rpm durante 10 min. Posterior a esto, se transfirió la fase superior (fase turbia) a un tubo nuevo, este paso se repitió tres veces. Se agregaron 4 mL de isopropanol y mezcló por inversión hasta precipitar el ADN en forma de fibra, se pasó ésta a un tuvo nuevo. El ADN se lavó con 1 mL de etanol al 70 %. Como último paso, se resuspendió el ADN en 3 mL de buffer TE(adaptado de Healey et al. 2014).

Secuenciación y ensamble del genoma de *P. chlororaphis* DSM 50083

El ADN cromosomal extraído se envió a la Unidad de Secuenciación Masiva del Instituto Nacional de Medicina Genómica (INMEGEN). El método de secuenciación utilizado fue de lecturas cortas de 300 bp obtenidas mediante MiSeq Illumina Inc. GAIIx. El análisis de la calidad de las lecturas se realizó utilizando TRIMMGALORE y el posterior ensamble se realizó con los ensambladores SPADES e IDBA_UD (Bankevich et al. 2012; Peng et al. 2012). Los "contigs" obtenidos fueron sometidos a un proceso de unificación (merging) con el protocolo de METASSEMBLERv1.5 (Wences et al. 2015). Posteriormente, se mejoró el ensamble original utilizando SPADESv3.12 en conjunto con datos de secuenciación obtenidos con la tecnología PacBio depositados en NCBI bajo el identificador ASM394576v1 (https://www.ncbi.nlm.nih.gov/assembly /GCF_003945765.1) (Rhoads et al. 2015).

Anotación del genoma de *P. chlororaphis* DSM 50083

La anotación del genoma se realizó mediante tres algoritmos diferentes para obtener una mayor calidad en la predicción de la función de los genes encontrados: RAST (Rapid Annotation using Subsystem Technology), Prokkav1.14.5 y PGAP (NCBI prokaryotic annotation pipeline) (Aziz et al. 2008; Seemann 2014; Tatusova et al. 2016).

La aproximación implementada en PGAP mapea dentro del genoma analizado a proteínas de organismos que pertenecen a la misma rama filogenética (Tatusova et al. 2016). Estas proteínas se seleccionan con base en un análisis pangenómico de un clado filogenético. Esta estrategia asume que si una proteína se encuentra en el 80 % de los miembros de un clado, existe una alta probabilidad de encontrar esa proteína en un nuevo miembro (Tatusova et al. 2016). Posteriormente utiliza el predictor de marcos de lectura GeneMarkS+ que es capaz de incorporar la información de las proteínas alineadas en el paso anterior. Por último, utiliza BLASTp para el alineamiento de las nuevas secuencias de proteínas contra todas las proteínas del resto de las ramas filogenéticas, auxiliándose de las bases de datos Uniprot-SwissProt

Protein Knowledgebase y RefSeq-collection (Camacho et al. 2009; Tatusova et al. 2016). La anotación de NCBI fue obtenida directamente de dicha base de datos bajo el identificador de GenBank GCA_002095925.1.

Prokka utiliza Prodigal para la predicción de las coordenadas de posibles secuencias codificantes (Hyatt et al. 2010; Seemann 2014). Posteriormente, se comparan las secuencias predichas con distintas bases de datos y en orden jerárquico utilizando BLASTp (Camacho et al. 2009). Inicialmente realiza una comparación de secuencias con proteínas con evidencia bioquímica en la base de datos Uniprot, seguida por una búsqueda en bases de datos específicas del género en cuestión en RefSeq (Seemann 2014). Por último, se realiza una búsqueda en bases de datos de dominios y familias proteicas conocidas como Pfam y TIGRFAMs (Bateman 2004; Haft et al. 2012). Las secuencias a las que no se haya podido asignar función se les asigna el encabezado "hypothetical protein" (Seemann 2014).

Para RAST se usó el servidor https://rast.nmpdr.org/rast.cgi, el cual realiza una anotación completamente automática haciendo uso de la base datos genómicos SEED. Esta base de datos es curada manualmente por expertos en metabolismo y fisiología (Aziz et al. 2008; Overbeek et al. 2014). Este sistema de anotación de genomas utiliza familias derivadas de subsistemas (conocidas como FIGfams). Según esta estrategia, un subsistema consiste en un conjunto de actividades funcionales que interactúan para dar lugar a una función celular o metabólica (Aziz et al. 2008). Dentro de este subsistema se crean asociaciones entre estas actividades y genes en genomas específicos, y a su vez, con las proteínas codificadas por estos genes (Aziz et al. 2008). Las diferentes proteínas que son asociadas a la misma actividad dan lugar a las FIGfams. La construcción de estas familias es realizada por múltiples curadores expertos en el subsistema en cuestión y debe seguir unas de las siguientes reglas: 1) las proteínas dentro de una misma FIGfam deben haber sido reportadas con la misma función y su secuencia de aminoácidos debe tener 70 % de similitud; 2) las proteínas pertenecen a un organismo filogenéticamente cercano, es decir, tienen más de 90 % de identidad y los genes asociadas a éstas poseen un contexto genómico similar (Aziz et al. 2008). RAST utiliza GLIMMER2 para la predicción de genes putativos y tRNAscan-SE para la predicción de tRNAs (Delcher et al. 1999; Chan et al. 2019). RAST utiliza las secuencias predichas para compararlas con un pequeño conjunto de proteínas dentro de las

FIGfams y establecer una cercanía filogenética de estos genes putativos mediante BLAST (Aziz et al. 2008). Una vez se conoce qué organismos dentro de las FIGfams son cercanos filogenéticamente al nuevo organismo a anotar, se comparan las nuevas proteínas putativas con las proteínas de estos organismos (Aziz et al. 2008). Este paso agiliza el proceso de anotación e incrementa la probabilidad de que la función asignada a la nueva proteína sea correcta.

Para la generación del modelo se revisaron meticulosamente las anotaciones generadas con las diferentes estrategias antes mencionadas.

Determinación del factor de conversión entre densidad óptica a 600 nm y gramos de peso de células secas por litro

Para calcular el factor de conversión entre DO_{600nm} y gramos de pesos seco por litro (gDCW/L) se llevaron a peso constante seis tubos de centrífuga de 1.5 mL en una estufa a 80 °C. Simultáneamente se realizó un cultivo en un matraz Erlenmeyer de 500 mL con 50 mL de medio LB (Apéndice 2). Al alcanzar aproximadamente una $DO_{600nm} = 8$ (medido con una dilución 1:20 en un espectrofotómetro Beckman DU700), se cosecharon las células y lavaron con NaCl 0.1 M. Posteriormente se resuspendieron en la misma solución salina alcanzando una DO_{600nm} cercana a 100 unidades de absorbancia. A partir de esta suspensión celular concentrada se realizaron cinco diluciones (1:1.5, 1:4, 1:5, 1:20 y 1:40) y se agregó un mililitro de cada de una de ellas a un respectivo tubo seco. Dichos tubos, fueron secados a 60°C durante 16 h en un secador al vacío Eppendorf 5301 Vacufuge Concentrator System. Para finalizar, fueron llevados a peso constante en una estufa a 80 °C. Se pesaron los tubos con la muestra seca y se restó a cada uno de ellos el peso del tubo vacío para finalmente obtener los gramos de células secas por litro (dried cell weight per liter; gDCW/L). Mediante una regresión lineal de DO_{600nm} contra gDCW/L se obtuvo el factor de conversión que corresponde a la función lineal obtenida.

Recopilación de información y obtención de datos bioquímicos de _P. chlororaphis_ DSM 50083

**Evaluación de consumo de diferentes fuentes de carbono por P. chlororaphis**

En este estudio se utilizó una microplaca PM1 de BIOLOG (PM1 MicroPlate™ Carbon Sources, 00A 042 Rev. C), la cual contiene 96 pozos con 95 fuentes de carbono distintas. Esta placa contiene aminoácidos, ácidos orgánicos, carbohidratos, y algunos otros compuestos más complejos. En cada pozo se encuentra liofilizada una fuente de carbono distinta. El pozo A1 de la placa no contiene ninguna fuente de carbono y se usó como control negativo para el crecimiento. Para este experimento se partió de un preinóculo en medio LB con 12 h de incubación, se tomaron 100 µL del mismo e inocularon 25 mL de medio M9 (Apéndice 2) con 2 g/L de glucosa. Éste cultivo se llevó hasta una $DO_{600nm} = 1$ y se tomó el volumen necesario para inocular 25 mL de medio M9 fresco sin fuente de carbono e iniciar a una $DO_{600nm} = 0.01$. A partir del medio M9 sin fuente de carbono inoculado, se vertieron 150 µL en cada uno de los pozos. Esto se realizó en un lector de microplaca (BioTek ELx808IU Ultra Microplate Reader) a 30 °C, agitación rápida (revoluciones por minuto no reportadas por el equipo) y constante, se leyó la absorbancia a 630 nm durante 48 h. En ninguno de los caso el pozo control A1 presentó crecimiento. El resultado de cada pozo fue reportado como positivo en caso de mostrar crecimiento y negativo en el caso contrario.

**Cálculo de velocidad específica de crecimiento en medio M9 con glucosa, glicerol y sacarosa como única fuente de carbono**

Se evaluó el crecimiento en fuentes de carbono utilizadas a nivel industrial como glucosa, glicerol y sacarosa. Se realizaron cultivos de 50 mL en matraces Erlenmeyer de 500 mL, utilizando como base medio de cultivo M9 con las fuentes de carbono anteriormente mencionadas a una concentración de 2 g/L. Estos matraces se incubaron a 30°C con una agitación de 300 rpm.

Se monitoreó el crecimiento celular mediante la medición de la absorbancia del medio de cultivo. Para esto, se tomaron muestras de 0.5 mL cada hora y colocaron en cubetas de plástico para espectrofotómetro. Las mediciones se realizaron en un espectrofotómetro Beckman DU700 ajustando la longitud de onda para una lectura a 600 nm. La velocidad específica de crecimiento (μ) se calculó mediante una regresión lineal utilizando la forma linealizada de la ecuación de crecimiento exponencial:

$$lnX = lnXo + \mu t$$

Donde X es la densidad celular en el tiempo t y Xo la densidad óptica en el tiempo cero.

A partir de estos cultivos también se calcularon las velocidades de consumo de las tres fuentes de carbono mencionadas. Para esto, se calculó el rendimiento de biomasa alcanzado en cada fuente de carbono (Y_{xs}). Esto rendimiento de biomasa se calculó mediante la siguiente formula:

$$Y_{xs} = \frac{\Delta X}{\Delta S}$$

Donde ΔX es el cambio de biomasa medido en gramos y ΔS es el cambio de sustrato en gramos para el mismo intervalo de tiempo. A partir de estos parámetros es posible calcular la velocidad específica de consumo de la fuente de carbono (q_s). Los valores de q_s se calcularon para todas las fuentes de carbono probadas en este experimento con la siguiente formula:

$$q_s = \frac{\mu}{Y_{xs}}$$

Cultivos continuos de P. chlororaphis en medio M9 para determinar el coeficiente de mantenimiento y la velocidad específica de consumo de oxigeno

Se realizaron cultivos continuos en reactores de 1 L (Applikon controlador ADI 1010 BIO, consola ADI 1025 BIO y controlador de agitación ADI 1032) con 500 mL de volumen de operación utilizando medio M9 con 2 g/L de glucosa como fuente de carbono. La aireación se realizó manteniendo un flujo de aire constante igual a 1 vvm y una agitación de 600 rpm

con un impulsor Rushton. El pH se mantuvo en 7.0 alimentando NaOH 2 M y la temperatura a 30 °C. La alimentación se inició al alcanzar una biomasa de 0.8 g/L, concentración celular a la cual la glucosa era cercana a cero. Para la determinación del coeficiente de mantenimiento (m) en glucosa se utilizaron 3 velocidades de dilución (D) diferentes: 0.37, 0.25 y 0.1 h^{-1}. La alimentación se realizó con una manguera de silicón del número 13 y dos bombas peristálticas Masterflex (L/S drive 07522-20, Easy-Load® II pump head 77200-62) con las revoluciones correspondientes para cada D. Se monitoreó el estado estacionario a partir de tres tiempos de retención (t) y se mantuvo durante cinco t.

El coeficiente de mantenimiento se calculó mediante una regresión lineal utilizando la siguiente ecuación derivada de la ecuación de Pirt:

$$\frac{1}{Y_{app}} = \frac{1}{Y_G} + \frac{m}{\mu}$$

Donde Y_{app} es el rendimiento de biomasa aparente y Y_G es el rendimiento de biomasa (van Bodegom 2007).

Para calcular la velocidad específica de consumo de oxígeno se analizó la concentración de oxígeno en la tubería de salida de los gases del biorreactor. La composición de este aire se realizó utilizando un sensor de oxígeno (Bluesense BCP-O2). Los valores obtenidos de concentración de oxígeno en la entrada y salida de aire se introdujeron en la siguiente fórmula para calcular la velocidad específica de consumo de oxígeno:

$$OUR = \frac{Q}{V}(O_{2in} - O2_{out}) - \frac{\Delta C_L}{\Delta t}$$

Donde OUR es la velocidad de consumo de oxígeno (mol/Lh), Q es el flujo de aire suplementado al bioreactor (L/h), V es el volumen del biorreactor, O_{2in} y O_{2out} son la concentración de oxígeno en la entrada y salida de aire respectivamente, y $\Delta C_L/\Delta t$ es la velocidad de acumulación de oxígeno disuelto en el medio de cultivo. Este último término se puede considerar cero en estado estable (Garcia-Ochoa et al. 2010).

Construcción de la red metabólica de *P. chlororaphis* DSM 50083

Como punto de partida, se usó la anotación del genoma de *P. chlororaphis* DSM 50083 para la búsqueda de reacciones enzimáticas. Se realizó una comparación con organismos cercanos muy bien estudiados como *P. aeruginosa* PAO1 y *P. putida* KT2440 mediante BLAST recíproco y con la herramienta OrthoVenn (Xu et al. 2019). OrthoVenn es servidor web que facilita la búsqueda de genes ortólogos e ilustra el resultado en forma de un diagrama de Venn (Nichio et al. 2017). OrthoVenn usa OrthoAgogue para la identificación de genes ortólogos, esta herramienta utiliza los pares de genes mejor puntuados por BLAST para establecer los grupos de ortólogos (Ekseth et al. 2014). Las listas de genes ortólogos obtenidas se revisaron manualmente en búsqueda de errores o irregularidades. Adicionalmente, se repitió esta metodología usando el genoma de *E. coli* MG1655. Una vez identificadas las enzimas involucradas en el metabolismo, se utilizó MrBac con el modelo metabólico de *P. putida* KT2440 (iJN1462) como guía para convertir la información genómica a un modelo del tipo SBML (Systems Biology Markup Language) (Liao et al. 2011; Nogales et al. 2020). En este proceso, MrBac utiliza listas de proteínas ortólogas para identificar las reacciones enzimáticas asociadas a esos genes y copiar dichas reacciones desde la base de datos BiGG hacia un nuevo modelo (King et al. 2016; Norsigian et al. 2019). Este proceso omite todas las reacciones enzimáticas que no hayan sido incluidas previamente en un modelo metabólico, por lo que se requiere de un trabajo manual para incluir todas las reacciones restantes.

El modelo metabólico construido en este trabajo está compuesto por un directorio de reacciones enzimáticas que pueden o no, estar asociada a uno o más genes. Este directorio contiene un identificador de reacción, nombre de la reacción, estequiometría de la reacción, asociación gen-proteína-reacción (GPR) y los límites superiores e inferiores de flujo para esta reacción. El parámetro GPR sirve para asociar una reacción metabólica a una o más secuencias codificantes dentro del genoma de *P. chlororaphis*. Esto permite simular inactivación de genes y analizar su efecto dentro de la red metabólica. Los límites de flujo superior e inferior se utilizan como restricciones dentro del problema de optimización de la función objetivo. Esto permite eliminar cualquier resultado que exceda dichos límites de flujo.

Para la creación de este modelo metabólico se construyeron todos los metabolitos asociados a todas las reacciones. Para ejemplificar esta tarea, utilizaremos al metabolito glucosa-6-fosfato:

1. Se asignó un identificador único para este metabolito: glc_D_c
   Este consiste en una simplificación arbitraria del nombre del metabolito y se asignó el sufijo "c" para los metabolitos citosólicos, "e" para los metabolitos extracelulares y "p" para los metabolitos periplásmicos.
2. Se agregó la formula química al metabolito glc_D_c de la siguiente forma: C6H12O6. Metabolitos que se encuentren protonados o desprotonados a pH intracelular fueron incluidos con la respectiva modificación en su fórmula química.
3. Se asignó el nombre completo del compuesto para propósitos de revisión:
   Glucose

Simulación del metabolismo de *P. chlororaphis* mediante análisis de balance de flujos (ABF)

Un modelo metabólico es generalmente una lista de reacciones balanceadas y con anotaciones para especificar características de éstas, por ejemplo, identificador de reacción, nombre de la enzima que cataliza la reacción, direccionalidad y localización celular (Figura 3). Este formato legible para los usuarios debe ser convertido a un problema matemático interpretable por los programas de simulación de flujos metabólicos. En este caso, se utilizó COBRApy para la conversión del modelo a una matriz estequiométrica (S) (Ebrahim et al. 2013). COBRApy junto con su versión de MATLAB han sido utilizados para el análisis de más de 108 modelos metabólicos contenidos en BiGG (Norsigian et al. 2019).

A partir de la información de todas las reacciones enzimáticas contenidas en el modelo metabólico, se construyó una matriz estequiométrica S (ejemplificado en la figura 4). Esta matriz contiene la red metabólica donde las filas representan a los metabolitos y las columnas a las reacciones enzimáticas. Cada elemento de la matriz representa el coeficiente estequiométrico del metabolito "m_i" en la reacción "n_i". La velocidad de producción o consumo de los compuestos dentro de la red, es igual al producto de la matriz S por el vector de flujos (v= v1, v2,...):

$$\frac{dx}{dt} = Sv$$

Donde x es la concentración del metabolito, t tiempo y v el vector de flujos.

Dado que no se conocen las constantes cinéticas de las reacciones dentro de la célula, se asume un estado estacionario, donde las concentraciones de los metabolitos no cambian respecto al tiempo, por lo tanto:

$\frac{dx}{dt} = Sv = 0$

Eso convierte el problema en un sistema de ecuaciones lineales.

Reaction identifier	HEX1
Name	R_hexokinase__D_glucoseATP_
Memory address	0x02ab6a4edeb8
Stoichiometry	atp_c + glc_D_c --> adp_c + g6p_c + h_c M_ATP_C10H12N5O13P3 + M_D_Glucose_C6H12O6 --> M_ADP_C10H12N5O10P2 + M_D_Glucose_6_phosphate_C6H11O9P + M_H_H
GPR	ORM46639.1
Lower bound	0.0
Upper bound	999999.0

Figura 3. Estructura de una reacción enzimática dentro de un modelo metabólico SBML. En la imagen se muestra como ejemplo la reacción realizada por la enzima glucoquinasa (HEX1 por convención de la base de datos BiGG). En este directorio podemos encontrar su nombre en extenso, la reacción enzimática que realiza, la secuencia codificante a la que está asociada y los límites de flujo establecidos para dicha reacción.

Para disminuir el conjunto de vectores que satisfagan esta igualdad es necesario establecer restricciones al sistema de ecuaciones. En este caso se utilizan límites superiores e inferiores para el flujo de cada reacción. Las reacciones irreversibles tienen como restricción $v_{min} > 0$ y $v_{max} > 0$. Las reacciones reversibles pueden tomar valores satisfaciendo la siguiente restricción $v_{min} < 0$ y $v_{max} > 0$. Es necesario destacar que al existir infinitos valores mayores o menores que cero se adiciona un límite superiror para el valor de todas las reacciones. De acuerdo con los modelos en la base de datos de BiGG, ninguna reacción puede tener un flujo mayor a 999999.0 ni menor a -999999.0 bajo ninguna condición (King et al. 2016). Todas las velocidades de los flujos (excepto la reacción de biomasa) están dadas en mmol/gDCWh.

	rxn1	rxn2	rxn3	rxn4	At	Bt	Ft	EX_A(e)	EX_B(e)	EX_F(e)	Sink
A[e]	0	0	0	0	1	0	0	−1	0	0	0
B[e]	0	0	0	0	0	1	0	0	−1	0	0
A[c]	−1	0	0	0	−1	0	0	0	0	0	0
B[c]	−1	0	0	0	0	−1	0	0	0	0	0
C[c]	1	−1	−1	0	0	0	0	0	0	0	0
D[c]	0	0	1	0	0	0	0	0	0	0	−1
E[c]	0	0	1	−1	0	0	0	0	0	0	0
F[c]	0	1	0	1	0	0	−1	0	0	0	0
F[e]	0	0	0	0	0	0	1	0	0	−1	0

Figura 4. Estructura de una matriz estequiométrica. En la figura se muestra un ejemplo de matriz estequiométrica como las implementadas en análisis de balance de flujos. Cada columna y fila dentro de esta matriz representa una reacción y un metabolito de la red metabólica. En este ejemplo se muestran seis metabolitos (A, B, C, D, E y F) y están acompañados del indicador del compartimento celular al cual pertenecen (c: citosol y e: extracelular). También se muestran 11 reacciones de las cuales, cuatro son genéricas (rxn1, rxn2, rxn3 y rxn4), 3 de transporte transmembranal (At, Bt y Ct) y cuatro son reacciones de intercambio (EX_A, EX_B, EX_F y Sink). Los valores dentro de la matriz se asignan de acuerdo con la estequiometria de la reacción y metabolito en cuestión. Los reactantes reciben coeficientes negativos y los productos coeficientes positivos (Heirendt et al. 2019).

Los flujos de la red se obtienen utilizando un método de optimización llamado programación lineal. Este algoritmo busca optimizar la función establecida como objetivo, en este caso dicha función es la reacción de generación de biomasa.

Construcción de un modelo cerrado de *P. chlororaphis* mediante simulaciones

Es común encontrar problemas para la simulación del crecimiento utilizando un MEG, debido a reacciones bloqueadas y reacciones faltantes (Thiele et al. 2010). Las reacciones bloqueadas son aquellas cuyos productos no son consumidos en ninguna reacción y por consecuencia, no son eliminadas del sistema e impiden la resolución del sistema de ecuaciones como un problema de optimización.

Por esto fue necesario crear reacciones especiales conocidas como reacciones de intercambio. Estas reacciones no tienen un trasfondo fisiológico, pero son esenciales para la construcción

de un modelo metabólico. Existen tres tipos de reacciones de intercambio denominadas como "Exchanges" (EX), "Demand" (DM) y "Sinks" (Sink). En la figura 3 se ejemplifican estos tres tipos de reacciones. Las reacciones EX fueron construidas para el suministro y eliminación de compuestos en el medio de cultivo. La composición del medio de cultivo depende de que las reacciones EX de los componentes del medio simulado estén activas. Las reacciones del tipo EX poseen la siente estructura:

$$EX\_glc\_c:\ \text{glc\_D\_e} \rightarrow$$
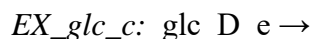
Como se puede observar la eliminación o suministro del metabolito en cuestión depende de la direccionalidad de la reacción. Si el flujo es negativo, el compuesto indicado como reactantes es suministrado al medio, de lo contrario, este metabolito es eliminado del espacio extracelular. Por otro lado, las reacciones DM y Sink cumplen con un objetivo distinto. Las reacciones DM se encargan de eliminar aquellos metabolitos para los cuales se desconoce una reacción enzimática que los utilice como sustrato. Las reacciones DM eliminan al metabolito en cuestión desde el citosol sin pasar por periplasma o espacio extracelular. Las reacciones Sink cumplen la tarea opuesta a las reacciones DM. Los Sink son utilizados cuando se desconoce el origen enzimático de un metabolito que participa en una vía metabólica que se desea estudiar. Estas reacciones suministran dicho compuesto directamente en el compartimento requerido. Los flujos de este tipo de reacciones dependen totalmente de los flujos de las reacciones subsecuentes en la vía metabólica.

Para la identificación de reacciones bloqueadas por la presencia de un metabolito no consumido, se diseñó un breve algoritmo de búsqueda y construcción de reacciones temporales para el consumo de estos. Una vez identificadas las reacciones y metabolitos que impedían la optimización, se buscaron genes en la anotación cuyos productos metabolizaran dichos compuestos o trasportaran este metabolito al espacio extracelular. Posteriormente, se construyeron reacciones con base en el genoma y se remplazaron las reacciones temporales. De igual forma, se buscaron reacciones en las cuales hacía falta la producción de uno de los sustratos y se construyeron las reacciones de tipo Sink necesarias.

Algoritmos para diseño celular

Se aplicaron tres algoritmos diferentes para la predicción de fenotipos sobreproductores de fenazina-1-carboxiamida (PCN), estos fueron OptKnock, OptGene y FSEOF (en español, Barrido de Flujos basado en Flujo del Objetivo Forzado, BFFOF) (Burgard et al. 2003; Patil et al. 2005; Choi et al. 2010). OptKnock se basa en la idea de que el sistema celular tiende a contrarrestar las perturbaciones metabólicas impuestas por modificaciones genéticas, es decir, trata de seleccionar valores de flujos metabólicos que restablezcan la máxima velocidad de crecimiento. Para lograr esto, se crea un nuevo problema de optimización binivel, en el cual se establecen dos funciones objetivo (a diferencia de ABF que sólo permite una a la vez). Este problema se construye anidando un problema dentro de otro. En este caso, se suele utilizar la maximización del crecimiento celular como problema interior, que a su vez, funciona como restricción para el problema exterior que busca maximizar una reacción seleccionada. Esto permite que la solución obtenida esté sujeta a una mínima velocidad de crecimiento. Por otro lado, OptGene utiliza un algoritmo genético que genera poblaciones de mutantes sencillas aleatoriamente y evalúa su desempeño en las tareas asignadas como objetivo, en este caso, la producción de un compuesto y el crecimiento celular. Con las mejores candidatas genera combinaciones y sigue evaluándolas iterativamente hasta cumplir con un número definido de ciclos. Si bien este método no garantiza la convergencia en los resultados obtenidos (a diferencia OptKnok), en la práctica resulta muy efectivo. Finalmente, BFFOF se basa en múltiples Análisis de Variabilidad de Flujos, en el cual se muestrean los intervalos de flujo obtenidos para cada reacción en función de la modificación de la reacción fijada como objetivo (producción de un metabolito). Aquellas reacciones que deban modificar su flujo de manera más significativa son candidatas a modificación genética. Este método resulta ser computacionalmente más económico, pero arroja resultados similares a los métodos antes mencionados. Adicionalmente, BFFOF aporta información no sólo para inactivación de genes sino también para sobre expresión.

RESULTADOS Y DISCUSIÓN

Secuenciación, anotación del genoma y construcción del modelo metabólico de *P. chlororaphis* DSM 50083

Los resultados de la secuenciación del genoma de *P. chlororaphis* DSM 50083 arrojaron una cobertura de 120x. La unificación de los "contigs" obtenidos por SPADES e IDBA_UD resultó en un genoma de 6,783,030 bp separados en 35 "contigs" con un contenido de GC de 63 %. La longitud promedio de los "contigs" fue de 15,918 con un N50 de 843,135 bp (Moreno-Avitia et al. 2017). Como se mencionó en la metodología, este ensamble se mejoró usando datos obtenidos por secuenciación PacBio en un proyecto realizado por otro grupo de investigación (identificador en NCBI ASM394576v1), reduciéndose de 35 a 14 "contigs" con una longitud total de 6,811,866 bp, donde el 99.9 % del genoma está contenido en un solo "contig".

Las secuencias codificantes encontradas según NCBI pipeline, RAST y PROKKA fueron 6128, 6122 y 6100 respectivamente (Aziz et al. 2008; Seemann 2014; Tatusova et al. 2016). De estas secuencias codificantes se encontraron, 81 pseudogenes, 58 tRNAs y 41 genes con cambio de marco de lectura. Estos resultados pueden ser revisados a mayor profundidad en NCBI bajo el identificador de GenBank GCA_002095965.1 o bajo el proyecto WGS NBAT01. En la figura 2 se muestra la distribución de genes encontrados clasificados por subsistemas en el genoma de *P. chlororaphis* DSM 50083 realizado con RAST. Esta agrupación de genes nos ayuda a observar más rápidamente la estructuración del genoma y ver que subsistemas poseen una mayor cantidad de genes. Como se puede observar en la figura 5, la anotación con RAST no fue capaz asociar una gran cantidad de genes al subsistema del metabolismo secundario, sin embargo, esto significa que podría estar constituido por genes poco caracterizados. Cuando se habla de metabolismo secundario, se habla de una gran diversidad de enzimas y reacciones, lo que ocasiona problemas para anotarlos, puesto que se carece de disponibilidad de reportes (Lobb et al. 2020).
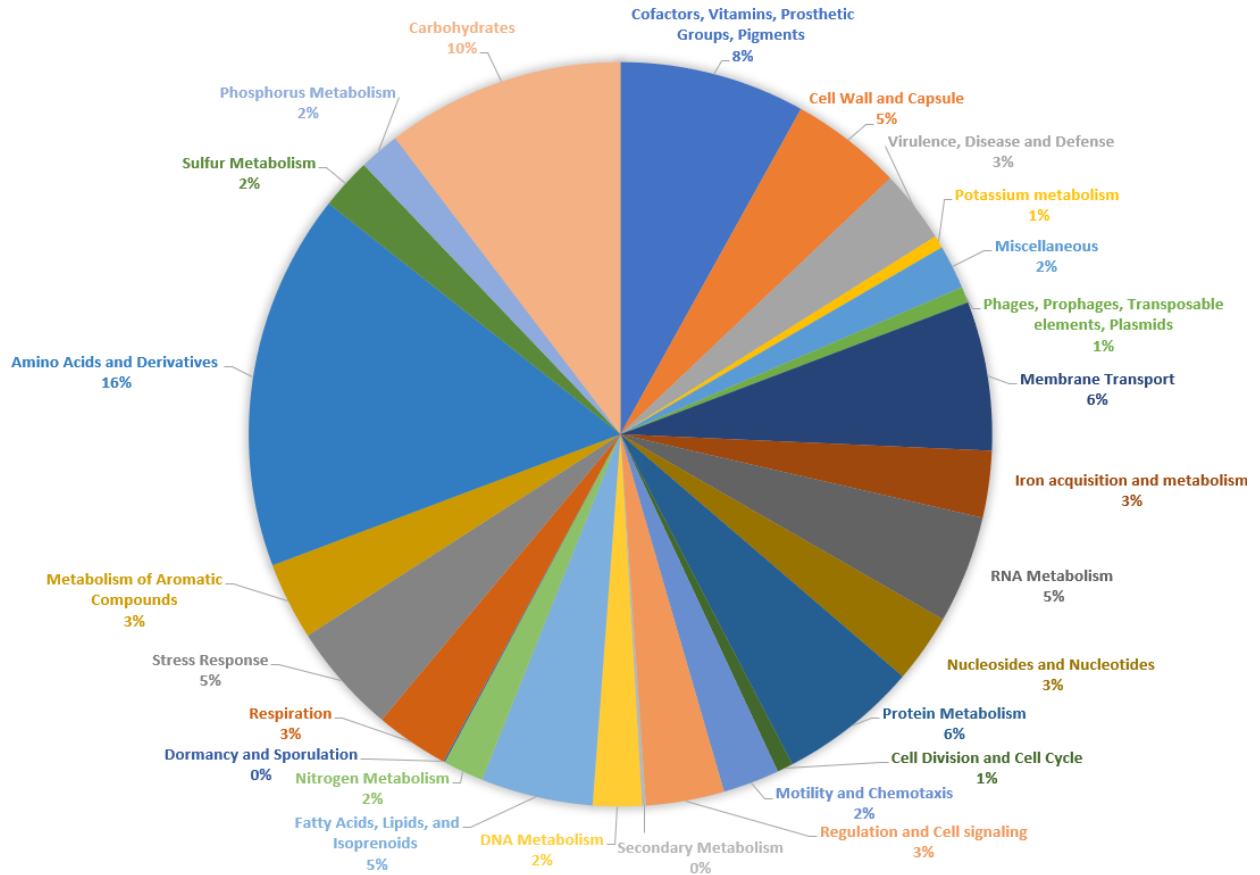
Figura 5. Distribución de los genes de *P. chlororaphis* DSM 50083 entre los diferentes subsistemas. En el gráfico se muestra el porcentaje de genes que fueron asociados a cada subsistema mediante la predicción automática de genes realizada con RAST.

Como se mencionó anteriormente, las fenazinas y pirrolnitrina son parte de los metabolitos más importantes producidos por *P. chlororaphis*. Entre los organismos productores de fenazinas existen variaciones en los compuestos derivados de fenazinas que pueden producir, esto se debe a los genes incluidos en sus operones biosintéticos (Bilal et al. 2017). El operón encargado de la producción de dichos compuestos fue identificado e incluido en el modelo mediante inspección manual usando herramientas como BLAST y sus variantes. En este caso, *P. chlororaphis* DSM 50083 posee el operón *phzABCDEFGH* (figura 6), lo cual indica que la principal fenazina producida es fenazina-1-carboxiamida (PCN).
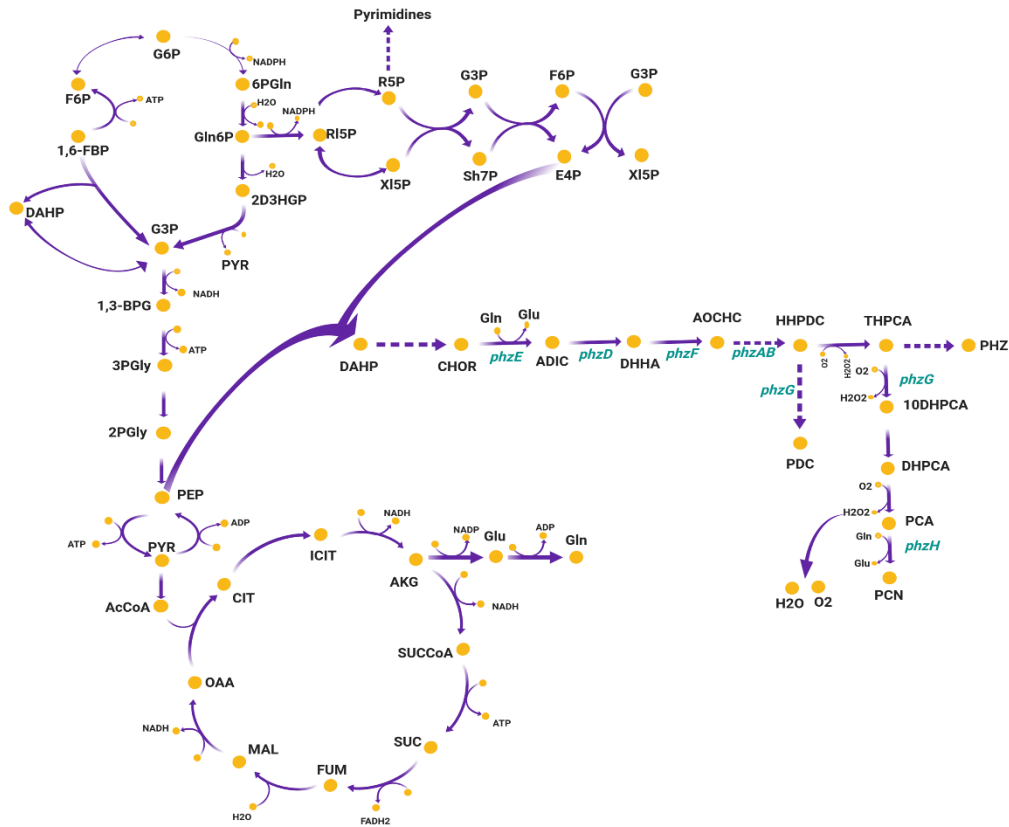
Figura 6. Operón encargado de la producción de fenazinas. La figura a) muestra el operón para la biosíntesis de fenazinas en *P. chlororaphis* DSM 50083, cuyos productos se encargan de la conversión de corismato en fenazina-1-carboxiamida. En la figura b) se puede observar cómo se conecta la vía de síntesis de las fenazinas con la vía del shikimato y a su vez con el metabolismo central. En negro se muestran las abreviaciones de los metabolitos involucrados y en azul los genes pertenecientes al operón *phz*. Las líneas moradas representan las reacciones químicas y las líneas punteadas representan dos o más reacciones consecutivas. G6P: glucosa-6-fosfato, F6P: fructosa-6-fosfato, 1,6-fructosa-bifosfato, DAHP: 3-Deoxi-arabino-heptulosonato-7-fosfato, CHOR: corismato, G3P: gliceraldehído-3-fosfato, 1,3-BPG: 1,3-bisfosfoglicerato, 3PGly: 3-fosfoglicerato, 2PGly: 2-fofoglicerato, PEP: fosfoenolpiruvato, PYR: piruvato, AcCoA: acetil-coenzima A, CIT: citrato, ICIT: isocitrato, AKG: α-cetoglutarato, SUCCoA: succinil-coenzima A, SUCC: succinato, FUM: fumarato, MAL: malato, OAA: oxaloacetato, Glu: glutamato, Gln: glutamina, 6PGln: 6-fosfo-D-glucono-1,5-lactona, Gln6P: gluconato-6-fosfato, 2D3HGP: 2-dehidro-3-deoxi-gluconato-6-fosfato, RL5P: ribulosa-5-fosfato, R5P: ribosa-5-fosfato, Xl5P: xilulosa-5-fosfato, Sh7P: sedoheptulosa-7-fosfato, E4P: eritrosa-4-fosfato, ADIC: 2-amino-2-deoxiisocorismato, DHHA: trans-2,3-dihidro-3-hidroxiantranilato, AOCHC: 6-amino-5-oxociclohex-2-ene-1-carboxilato, HHPDC: 1,4,5,5a,6,9-hexahidrofenazina-1,6-dicarboxilato, PHZ: fenazina, PDC: fenazina-1,6-dicarboxilato, THPCA: 1,4,5,10-tetrahidrofenazina-1-carboxilato, DHPCA: 5,10-dihidrofenazina-1-carboxilato , 10HPCA: dihidrofenazina-1-carboxilato, PCA: fenazina-1-carboxilato, PCN: fenazina-1-carboxiamida.

En cuanto al operón de pirrolnitrina, se encontraron errores en la anotación automática, pero con la ayuda del servidor ANTISMASH dichos genes fueron identificados (Medema et al. 2011). Este operón está compuesto por los genes *prnC, prnB, prnA* y *prnD,* los cuales son responsables de la conversión de triptófano en pirrolnitrina. Este proceso de identificación manual de los genes omitidos por la anotación y curación de estos se realizó repetidamente para tratar de obtener la información genómica más confiable.

Una vez que se contó con una anotación como base, el modelo inicial iMAd fue obtenido de la comparación directa de genes ortólogos entre *P. chlororaphis* DSM 50083 y los genes incluidos en el modelo iJN1462 (Nogales et al. 2020). Dado que el modelo iJN1462 fue meticulosamente curado, este paso ofrece un fuerte soporte para la construcción de un modelo preciso. Para la extensión del modelo se buscaron genes ortólogos en otras *Pseudomonas* bien estudiadas mediante el uso de OthoVenn y BLAST recíproco. La figura 7 muestra un diagrama de Venn los genes compartidos entre *P. aeruginosa* PAO1*, P. putida* KT2440, *E. coli* MG1655 y *P. chlororaphis* DSM 50083. En esta figura podemos observar que estos cuatro organismos comparten 1359 genes*,* los cuales contienen la mayor parte del metabolismo central y síntesis de aminoácidos. Las tres *Pseudomonas* analizadas comparten 1675 genes, los cuales podrían ser compartidos por otras especies del mismo género. Con esta comparación se encontraron 1540 genes únicos de *P. chlororaphis* DSM 50083, donde se incluyen algunos de los genes relacionados con la síntesis de metabolitos secundarios. Se realizó una búsqueda meticulosa para seleccionar los genes presentes únicamente en *P. chlororaphis* DSM 50083*.*

Posterior a la adición de los genes y reacciones manualmente curadas, y la modificación de la reacción de biomasa, el nuevo modelo iMA1267 consistió en 1267 genes, 2289 reacciones y 1810 metabolitos. Este es el primer modelo reportado para *P. chlororaphis* y cubre el 20 % de los 6,122 secuencias codificantes predichas con RAST para el genoma de *P. chlororaphis* DSM 50083. La estructura de este modelo lo coloca cerca de la cobertura de MEGs de organismos mucho más estudiados como *E. coli* y *P. putida,* con una representación de 25 % y 30 % de su genoma respectivamente. Este avance nos indica que siguiendo un protocolo cuidadoso y bien estructurado, es posible obtener información metabólica detallada de un organismo no modelo rápidamente. Esto cobra más relevancia si se considera la gran

cantidad de información genómica que se genera cada vez más rápido y que no siempre es precedida de corroboración bioquímica.
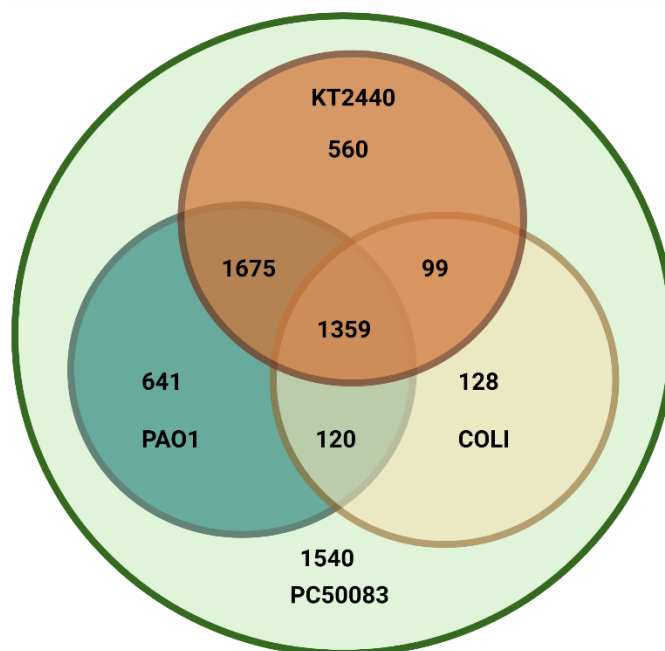
Figura 7. Genes compartidos por *P. chlororaphis, P. putida* y *P. aeruginosa*. El diagrama de Venn muestra los genes compartidos por *P. chlororaphis* (PC50083)*, P. putida* (KT2440), *P. aeruginosa* (PAO1) y *E. coli* MG1655.

Cálculos de parámetros fenotípicos

Fuentes de carbono metabolizables in vivo e *in silico por P. chlororaphis DSM 50083*

Para explorar las capacidades metabólicas de *P. chlororaphis* DSM 50083 para metabolizar diferentes fuentes de carbono, se probó el desempeño de esta cepa en placas de Biolog (PM1 MicroPlate™ Carbon Sources) con 95 fuentes de carbono. Esta determinación fue cualitativa, sólo se evaluaron los resultados como positivo o negativo. La tabla 2 muestra las fuentes de carbono metabolizables por *P. chlororaphis* DSM 50083. Los resultados muestran la oxidación de 49 compuestos, el 50 % de los compuestos probados. Varios de estos resultados concuerdan con lo reportado en la literatura (Burr et al. 2010).

Tabla 2. Fuentes de carbono metabolizables por *P. chlororaphis* DSM 50083.

Nombre del compuesto	Resultado	Nombre del compuesto	Resultado	Nombre del compuesto	Resultado	Nombre del compuesto	Resultado
L-arabinosa	-	D-alanina	+	m-tartrato	-	Adonitol	-
N-acetil-glucosamina	+	D-trehalosa	+	D-glucosa-1-fosfato	-	Maltotriosa	-
D-glucarato	+	D-manosa	+	D-fructosa-6-fosfato	-	2-Deoxiadenosina	-
Succinato	+	Dulcitol	-	Tween 80	+	Adenosina	-
D-galactosa	+	α-glicerol fosfato	-	α-hidroxiglutarato-γ-lactona	+	Propionato	+
D-serina	+	D-xilosa	-	Glicil-L-aspartato	-	Mucato	+
D-sorbitol	-	L-lactato	+	Citrato	+	Glicolato	-
Glicerol	+	Formato	-	m-inositol	+	Glioxilato	-
L-fucosa	-	D-manitol	+	D-treonina	-	D-celobiosa	-
D-glucoronato	-	L-glutamato	+	Fumarato	+	Inosina	+
D-gluconato	+	D-fructosa	+	Glicil-L-glutamato	+	Acetoacetato	-
D-glucosa-6-fosfato	-	Acetato	+	Tricabalilato	-	N-acetil-β-D-manosamina	-
D-galactonato-6-fosfato	+	α-D-glucosa	+	L-serina	+	Monometilsuccinato	-
D-malato	+	Maltosa	+	L-treonina	+	Metilpiruvato	-
L-malato	-	D-melobiosa	-	L-alanina	+	Glucuronato	-
Tween 20	+	Timidina	-	L-alanil-glicina	-	Piruvato	+
L-ramnosa	-	α-cetobutirato	-	Glicil-L-prolina	+	L-galactonato-γ-lactona	-
L-asparagina	+	α-metil-galactosido	-	p-hidroxifenilacetato	+	D-galacturonato	-
D-aspartato	+	α-D-lactosa	-	m- hidroxifenilacetato	-	Fenilmetilamina	-
D-glucosaminato	+	Lactulosa	-	Tiramina	+	2-aminoetanol	+
1,2-propanediol	-	Sacarosa	+	D-psicosa	-	L-prolina	+
Tween 40	+	Uridina	+	L-lixosa	-	L-glutamina	-
α-cetoglutarato	+	α-hidroxibutirato	-	L-asparto	+	β-metill-D-glucosido	-

-, No mostró crecimiento; +, mostró crecimiento

Como podemos observar en la tabla 2, *P. chlororaphis* DSM 50083 es capaz de metabolizar sacarosa, maltosa, trehalosa, manosa y galactosa, carbohidratos raramente consumidos por las *Pseudomonas* mejor estudiadas (Oberhardt et al. 2011; Löwe et al. 2017; Nogales et al. 2020). La utilización de estos compuestos fue confirmada mediante cultivos en tubo de ensayo con 3 mL medio M9 con 2 g/L de cada azúcar. Para adicionar al modelo iMA1267 la capacidad de metabolizar estos compuestos, se buscaron errores de anotación y genes relacionados en el genoma. Existen dos sistemas bien estudiados en bacterias Gram-negativas para el consumo de sacarosa, el *csc* y *scr* (Löwe et al. 2018). El operón *csc* se encuentra en algunas cepas de *E. coli* y el operón *scr* en cepas de *Salmonella*. Interesantemente, la comparación de secuencias de estos genes reveló que dichos operones no se encuentran presentes en *P. chlororaphis* DSM 50083. Otro operón menos estudiado es el *cscRABY* ha sido reportado en *P. protegens* Pf-5 y *P. fluorescences* AU1385 y AU20219, el cual presenta similitudes con ambos operones mencionados anteriormente. Nuevamente, la comparación de secuencias del operón *cscRABY* con el genoma de *P. chlororaphis* DSM 50083 no mostró similitudes mediante BLAST (Löwe et al. 2018).

El único sistema capaz de transportar disacáridos es el sistema fosfotransferasa (phosphotransferase system; PTS) de trehalosa. El trabajo realizado por Steen y colaboradores sugiere cierto nivel de promiscuidad en dicho sistema en *E. coli* por sacarosa (Steen et al. 2014). Aquí se hipotetiza un fenómeno similar para el transporte de sacarosa en la cepa DSM 50083.

A partir de datos genómicos es posible observar la presencia o ausencia de transportadores y enzimas asociadas a la degradación de determinado compuesto. Una vez que se posee un MEG se puede agilizar esta revisión. Estos modelos contienen la integración de estas vías de degradación con el resto del metabolismo, pudiendo así estimar si el compuesto de interés es capaz de sustentar el crecimiento del organismo estudiado. El primer análisis realizado con el modelo iMA1267 fue la predicción de fuentes de carbono metabolizables mediante múltiples ABF usando COBRApy. Para evaluar la utilización de fuentes de carbono, la única restricción usada fue igualar a cero todas las velocidades específicas consumo de carbono excepto la analizada. Esta última fue fijada a 10 mmol/gDCWh mientras que el consumo de oxígeno se dejó sin restricciones. Los minerales nitrógeno y fosfatos fueron mantenidos

igualmente sin restricciones. El modelo iMA1267 fue capaz de crecer en 96 fuentes de carbono diferentes. Discrepancias entre los metabolitos contenidos en la placa de BIOLOG y anotados en el modelo no permitieron la comparación de algunos compuestos. De los 49 compuestos con resultados de crecimiento positivos 26% no están incluidos en el modelo debido a falta de información inicial. Esta nueva información fenotípica puede ser utilizada en futuras versiones del modelo, ejemplificando el enfoque iterativo de la construcción de los MEGs. Los 46 metabolitos con resultados negativos fueron consistentes con los resultados *in silico,* no se encontraron falsos positivos. La tabla 3 compara los resultados de crecimiento *in vivo* e *in silico.*

Determinación del factor de conversión DO$_{600}$nm - gDCW/L

El factor obtenido para la conversión de unidades de densidad óptica a 600 nm y gramos de pesos seco fue de 0.52, con una R^2 de 0.99 (figura 8). Esto nos permite ahora realizar todos aquellos cálculos relacionados con la biomasa generada utilizando la correlación que existe entre la concentración celular en g/L y la densidad óptica medida a 600 nm.
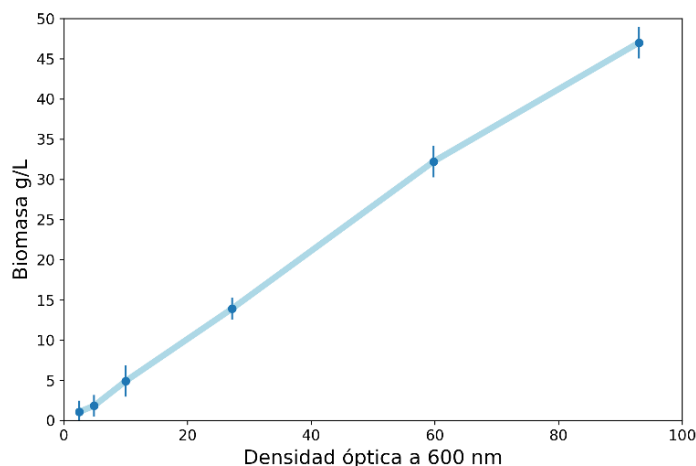
Figura 8. Curva de peso seco contra densidad óptica. El gráfico muestra la relación lineal entre los parámetros de densidad óptica a 600 nm (eje X) y concentración de biomasa (eje Y) de *P. chlororaphis.*

Tabla 3. Comparación *in vivo* e *in silico* de fuentes de carbono metabolizables por *P. chlororaphis* DSM 50083

Fuente de carbono	In vivo	In silico	Fuente de carbono	In vivo	In silico
2-aminoetanol	O	O	Inosina	O	O
Acetato	O	O	L-alanina	O	O
α-D-glucosa	O	O	L-alanil-glicina	O	X
α-hidroxiglutarato-γ-lactona	O	O	L-asparagina	O	O
α-cetoglutarato	O	O	L-aspartato	O	O
Citrato	O	O	L-glutamato	O	O
D-alanina	O	O	L-glutamina	O	O
D-aspartato	O	O	L-lactato	O	O
D-fructosa	O	O	L-malato	O	O
D-galactonato-γ-lactona	O	X	L-prolina	O	O
D-galactosa	O	O	L-serina	O	O
D-gluconato	O	O	L-treonina	O	O
D-glucosaminato	O	X	m-inositol	O	O
D-malato	O	O	Mucicato	O	O
D-manitol	O	O	N-acetil-D-glucosamina	O	O
D-manosa	O	O	p-hidroxifenil acetato	O	O
D-glucarato	O	X	Propionato	O	O
D-serina	O	O	Piruvato	O	O
D-trehalosa	O	O	Succinato	O	O
D,L-malato	O	O	Sacarosa	O	O
Fumarato	O	O	Tween 20	O	X
Glicerol	O	O	Tween 40	O	X
Glcil-L -glutamato	O	X	Tween 80	O	X
Glicil-L-prolina	O	X	Tiramina	O	X
			Uridina	O	O

X, No mostró crecimiento; O, mostró crecimiento

Determinación de parámetros cinéticos de P. chlororaphis

Con el propósito de evaluar la capacidad del modelo para predecir el crecimiento celular, se compararon las velocidades específicas de crecimiento (μ) en glucosa, sacarosa y glicerol. Adicionalmente, se evaluaron diferentes velocidades de consumo de glucosa usando cultivos continuos. *P. chlororaphis* fue capaz de crecer a relativamente a altas velocidades en las tres

fuentes de carbono probadas, siendo glucosa la que produce la mayor velocidad específica de crecimiento (0.44 h^{-1}), seguida por sacarosa (0.39 h^{-1}) y finalmente por glicerol (0.30 h^{-1}) (figura 9). El glicerol proporciona un mayor rendimiento de biomasa, cerca de un 8% superior a los carbohidratos probados. Esta característica ha sido probada previamente en procesos de producción de fenazinas (Hu et al. 2017; Yue et al. 2020). Las velocidades de crecimiento simuladas usando las tasas de consumo experimentales arrojaron valores cercanos a los datos *in vivo* con velocidades ligeramente superiores (tabla 4).

Las figura 10 muestra las cinéticas de crecimiento, consumo de glucosa y oxígeno disuelto (OD) de tres cultivos continuos de *P. chlororaphis* en medio M9 con 2 g/L de glucosa como fuente de carbono. Para poder calcular el coeficiente de mantenimiento, estos cultivos se realizaron a tres velocidades de dilución distintas, 12 h^{-1}, 0.25 h^{-1}, 0.37 h^{-1}. En todos los experimentos se dejó un margen de tres tiempos de retención posteriores al inicio de la alimentación para asegurar la estabilidad del cultivo en estado estacionario. Esto se puede observar con la estabilización de la concentración de la biomasa y el oxígeno disuelto (DO) a lo largo de la etapa con alimentación. El coeficiente de mantenimiento calculado a partir de estos datos es de 0.123 mmol/gDCWh. Este coeficiente se utilizó como restricción para simular la producción de ATP con el modelo iMA1267 y así obtener el ATP de mantenimiento (ATPM = 3.41 mmol/gDCWh). Este valor indica que independientemente la velocidad específica de crecimiento de *P. chlororaphis,* ésta debe reservar 3.41 mmol/gDCWh de ATP para mantenimiento y no pueden ser utilizados en ninguna otra reacción. Por definición, el ABF busca maximizar una función objetivo y necesita de parámetros que restrinjan el espacio de soluciones para lograr obtener valores de flujos metabólicos más realistas. En este caso, el parámetro ATPM se utiliza como restricción para la optimización del modelo y reducir las diferencias entre los valores *in silico* e *in vivo*. Los resultados de las simulaciones mediante ABF se acercan a los valores experimentales en cultivos continuos con una tasa de dilución de 0.12, 0.25 y 0.37 h^{-1} (tabla 5). Estos resultados nos muestran el poder de predicción del modelo iMA1267 en una primera versión y de la confianza que podemos tener en él para realizar diversos análisis.
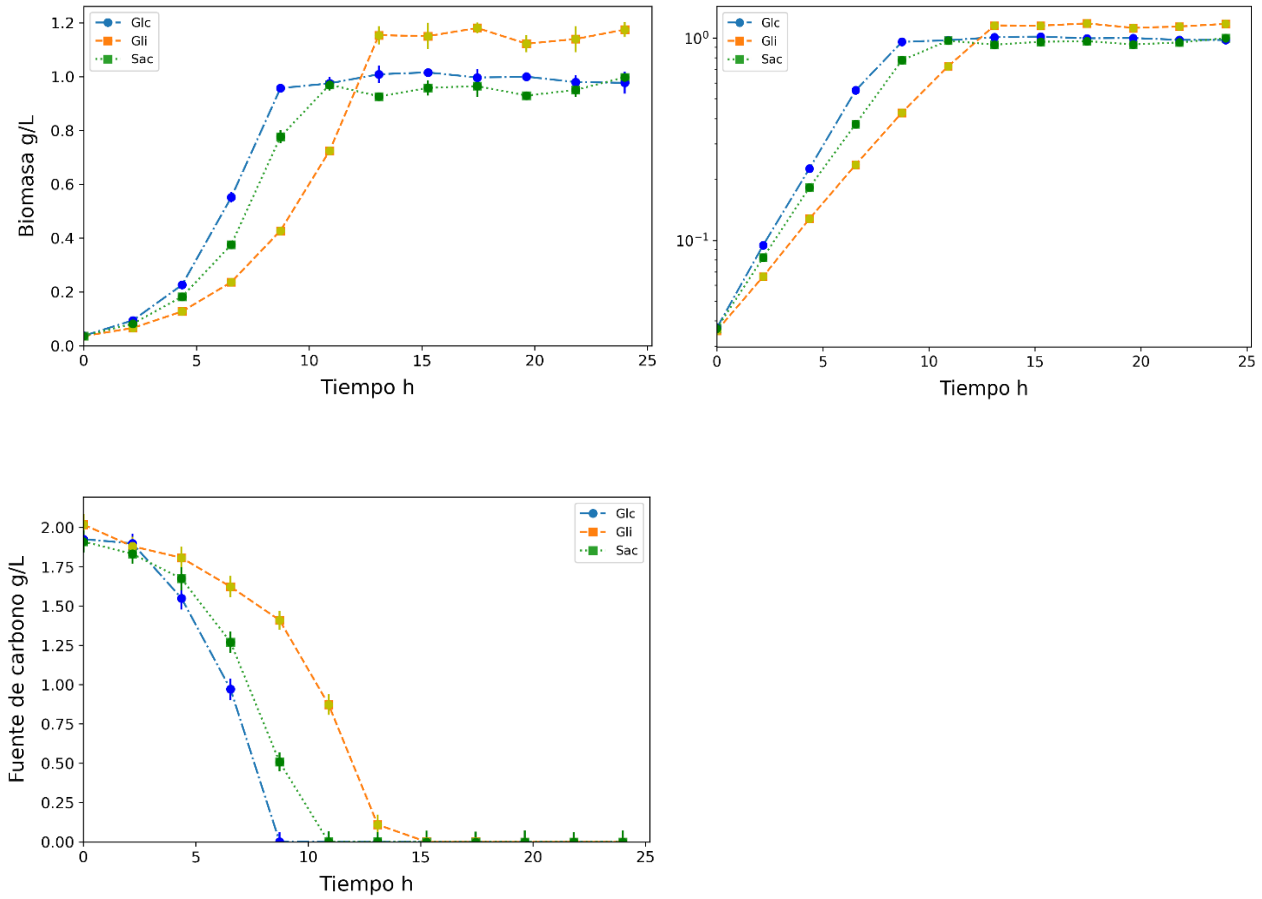
Figura 9. Cinética crecimiento de *P. chlororaphis* en cultivos en lote con diferentes fuentes de carbono. En las imágenes se muestran las curvas de crecimiento de *P. chlororaphis* en cultivos en lote con glucosa (puntos azules), sacarosa (cuadrados verdes) y glicerol (cuadrados anaranjados) como fuente de carbono. Del lado izquierdo se muestra la biomasa en escala lineal y del lado derecho en escala logarítmica. En la parte inferior de la figura se observan las curvas de consumo de las tres fuentes de carbono.
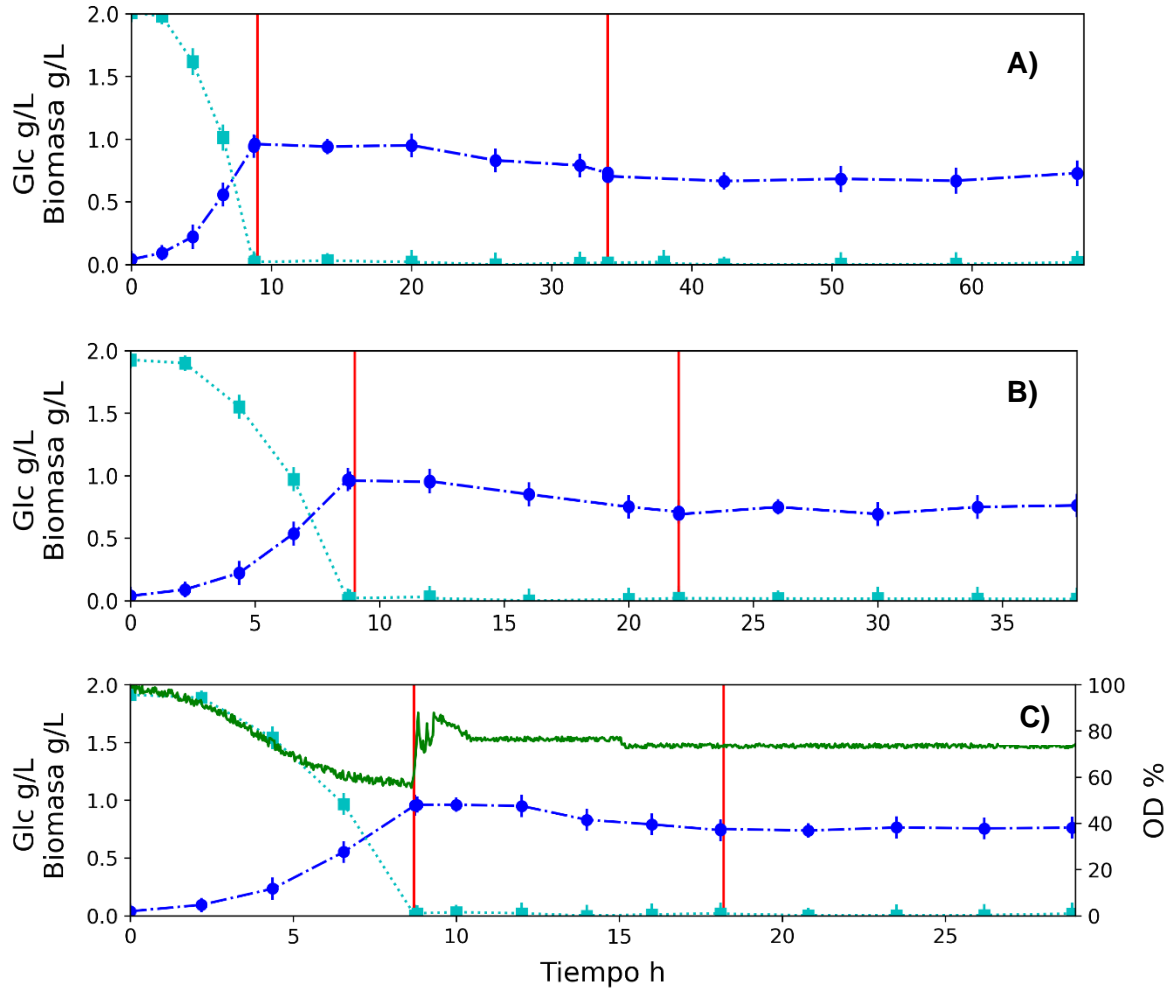
Figura 10. Cultivos continuos de *P. chlororaphis* en medio M9 con glucosa a diferentes velocidades de dilución. En la figura se muestran las curvas de crecimiento (círculos azules), consumo de glucosa (cuadros azul celeste) y oxígeno disuelto (línea verde) de tres cultivos continuos de *P. chlororaphis* en medio M9 con 2 g/L de glucosa como fuente de carbono. Las figuras A), B) y C) pertenecen a las velocidades de dilución 0.12 h^{-1}, 0.25 h^{-1}, 0.37 h^{-1} respectivamente. El tiempo transcurrido entre las líneas rojas verticales es el periodo de estabilización del cultivo continuo.

Tabla 4. Comparación de parámetros cinéticos *in vivo* e *in silico* de *P.chlororaphis* en medio M9

Condición	O_2 mmol/gDCWh	q_s mmol/gDCWh	μ h^{-1}
In vivo Glc	NA	5.16 ± 0.02	0.44 ± 0.01
In silico Glc	10.3	5.16	0.49
In vivo Sac	NA	2.3 ± 0.01	0.39 ± 0.01
In silico Sac	7.92	2.3	0.44
In vivo Gli	NA	5.9 ± 0.01	0.30 ± 0.02
In silico Gli	5.63	5.9	0.34

NA, No analizado, Glc: glucosa, Sac: sacarosa, Gli : glicerol.

Tabla 5. Comparación de las velocidades de crecimiento en estado estacionario

D (h^{-1})	q_s (mmol/gDCWh)	q_{O2} (mmol/gDCWh)	μ (h^{-1})*	q_{O2} (mmol/gDCWh)*
0.37	4.30 ± 0.03	9.01 ± 0.04	0.41	8.20
0.25	3.48 ± 0.07	NA	0.33	6.67
0.12	2.08 ± 0.04	NA	0.18	4.05

NA, No analizado, * *In silico*

Desnitrificación

Otra característica relevante para incluir en el modelo es la capacidad de desnitrificación en *P. chlororaphis.* Esta bacteria es capaz de reducir nitrato a óxido nitroso, pero es incapaz de reducir este último en nitrógeno molecular. El operón *narHGI* es el responsable de la conversión de nitrato en nitrito, posteriormente NarK reduce el nitrito en óxido nítrico y finalmente el óxido nitroso es producido por NorBC. Estas reacciones fueron agregadas al modelo para conferir esta capacidad. La producción ATP del modelo iMA1267 fue comparada con datos experimentales de *P. denitrificans* debido a la falta de información de *P. chlororaphis* DSM 50083 y fueron expresados en porcentaje de ATP obtenido respecto al uso de nitrato como aceptor de electrones (tabla 6) (Koike et al. 1975).

Tabla 6. Comparación de rendimiento de ATP con diferentes aceptores de electrones

Aceptor de electrones	molATP/molA	% Respect NO_3^-	% Respect NO_3^- *
NO_3^-	2.8	100	100
NO_2^-	1.4	50	65
NO	0.7	25	35

*Datos experimentales (Koike et al. 1975).

Optimización del modelo para la producción de fenazinas

Se analizó *in silico* el metabolismo de *P. chlororaphis* DSM 50083 para encontrar la fuente de carbono que brindara el mayor rendimiento para la producción de PCN. Se probaron todas las fuentes de carbono capaces de sustentar el crecimiento y se comparó su rendimiento de PCN (figura 11). El mayor rendimiento fue obtenido usando ácidos grasos de longitud media y algunos dipéptidos. Entre los ácidos grasos encontramos el hexanoato, heptanoato, octanoato y nonanoato. Los dipéptidos leucil-leucina y alanil-leucina se encuentran dentro de las fuente de carbono con mejor rendimiento para la síntesis de PCN. Esto se debe a que estos compuestos pueden ser utilizados como fuente de carbono y nitrógeno. Los mejores rendimientos fueron obtenidos con los compuestos mencionados, sin embargo, esto no significa que sean los más adecuados para su implementación en un proceso de producción. La falta de evidencia experimental implica que estas fuentes de carbono deben ser probadas *in vivo* para evaluar sus velocidades de consumo y productividad de PCN. Sin embargo, el glicerol se encontró entre los compuestos que mostraron un mejor rendimiento y además es considerado una fuente de carbono de bajo costo por lo que su utilización en procesos a gran escala resulta más factible.
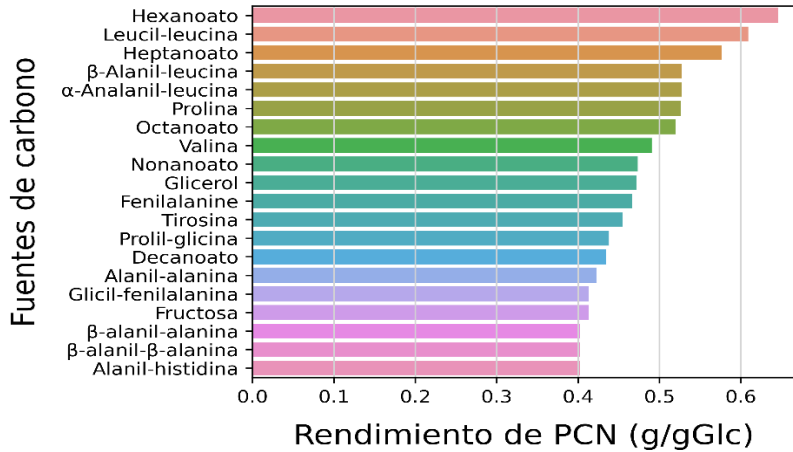
Figura 11. Rendimientos de PCN con diferentes fuentes de carbono. En el gráfico se muestran los 20 más altos rendimientos de PCN obtenidos mediante simulación utilizando todas las fuentes de carbono metabolizables por el modelo iMA1267.

La construcción de un modelo metabólico brinda la ventaja de acelerar la exploración del metabolismo bajo diferentes condiciones, siendo la sobreproducción de algún compuesto de interés uno de las aplicaciones más explotadas. En este caso decidimos analizar el metabolismo con la finalidad de encontrar modificaciones a la red metabólica que resultaran en una mejora de la producción de PCN usando glicerol y glucosa como substrato. Para esto se implementó el algoritmo de Barrido de Flujos basado en Flujo del Objetivo Forzado (BFFOF), el cual es una variante del algoritmo de Variabilidad de Flujos (Choi et al. 2010) (figura 12, figura 13).

Figura 12. Reacciones cuyo incremento de flujo mejora la producción de PCN. En azul obscuro se muestran las celdas cuyas reacciones deben incrementar su flujo en un factor de 10 o mayor, en azul claro entre 5 y 9, en verde entre 1.5 y 4
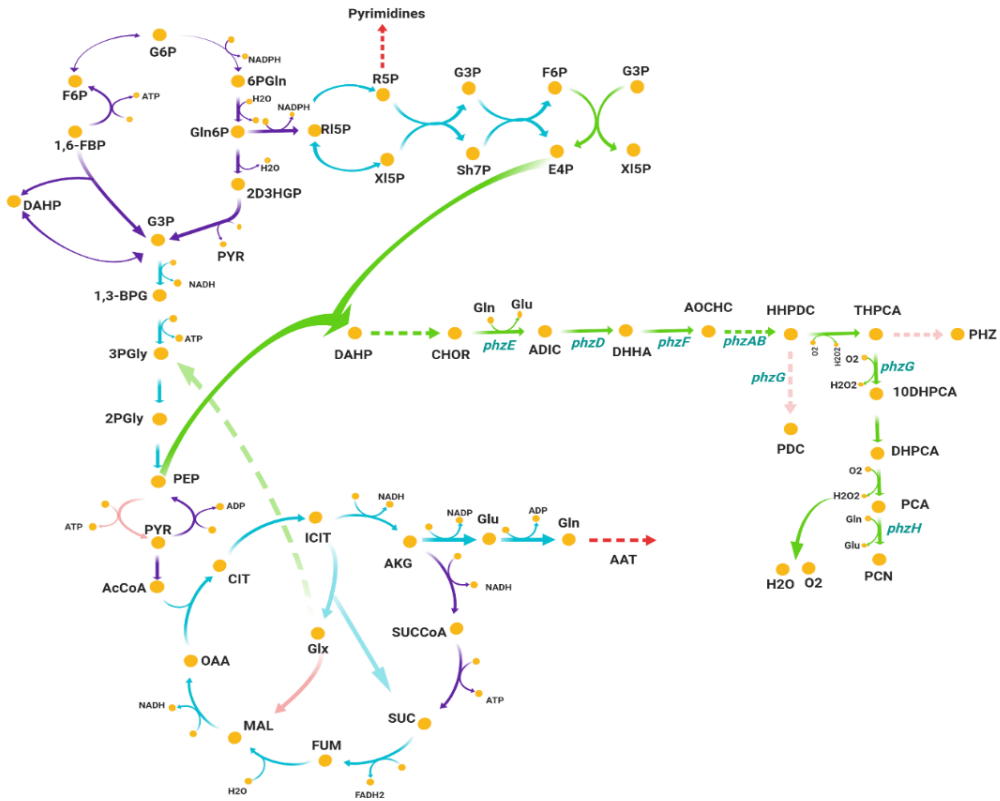
Figura 13. Modificaciones necesarias para optimizar la producción de PCN. En verde se muestran las reacciones enzimáticas que deben incrementar su flujo en un factor de 10 veces o más. En azul claro encontramos las reacciones enzimáticas que deben mantener un flujo entre 2 y 9 veces respecto al flujo inicial. En rojo se muestran las reacciones que deben reducir su flujo. En negro se muestran las abreviaciones de los metabolitos involucrados y en azul los genes pertenecientes al operón *phz*. Las líneas moradas representan las reacciones enzimáticas y las líneas punteadas representan dos o más reacciones consecutivas. G6P: glucosa-6-fosfato, F6P: fructosa-6-fosfato, 1,6-fructosa-bifosfato, DAHP: 3-Deoxi-arabino-heptulosonato-7-fosfato, CHOR: corismato, G3P: gliceraldehído-3-fosfato, 1,3-BPG: 1,3-bisfosfoglicerato, 3PGly: 3-fosfoglicerato, 2PGly: 2-fofoglicerato, PEP: fosfoenolpiruvato, PYR: piruvato, AcCoA: acetil-coenzima A, CIT: citrato, ICIT: isocitrato, AKG: α-cetoglutarato, SUCCoA: succinil-coenzima A, SUCC: succinato, FUM: fumarato, MAL: malato, OAA: oxaloacetato, Glu: glutamato, Gln: glutamina, 6PGln: 6-fosfo-D-glucono-1,5-lactona, Gln6P: gluconato-6-fosfato, 2D3HGP: 2-dehidro-3-deoxi-gluconato-6-fosfato, RL5P: ribulosa-5-fosfato, R5P: ribosa-5-fosfato, Xl5P: xilulosa-5-fosfato, Sh7P: sedoheptulosa-7-fosfato, E4P: eritrosa-4-fosfato, ADIC: 2-amino-2-deoxiisocorismato, DHHA: trans-2,3-dihidro-3-hidroxiantranilato, AOCHC: 6-amino-5-oxociclohex-2-ene-1-carboxilato, HHPDC: 1,4,5,5a,6,9-hexahidrofenazina-1,6-dicarboxilato, PHZ: fenazina, PDC: fenazina-1,6-dicarboxilato, THPCA: 1,4,5,10-tetrahidrofenazina-1-carboxilato, DHPCA: 5,10-dihidrofenazina-1-carboxilato , 10HPCA: dihidrofenazina-1-carboxilato, PCA: fenazina-1-carboxilato, PCN: fenazina-1-carboxiamida.

En ambas condiciones las modificaciones requeridas para incrementar diez veces la producción de PCN están relacionadas con las mismas reacciones enzimáticas. Entre las modificaciones más relevantes se encuentra el incremento del flujo a través de la vía de las

pentosas fosfato (VPF), específicamente en las reacciones trancetolsasa (TKT1/2) y transaldolasa (TALA) en un factor de 5. Por otro lado, reacciones que compiten por la ribosa-5-fosfato deben disminuir su flujo (Fosfomutasa, PPM, ribosa-1,5-bifosfoquinasa, R15BPK, ribosa-1-fosfoquinasa, R1PK). Adicionalmente, las reacciones ribosa-5-fosfato isomerasa (RPI y ribulosa-5-fosfato 3-epimerasa, RPE), las cuales producen ribosa-5-fosfato y D-ribosa-5-fosfato respectivamente deben incrementar su actividad. La eliminación de reacciones que consumen ribosa-5-fosfato e incrementar la actividad a través de la VPF promueven una mayor disponibilidad de eritrosa-4-fosfato. Como se puede observar en la figura 12, la eritrosa-4-fosfato es un compuesto esencial para iniciar la síntesis de compuestos aromáticos, necesarios para producir PCN.

Debido a los diferentes mecanismos de internalización de las fuentes de carbono utilizadas en la simulación, la condición con glucosa como substrato mostró un incremento en la fosfogluconato deshidrogenasa (GND). La reacción GND no es tan demandante con glicerol como substrato, ya que la vía de oxidación de este compuesto no requiere esta reacción enzimática. Bajo esta misma idea, la optimización con glicerol utiliza la gluconeogénesis para producir glucosa-6-fosfato y mantener activa la VPF. Junto con estos cambios, la piruvato quinasa (PYK) debe disminuir su flujo usando glucosa. La disminución de en la actividad de esta enzima incrementa la disponibilidad de fosfoenolpiruvato, el cual se condensa con eritrosa-4-fosfato en el primer paso de la vía del shikimato (VSHK). Este resultado se puede confirmar con reportes previos para la producción de compuestos aromáticos en otros organismos (Meza et al. 2012; Martínez et al. 2015). Esta modificación parece ser no esencial en la simulación con glicerol. Reacciones relacionadas directa o indirectamente con la parte baja del glucólisis muestran mayor actividad en ambas condiciones. Tal es el caso de la reacción glicerol deshidrogenasa (ALCD19) y glicerol quinasa (GLYK). Un mayor flujo en estas reacciones ayudaría a redirigir el flujo de carbono hacia fosfoenolpiruvato. La reacción fosfoglicerato quinasa requiere un incremento de flujo sólo cuando se usa glucosa.

La producción de PCN requiere el incremento del flujo dentro de la VSHK y la vía de las fenazinas. La glutamina es un metabolito importante en la síntesis de PCN, el cual es el donador de dos grupos amino en dos reacciones de la síntesis de fenazinas. Las reacciones

enzimáticas involucradas en la producción de intermediarios (α-cetoglutarato y glutamato) y la síntesis directa de glutamina requieren un incremento en el flujo. La reacción α-cetoglutarato deshidrogenasa (AKGDbr/AKGDBa) conducen a un incremento en la disponibilidad de α-cetoglutarato para la síntesis de glutamato. Este aminoácido es convertido a glutamina por la glutamina sintetasa (GLNS) en un paso subsecuente. La disponibilidad de glutamina se vio incrementada por una disminución de flujo a través de la valina y leucina transaminasas que usan a este metabolito como donador de grupos amino.

La simulación sugiere un incremento de flujo en el ciclo de los ácidos tricarboxilicos (TCA) y la producción de succinato a través de la isocitrato liasa (ICL). Este cambio está acompañado de una mayor incorporación de glioxilato al metabolismo central en forma de 3-fosfoglicerato. En este caso el glioxilato es convertido a 2-hidroxi-3-oxopropanoato por la glioxilato carboligasa (GLXCL) y posteriormente convertido a hidroxipiruvato por la hidroxipiruvato isomerasa (HPYRI). Finalmente, el hidroxipiruvato es convertido a glicerol-3-fosfato por las hidroxipiruvato reductasa (HPYRRy /GLYD) y glicerol quinasa (GLYK) secuencialmente. Un resultado similar se obtiene con la simulación usando glicerol como fuente de carbono.

La producción de PCN resulta en una alta producción de peróxido de hidrógeno (H_2O_2) independientemente de la fuente de carbono utilizada. Es sabido que las fenazinas están relacionadas con el mecanismo celular para sobrellevar el estrés oxidativo bajo diferentes ambientes (Girard et al. 2011; Xie et al. 2013). Este mecanismo promueve la degradación de fenazinas para la reducir la concentración de radicales libres, afectando la concentración de las primeras. En la simulación realizada, el modelo tiende a combatir con la producción del H_2O_2 con un incremento de la actividad catalasa (CAT). Con el propósito de entender el comportamiento del H_2O_2 dentro del metabolismo se realizó un análisis de fase plana para observar la variabilidad de los flujos mientras se incrementa la producción de PCN. En la figura 14 se observa como los límites superiores e inferiores del flujo del H_2O_2 cambian y reducen su rango de variabilidad debido a la producción de PCN. Este comportamiento puede tener implicaciones en el fenotipo debido al estrés oxidativo.
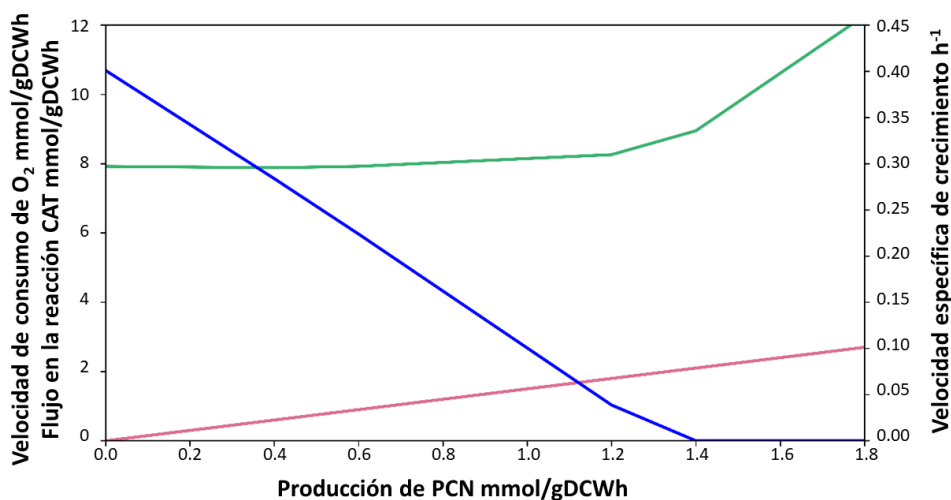
Figura 14. Análisis de fase plana para la actividad catalasa y producción de PCN. El grafico muestra el incremento de la velocidad de consumo de oxígeno (verde) resultado del incremento en la producción de PCN. El mismo efecto se muestra para la reacción de la catalasa (CAT, rojo). La velocidad de crecimiento (azul) es disminuida por el redireccionamiento de la fuente de carbono hacia la síntesis de PCN.

Recientemente y de forma paralela a este proyecto, se han reportado estrategias similares a las aquí propuestas para mejorar la producción de fenazinas. Un estudio demostró el estrés oxidativo que sufre *P. chlororaphis* al producir fenazinas, algo que no había sido tomado en consideración desde el punto de vista de producción (Yue et al. 2020). En dicho reporte se observa que la adición de un agente reductor, como ditiotreitol, es suficiente para incrementar la producción de fenazinas. Esto nos habla de la importancia del estudio no sólo de la vía metabólica estudiada, sino que es necesario estudiar el contexto celular, el cual puede ser estudiado utilizando modelos metabólicos como el que aquí se presenta. El análisis del modelo iMA1267 fue una herramienta poderosa para inferir el estrés oxidativo durante la producción de fenazinas, análisis que puede ser extrapolado a cualquier vía metabólica.

CONCLUSIONES

El uso de información metabólica de organismos cercanos a nuestro objeto de estudio brinda un sólido punto de partida para la reconstrucción de modelos metabólicos. Además, acelera el proceso de curación y aporta confianza a las predicciones. Este trabajo muestra que es posible realizar análisis de flujos metabólicos de organismos no modelo con la estrategia aquí descrita, agilizando el proceso de adquisición de información.

Las características de *P. chlororaphis* DSM 50083 la colocan como un buen modelo de producción de metabolitos. Esta bacteria posee un metabolismo robusto similar al de otras *Pseudomonas,* un alto rendimiento de NADPH y resistencia a acumulación de metabolitos secundarios. Estas características en combinación con el amplio repertorio de compuestos metabolizables permiten que este organismo pueda ser utilizado para diferentes propósitos biotecnológicos.

La simulación del metabolismo de *P. chlororaphis* DSM 50083 brinda un panorama amplio de su potencial metabólico. El modelado de flujo metabólicos permite analizar el crecimiento y producción de metabolitos con un enfoque integrativo. Adicionalmente, este análisis proporciona una herramienta de estudio para optimizar la producción de metabolitos secundarios, sugiriendo modificaciones genéticas o fuentes de carbono con mejores rendimientos.

El estrés oxidativo tiene un papel muy importante en la producción de PCN. La producción de peróxido de hidrógeno está fuertemente ligada a la producción de fenazinas, lo que podría provocar un ambiente celular oxidativo. Esta condición predicha es un blanco de mejora para las cepas productoras de esta familia de compuestos.

La optimización metabólica utiliza rutas más económicas que los mecanismos *in vivo* contra el estrés oxidativo, por lo que su implementación experimental podría ayudar a mejorar el fenotipo en condiciones de producción.

PERSPECTIVAS

Construcción de una cepa que sobre exprese los genes para la síntesis de PCN y que contenga las modificaciones aquí sugeridas para incrementar la producción este compuesto. Adicionalmente, deberá sobre expresar el gen de la catalasa para analizar el efecto de ésta sobre la producción de PCN.

Estrategias alternativas para incrementar la actividad de la catalasa, la disponibilidad de NADPH podrían ayudar a combatir el estrés oxidativo y mejorar el desempeño celular durante la producción de PCN.

Complementar el modelado de flujos metabólicos con análisis transcriptómicos y proteómicos para implementar reducir el espacio de soluciones incrementará el poder de predicción del modelo bajo condiciones específicas.

La parametrización del modelo es muy importante, por lo que se requiere de una evaluación del modelo bajo diferentes condiciones para mejorar la calidad del mimos.

Incorporar los resultados el análisis de flujos metabólicos a modelos de crecimiento celular fuera del estado estacionario.

REFERENCIAS

Adamczyk PA, Reed JL (2017) *Escherichia coli* as a model organism for systems metabolic engineering. Curr Opin Syst Biol. 6:80–88 . https://doi.org/10.1016/j.coisb.2017.11.001

Agren R, Liu L, Shoaie S, Vongsangnak W, Nookaew I, Nielsen J (2013) The RAVEN toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum*. PLoS Comput Biol. 9:e1002980 . https://doi.org/10.1371/journal.pcbi.1002980

Azam MW, Khan AU (2019) Updates on the pathogenicity status of *Pseudomonas aeruginosa*. Drug Discov Today. 24:350–359 . https://doi.org/10.1016/j.drudis.2018.07.003

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O (2008) The RAST Server: Rapid annotations using subsystems technology. BMC Genomics. 9:75 . https://doi.org/10.1186/1471-2164-9-75

Babaei P, Ghasemi-Kahrizsangi T, Marashi S (2014) Modeling the differences in biochemical capabilities of *Pseudomonas* species by flux balance analysis: how good are genome-scale metabolic networks at predicting the differences?. Sci World J. 2014:1–11 . https://doi.org/10.1155/2014/416289

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin A V., Sirotkin A V., Vyahhi N, Tesler G, Alekseyev M a., Pevzner P a. (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19:455–477 . https://doi.org/10.1089/cmb.2012.0021

Bartell JA, Blazier AS, Yen P, Thøgersen JC, Jelsbak L, Goldberg JB, Papin JA (2017) Reconstruction of the metabolic network of *Pseudomonas aeruginos*a to interrogate virulence factor synthesis. Nat Commun. 8: . https://doi.org/10.1038/ncomms14631

Bateman A (2004) The Pfam protein families database. Nucleic Acids Res. 32:138D – 141 . https://doi.org/10.1093/nar/gkh121

Becker J, Zelder O, Häfner S, Schröder H, Wittmann C (2011) From zero to hero—design-based systems metabolic engineering of *Corynebacterium glutamicum* for l-lysine production. Metab

Eng. 13:159–168 . https://doi.org/10.1016/j.ymben.2011.01.003

Bertani G (1951) Studies on lysogenesis I. J Bacteriol. 62:293–300 .
https://doi.org/10.1128/jb.62.3.293-300.1951

Bilal M, Guo S, Iqbal HMN, Hu H, Wang W, Zhang X (2017) Engineering *Pseudomonas* for
phenazine biosynthesis, regulation, and biotechnological applications: a review. World J
Microbiol Biotechnol. 33:1–11 . https://doi.org/10.1007/s11274-017-2356-9

Borgos SE, Bordel S, Sletta H, Ertesvåg H, Jakobsen Ø, Bruheim P, Ellingsen TE, Nielsen J, Valla
S (2013) Mapping global effects of the anti-sigma factor MucA in *Pseudomonas fluorescens*
SBW25 through genome-scale metabolic modeling. BMC Syst Biol. 7:19.
https://doi.org/10.1186/1752-0509-7-19

Burgard AP, Pharkya P, Maranas CD (2003) Optknock: A bilevel programming framework for
identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng.
84:647–657. https://doi.org/10.1002/bit.10803

Burr SE, Gobeli S, Kuhnert P, Goldschmidt-Clermont E, Frey J (2010) *Pseudomonas chlororaphis*
subsp. piscium subsp. nov., isolated from freshwater fish. Int J Syst Evol Microbiol. 60:2753–
2757. https://doi.org/10.1099/ijs.0.011692-0

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009)
BLAST+: architecture and applications. BMC Bioinformatics. 10:421 .
https://doi.org/10.1186/1471-2105-10-421

Catone M V., Ruiz JA, Castellanos M, Segura D, Espin G, López NI (2014) High
polyhydroxybutyrate production in *Pseudomonas* extremaustralis is associated with
differential expression of horizontally acquired and core genome polyhydroxyalkanoate
synthase genes. PLoS One. 9(6). https://doi.org/10.1371/journal.pone.0098873

Cavinato L, Genise E, Luly FR, Di Domenico EG, Del Porto P, Ascenzioni F (2020) Escaping the
phagocytic oxidative burst: The role of sodb in the survival of *Pseudomonas aeruginosa*
within macrophages. Front Microbiol. 11:1–12 . https://doi.org/10.3389/fmicb.2020.00326

Chan PP, Lowe TM (2019) tRNAscan-SE: Searching for trna genes in genomic sequences. In:
Kollmar M (ed) Methods Mol Biol. Springer New York, New York, NY, pp 1–14

Choi HS, Lee SY, Kim TY, Woo HM (2010) In silico identification of gene amplification targets
for improvement of lycopene production. Appl Environ Microbiol. 76:3097–3105.

https://doi.org/10.1128/AEM.00115-10

Chubukov V, Mukhopadhyay A, Petzold CJ, Keasling JD, Martín HG (2018) Synthetic and systems biology for microbial production of commodity chemicals. Syst Biol Appl. 2:1–11 . https://doi.org/10.1038/npjsba.2016.9

Ciskanik LM, Wilczek JM, Fallon RD (1995) Purification and characterization of an enantioselective amidase from *Pseudomonas chlororaphis* B23. Appl Environ Microbiol. 61:998–1003. 10.1128/aem.61.3.998-1003.1995

Delcher AL, Harmon D, Kasif S, White O, Salzberg SL (1999) Improved microbial gene identification with GLIMMER. Nucleic Acids Res. 27:4636–41 . https://doi.org/10.1093/nar/27.23.4636

Dimkpa CO, Zeng J, McLean JE, Britt DW, Zhan J, Anderson AJ (2012) Production of indole-3-acetic acid via the indole-3-acetamide pathway in the plant-beneficial bacterium *Pseudomonas chlororaphis* O6 is inhibited by ZnO nanoparticles but enhanced by CuO nanoparticles. Appl Environ Microbiol. 78:1404–1410 . https://doi.org/10.1128/AEM.07424-11

Dong H, Zhang D (2014) Current development in genetic engineering strategies of *Bacillus* species. Microb Cell Fact. 13:1–11 . https://doi.org/10.1186/1475-2859-13-63

Dunphy LJ, Papin JA (2018) Biomedical applications of genome-scale metabolic network reconstructions of human pathogens. Curr Opin Biotechnol. 51:70–79 . https://doi.org/10.1016/j.copbio.2017.11.014

Ebrahim A, Lerman JA, Palsson BO, Hyduke DR (2013) COBRApy: Constraints-based reconstruction and analysis for python. BMC Syst Biol. 7:74 . https://doi.org/10.1186/1752-0509-7-74

EFSA Panel on BIOHAZ (2009) Scientific opinion on the maintenance of the list of QPS microorganisms intentionally added to food or feed (2009 update). EFSA J 7:1431 . https://doi.org/10.2903/j.efsa.2009.1431

Ekseth OK, Kuiper M, Mironov V (2014) orthAgogue: an agile tool for the rapid prediction of orthology relations. Bioinformatics. 30:734–6 . https://doi.org/10.1093/bioinformatics/btt582

Famili I, Forster J, Nielsen J, Palsson BO (2003) *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. Proc Natl Acad Sci U S A. 100:13134–9 . https://doi.org/10.1073/pnas.2235812100

Fatma Z, Schultz JC, Zhao H (2020) Recent advances in domesticating non-model microorganisms. Biotechnol Prog. 36(5). https://doi.org/10.1002/btpr.3008

Feist AM, Scholten JCM, Palsson BØ, Brockman FJ, Ideker T (2006) Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosarcina barkeri*. Mol Syst Biol. 2:1–14 . https://doi.org/10.1038/msb4100046

Garcia-Ochoa F, Gomez E, Santos VE, Merchuk JC (2010) Oxygen uptake rate in microbial processes: An overview. Biochem Eng J. 49:289–307 . https://doi.org/10.1016/j.bej.2010.01.011

Ghosh S, Banerjee UC (2015) Generation of *aroE* overexpression mutant of *Bacillus megaterium* for the production of shikimic acid. Microb Cell Fact. 14:1–9 . https://doi.org/10.1186/s12934-015-0251-3

Girard G, Rigali S (2011) Role of the phenazine-inducing protein Pip in stress resistance of *Pseudomonas chlororaphis.* Microbiology. 157:398–407 . https://doi.org/10.1099/mic.0.043075-0

Gomila M, Peña A, Mulet M, Lalucat J, García-Valdés E (2015) Phylogenomics and systematics in *Pseudomonas*. Front Microbiol. 6. https://doi.org/10.3389/fmicb.2015.00214

Gross H, Loper JE (2009) Genomics of secondary metabolite production by *Pseudomonas* spp. Nat Prod Rep. 26:1408–1446 . https://doi.org/10.1039/b817075b

Gunther NW, Nuñez A, Fett W, Solaiman DKY (2005) Production of rhamnolipids by *Pseudomonas chlororaphis*, a nonpathogenic bacterium. Appl Environ Microbiol. 71:2288–2293. https://doi.org/10.1128/AEM.71.5.2288-2293.2005

Gupta SK, Srivastava SK, Sharma A, Nalage VHH, Salvi D, Kushwaha H, Chitnis NB, Shukla P (2017) Metabolic engineering of CHO cells for the development of a robust protein production platform. PLoS One. 12:1–23 . https://doi.org/10.1371/journal.pone.0181455

Guzmán GI, Utrilla J, Nurk S, Brunk E, Monk JM, Ebrahim A, Palsson BO, Feist AM (2015) Model-driven discovery of underground metabolic functions in *Escherichia coli*. Proc Natl Acad Sci. 112:929–934 . https://doi.org/10.1073/pnas.1414218112

Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, Beck E (2012) Tigrfams and genome properties in 2013. Nucleic Acids Res. 41:D387–D395 . https://doi.org/10.1093/nar/gks1234

Healey A, Furtado A, Cooper T, Henry RJ (2014) Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. Plant Methods 10:21 . https://doi.org/10.1186/1746-4811-10-21

Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, Haraldsdóttir HS, Wachowiak J, Keating SM, Vlasov V, Magnusdóttir S, Ng CY, Preciat G, Žagare A, Chan SHJ, Aurich MK, Clancy CM, Modamio J, Sauls JT, Noronha A, Bordbar A, Cousins B, El Assal DC, Valcarcel L V., Apaolaza I, Ghaderi S, Ahookhosh M, Ben Guebila M, Kostromins A, Sompairac N, Le HM, Ma D, Sun Y, Wang L, Yurkovich JT, Oliveira MAP, Vuong PT, El Assal LP, Kuperstein I, Zinovyev A, Hinton HS, Bryant WA, Aragón Artacho FJ, Planes FJ, Stalidzans E, Maass A, Vempala S, Hucka M, Saunders MA, Maranas CD, Lewis NE, Sauter T, Palsson B, Thiele I, Fleming RMT (2019) Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. Nat Protoc. 14:639–702 . https://doi.org/10.1038/s41596-018-0098-2

Höfte M, De Vos P (2007) Plant pathogenic *Pseudomonas* species. In: Gnanamanickam SS (ed) Plant-Associated Bacteria. Springer Netherlands, Dordrecht, pp 507–533

Huang HT, English AR. March (1966). Production of 6-aminopenicillanic

acid. US patent 3239427 A.

Hu H, Li Y, Liu K, Zhao J, Wang W, Zhang X (2017) Production of trans-2,3-dihydro-3-hydroxyanthranilic acid by engineered *Pseudomonas chlororaphis* GP72. Appl Microbiol Biotechnol. 101:6607–6613 . https://doi.org/10.1007/s00253-017-8408-0

Huccetogullari D, Luo ZW, Lee SY (2019) Metabolic engineering of microorganisms for production of aromatic compounds. Microb Cell. Fact 18:41 . https://doi.org/10.1186/s12934-019-1090-4

Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 11:119 . https://doi.org/10.1186/1471-2105-11-119

Jiménez JI, Miñambres B, Luis J, Díaz E (2002) Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. Environ Microbiol. 4:824–841. 10.1046/j.1462-2920.2002.00370.x

Kavšček M, Stražar M, Curk T, Natter K, Petrovič U (2015) Yeast as a cell factory: current state

and perspectives. Microb Cell Fact. 14:94 . https://doi.org/10.1186/s12934-015-0281-x

King, Z. A., Lloyd, C. J., Feist, A. M., & Palsson, B. O. (2015). Next-generation genome-scale models for metabolic engineering. Current opinion in biotechnology. 35, 23-29.

King ZA, Lu J, Dräger A, Miller P, Federowicz S, Lerman JA, Ebrahim A, Palsson BO, Lewis NE (2016) BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Res. 44:D515–D522 . https://doi.org/10.1093/nar/gkv1049

Klamt S, Müller S, Regensburger G, Zanghellini J (2018) A mathematical framework for yield (vs. rate) optimization in constraint-based modeling and applications in metabolic engineering. Metab Eng. 47:153–169 . https://doi.org/10.1016/j.ymben.2018.02.001

Koike I, Hattori A (1975) Energy yield of denitrification: An estimate from growth yield in continuous cultures of *Pseudomonas denitrificans* under nitrate, nitrite and nitrous oxide limited conditions. J Gen Microbiol. 88:11–19 . https://doi.org/10.1099/00221287-88-1-11

Kusumawardhani H, Hosseini R, de Winde JH (2018) Solvent tolerance in bacteria: Fulfilling the promise of the biotech era?. Trends Biotechnol. 36:1025–1039 . https://doi.org/10.1016/j.tibtech.2018.04.007

Lee WH, Kim MD, Jin YS, Seo JH (2013) Engineering of NADPH regenerators in *Escherichia coli* for enhanced biotransformation. Appl Microbiol Biotechnol. 97:2761–2772 . https://doi.org/10.1007/s00253-013-4750-z

Li J, Rong L, Zhao Y, Li S, Zhang C, Xiao D, Foo JL, Yu A (2020) Next-generation metabolic engineering of non-conventional microbial cell factories for carboxylic acid platform chemicals. Biotechnol Adv. 43:107605 . https://doi.org/10.1016/j.biotechadv.2020.107605

Liao YC, Chen JCY, Tsai MH, Tang YH, Chen FC, Hsiung CA (2011) MrBac: A web server for draft metabolic network reconstructions for bacteria. Bioeng Bugs. 2:284–287 . https://doi.org/10.4161/bbug.2.5.16113

Liu Z, Moradi H, Shi S, Darvishi F (2021) Yeasts as microbial cell factories for sustainable production of biofuels. Renew Sustain Energy Rev. 143:110907 . https://doi.org/10.1016/j.rser.2021.110907

Lobb B, Tremblay BJ-M, Moreno-Hagelsieb G, Doxey AC (2020) An assessment of genome annotation coverage across the bacterial tree of life. Microb Genomics. 6: 3. https://doi.org/10.1099/mgen.0.000341

Löbs A-K, Schwartz C, Wheeldon I (2017) Genome and metabolic engineering in non-conventional yeasts: Current advances and applications. Synth Syst Biotechnol. 2:198–207. https://doi.org/10.1016/j.synbio.2017.08.002

Loeschcke A, Thies S (2015) *Pseudomonas putida* a versatile host for the production of natural products. Appl Microbiol Biotechnol. 99:6197–6214 . https://doi.org/10.1007/s00253-015-6745-4

Löwe H, Schmauder L, Hobmeier K, Kremling A, Pflüger-Grau K (2017) Metabolic engineering to expand the substrate spectrum of *Pseudomonas putida* toward sucrose. Microbiologyopen. 6:1–9 . https://doi.org/10.1002/mbo3.473

Löwe H, Sinner P, Kremling A, Pflüger-Grau K (2018) Engineering sucrose metabolism in *Pseudomonas putida* highlights the importance of porins. Microb Biotechnol. 0:1–10 . https://doi.org/10.1111/1751-7915.13283

Magnúsdóttir S, Heinken A, Kutt L, Ravcheev DA, Bauer E, Noronha A, Greenhalgh K, Jäger C, Baginska J, Wilmes P, Fleming RMT, Thiele I (2017) Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. Nat Biotechnol. 35:81–89 . https://doi.org/10.1038/nbt.3703

Maniatis T., Fritsch E.F., Sambrook J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Martínez JA, Bolívar F, Escalante A (2015) Shikimic acid production in *Escherichia coli*: From classical metabolic engineering strategies to omics applied to improve its production. Front Bioeng Biotechnol. 3:1–16 . https://doi.org/10.3389/fbioe.2015.00145

McCully LM, Bitzer AS, Spence CA, Bais HP, Silby MW (2014) Draft genome sequence of rice isolate *Pseudomonas chlororaphis* EA105. Genome Announc. 2:e01342-14-e01342-14 . https://doi.org/10.1128/genomeA.01342-14

Medema MH, Blin K, Cimermancic P, De Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R (2011) AntiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 39:339–346 . https://doi.org/10.1093/nar/gkr466

Meza E, Becker J, Bolivar F, Gosset G, Wittmann C (2012) Consequences of phosphoenolpyruvate:sugar phosphotranferase system and pyruvate kinase isozymes

inactivation in central carbon metabolism flux distribution in *Escherichia coli*. Microb Cell Fact. 11:1–13 . https://doi.org/10.1186/1475-2859-11-127

Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, Orth JD, Feist AM, Palsson BO (2013) Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. Proc Natl Acad Sci. 110:20338–20343 . https://doi.org/10.1073/pnas.1307797110

Moreno-Avitia F, Lozano L, Utrilla J, Bolívar F, Escalante A (2017) Draft genome sequence of *Pseudomonas chlororaphis* ATCC 9446, a nonpathogenic bacterium with bioremediation and industrial potential. Genome Announc. 5:1–2 . https://doi.org/10.1128/genomeA.00474-17

Nelson KE, Weinel C, Paulsen IT, Dodson RJ, Hilbert H, Santos VAPM, Fouts DE, Gill SR, Pop M, Holmes M, Brinkac L, Beanan M, Deboy RT, Daugherty S, Kolonay J, Madupu R, Nelson W, White O, Peterson J, Khouri H, Hance I, Lee PC, Holtzapple E, Scanlan D, Tran K, Moazzez A, Utterback T, Rizzo M, Lee K, Kosack D, Moestl D, Wedler H, Lauber J, Stjepandic D, Hoheisel J, Straetz M, Heim S, Kiewitz C, Eisen J, Timmis KN, Düsterhöft A, Tümmler B, Fraser CM (2002) Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. Environ Microbiol. 4:799–808

Nichio BTL, Marchaukoski JN, Raittz RT (2017) New tools in orthology analysis: A brief review of promising perspectives. Front Genet 8:1–12 . https://doi.org/10.3389/fgene.2017.00165

Nikel PI, Fuhrer T, Chavarría M, Sánchez-Pascuala A, Sauer U, de Lorenzo V (2020) Redox stress reshapes carbon fluxes of *Pseudomonas putida* for cytosolic glucose oxidation and NADPH generation. bioRxiv. 1–31 . https://doi.org/10.1101/2020.06.13.149542

Nikel PI, Martínez-García E, de Lorenzo V (2014) Biotechnological domestication of pseudomonads using synthetic biology. Nat Rev Microbiol. 12:368–379 . https://doi.org/10.1038/nrmicro3253

Nishimori E, Kita-Tsukamoto K, Wakabayashi H. (2000) *Pseudomonas plecoglossicida* sp. nov., the causative agent of bacterial haemorrhagic ascites of ayu, *Plecoglossus altivelis*. Int J Syst Evol Microbiol. 50 1:83-89. doi: 10.1099/00207713-50-1-83. PMID: 10826790.

Nogales J (2014) A Practical Protocol for Genome-Scale Metabolic Reconstructions. In: McGenity T., Timmis K., Nogales B. (eds) Hydrocarbon and Lipid Microbiology Protocols. Springer Protocols Handbooks. Springer, Berlin, Heidelberg. https://doi.org/10.1007/8623_2014_12

Nogales J, Mueller J, Gudmundsson S, Canalejo FJ, Duque E, Monk J, Feist AM, Ramos JL, Niu W, Palsson BO (2020) High-quality genome-scale metabolic modelling of P*seudomonas putida* highlights its broad metabolic capabilities. Environ Microbiol. 22:255–269 . https://doi.org/10.1111/1462-2920.14843

Nogales J, Palsson BØ, Thiele I (2008) A genome-scale metabolic reconstruction of *Pseudomonas putida* KT2440: iJN746 as a cell factory. BMC Syst Biol. 2:79 . https://doi.org/10.1186/1752-0509-2-79

Norsigian CJ, Pusarla N, McConn JL, Yurkovich JT, Dräger A, Palsson BO, King Z (2019) BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. Nucleic Acids Res. 1–5 . https://doi.org/10.1093/nar/gkz1054

Oberhardt MA, Palsson BØ, Papin JA (2009) Applications of genome-scale metabolic reconstructions. Mol Syst Biol. 5:1–15 . https://doi.org/10.1038/msb.2009.77

Oberhardt MA, Puchałka J, dos Santos VAPM, Papin JA (2011) Reconciliation of genome-scale metabolic reconstructions for comparative systems analysis. PLoS Comput Biol. 7: . https://doi.org/10.1371/journal.pcbi.1001116

Oberhardt MA, Puchałka J, Fryer KE, Martins Dos Santos VAP, Papin JA (2008) Genome-scale metabolic network analysis of the opportunistic pathogen *Pseudomonas aeruginosa* PAO1. J Bacteriol. 190:2790–2803 . https://doi.org/10.1128/JB.01583-07

Ohnishi, S. Mitsuhashi, M. Hayashi, J (2002) A novel methodology employing *Corynebacterium glutamicum* genome information to generate a new L -lysine-producing mutant. Appl Microbiol Biotechnol. 58:217–223 . https://doi.org/10.1007/s00253-001-0883-6

Orth JD, Conrad TM, Na J, Lerman J a, Nam H, Feist AM, Palsson BØ (2011) A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism. Mol Syst Biol. 7:535 . https://doi.org/10.1038/msb.2011.65

Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R (2014) The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res. 42:206–214 . https://doi.org/10.1093/nar/gkt1226

Özcan E, Selvi SS, Nikerel E, Teusink B, Toksoy Öner E, Çakır T (2019) A genome-scale metabolic network of the aroma bacterium *Leuconostoc mesenteroides* subsp. cremoris. Appl

Microbiol Biotechnol 103:3153–3165 . https://doi.org/10.1007/s00253-019-09630-4

Palleroni NJ, Moore ERB (2004) Taxonomy of pseudomonads: Experimental approaches. In:
Ramos J-L (ed) Pseudomonas: Volume 1 Genomics, Life Style and Molecular Architecture.
Springer US, Boston, MA, pp 3–44

Palsson B (2009) Metabolic systems biology. FEBS Lett. 583:3900–3904 .
https://doi.org/10.1016/j.febslet.2009.09.031

Patil KR, Rocha I, Förster J, Nielsen J (2005) Evolutionary programming as a platform for in silico
metabolic engineering. BMC Bioinformatics. 6:1–12 . https://doi.org/10.1186/1471-2105-6-
308

Peña DA, Gasser B, Zanghellini J, Steiger MG, Mattanovich D (2018) Metabolic engineering of
Pichia pastoris. Metab Eng. 50:2–15 . https://doi.org/10.1016/j.ymben.2018.04.017

Peng H, Tan J, Bilal M, Wang W, Hu H, Zhang X (2018) Enhanced biosynthesis of phenazine-1-
carboxamide by *Pseudomonas chlororaphis* strains using statistical experimental designs.
World J Microbiol Biotechnol. 34:0 . https://doi.org/10.1007/s11274-018-2501-0

Peng Y, Leung HCM, Yiu SM, Chin FYL (2012) IDBA-UD: A de novo assembler for single-cell
and metagenomic sequencing data with highly uneven depth. Bioinformatics. 28:1420–1428 .
https://doi.org/10.1093/bioinformatics/bts174

Pharkya P, Burgard AP, Maranas CD (2004) OptStrain: a computational framework for redesign of
microbial production systems. Genome Res. 14(11):2367-76. doi: 10.1101/gr.2872004.

Pierson LS, Pierson EA (2010) Metabolism and function of phenazines in bacteria: impacts on the
behavior of bacteria in the environment and biotechnological processes. Appl Microbiol
Biotechnol. 86:1659–1670 . https://doi.org/10.1007/s00253-010-2509-3

Pizarro-Tobías P, De Genève J, Fernández M (2008) Mining GOLD and new model organisms in
biotechnology. Microb Biotechnol. 1:273–274 . https://doi.org/10.1111/j.1751-
7915.2008.00039.x

Puchałka J, Oberhardt MA, Godinho M, Bielecka A, Regenhardt D, Timmis KN, Papin JA, Martins
dos Santos VAP (2008) Genome-scale reconstruction and analysis of the *Pseudomonas putida*
KT2440 metabolic network facilitates applications in biotechnology. PLoS Comput Biol.
4:e1000210 . https://doi.org/10.1371/journal.pcbi.1000210

Raghunathan A, Reed J, Shin S, Palsson B, Daefler S (2009) Constraint-based analysis of metabolic capacity of *Salmonella typhimurium* during host-pathogen interaction. BMC Syst Biol 3:1–16 . https://doi.org/10.1186/1752-0509-3-38

Ramos JL, Duque E, Hai ALI (1995) Isolation and expansion of the catabolic potential of a pseudomonas putida strain able to grow in the presence of high concentrations of aromatic hydrocarbons. J Bacteriol. 177:3911–3916. 10.1128/jb.177.14.3911-3916.1995

Ranganathan S, Suthers PF, Maranas CD (2010) OptForce: An optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PLoS Comput Biol. 6:e1000744 . https://doi.org/10.1371/journal.pcbi.1000744

Rhoads A, Au KF (2015) PacBio sequencing and its applications. Genomics, Proteomics Bioinforma. 13:278–289 . https://doi.org/10.1016/j.gpb.2015.08.002

Ruffner B, Péchy-Tarr M, Höfte M, Bloemberg G, Grunder J, Keel C, Maurhofer M (2015) Evolutionary patchwork of an insecticidal toxin shared between plant-associated pseudomonads and the insect pathogens *Photorhabdus and Xenorhabdus*. BMC Genomics. 16:609 . https://doi.org/10.1186/s12864-015-1763-2

Sabarinathan D, Chandrika SP, Venkatraman P, Easwaran M, Sureka CS, Preethi K (2018) Production of polyhydroxybutyrate (PHB) from *Pseudomonas plecoglossicida* and its application towards cancer detection. Informatics Med Unlocked. 11:61–67 . https://doi.org/10.1016/j.imu.2018.04.009

Sanchez-Garcia L, Martín L, Mangues R, Ferrer-Miralles N, Vázquez E, Villaverde A (2016) Recombinant pharmaceuticals from microbial cells: a 2015 update. Microb Cell Fact. 15:33 . https://doi.org/10.1186/s12934-016-0437-3

Sardessai Y, Bhosle S (2002) Tolerance of bacteria to organic solvents. Res Microbiol. 153:263–8. https://doi.org/10.1016/S0923-2508(02)01319-0

Sardessai YN, Bhosle S (2004) Industrial potential of organic solvent tolerant bacteria. Biotechnol Prog 20:655–660 . https://doi.org/10.1021/bp0200595

Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S, Kang J, Hyduke DR, Palsson BØ (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc. 6:1290–1307 . https://doi.org/10.1038/nprot.2011.308

Schipper LA, Robertson WD, Gold AJ, Jaynes DB, Cameron SC (2010) Denitrifying bioreactors-An approach for reducing nitrate loads to receiving waters. Ecol Eng. 36:1532–1543 . https://doi.org/10.1016/j.ecoleng.2010.04.008

Seemann T (2014) Prokka: Rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069 . https://doi.org/10.1093/bioinformatics/btu153

Segre D, Vitkup D, Church GM (2002) Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci 99:15112–15117 . https://doi.org/10.1073/pnas.232349399

Sentchilo VS, Perebituk AN, Zehnder AJ, van der Meer JR (2000) Molecular diversity of plasmids bearing genes that encode toluene and xylene metabolism in *Pseudomonas* strains isolated from different contaminated sites in Belarus. Appl Environ Microbiol. 66:2842–52 . https://doi.org/10.1128/genomeA.01225-13

Shen X, Hu H, Peng H, Wang W, Zhang X (2013) Comparative genomic analysis of four representative plant growth-promoting rhizobacteria in *Pseudomonas*. BMC Genomics 14:271 . https://doi.org/10.1186/1471-2164-14-271

Sonoyama T, Kageyama B, Honjo T. November 1975. Process for pro ducing 2-keto-L-gulonic acid. US patent 3922194 A.

Steen JA, Bohlke N, Vickers CE, Nielsen LK. (2014) The trehalose phosphotransferase system (PTS) in *E. coli* W can transport low levels of sucrose that are sufficient to facilitate induction of the *csc* sucrose catabolism operon. PLoS One.9(2):e88688. doi:10.1371/journal.pone.0088688

Stephanopoulos G (1999) Metabolic Fluxes and Metabolic Engineering. Metab Eng. 1:1–11 . https://doi.org/10.1006/mben.1998.0101

Swainston N, Smallbone K, Hefzi H, Dobson PD, Brewer J, Hanscho M, Zielinski DC, Ang KS, Gardiner NJ, Gutierrez JM, Kyriakopoulos S, Lakshmanan M, Li S, Liu JK, Martínez VS, Orellana CA, Quek L-E, Thomas A, Zanghellini J, Borth N, Lee D-Y, Nielsen LK, Kell DB, Lewis NE, Mendes P (2016) Recon 2.2: from reconstruction to model of human metabolism. Metabolomics. 12:109 . https://doi.org/10.1007/s11306-016-1051-4

Tatusova T, Dicuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J (2016) NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 44:6614–6624 . https://doi.org/10.1093/nar/gkw569

Thiele I, Palsson BØ (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc. 5:93–121 . https://doi.org/10.1038/nprot.2009.203

Town J, Audy P, Boyetchko SM, Dumonceaux TJ (2016) Genome Sequence of *Pseudomonas chlororaphis* Strain 189. Genome Announc. 4. https://doi.org/10.1128/genomeA.00581-16

van Bodegom P (2007) Microbial Maintenance: A critical review on its quantification. Microb Ecol. 53:513–523 . https://doi.org/10.1007/s00248-006-9049-5

Varma A, Palsson BO (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. Appl Environ Microbiol. 60:3724–3731 . https://doi.org/PMC201879

Wang X, Mavrodi D V., Ke L, Mavrodi O V., Yang M, Thomashow LS, Zheng N, Weller DM, Zhang J (2015) Biocontrol and plant growth-promoting activity of rhizobacteria from Chinese fields with contaminated soils. Microb Biotechnol. 8:404–418 . https://doi.org/10.1111/1751-7915.12158

Wences AH, Schatz MC (2015) Metassembler : merging and optimizing de novo genome assemblies. Genome Biol. 1–10 . https://doi.org/10.1186/s13059-015-0764-4

Xiang M, Kang Q, Zhang D (2020) Advances on systems metabolic engineering of *Bacillus subtilis* as a chassis cell. Synth Syst Biotechnol 5:245–251 . https://doi.org/10.1016/j.synbio.2020.07.005

Xie K, Peng H, Hu H, Wang W, Zhang X (2013) OxyR, an important oxidative stress regulator to phenazines production and hydrogen peroxide resistance in *Pseudomonas chlororaphis* GP72. Microbiol Res. 168:646–653 . https://doi.org/10.1016/j.micres.2013.05.001

Xu L, Dong Z, Fang L, Luo Y, Wei Z, Guo H, Zhang G, Gu YQ, Coleman-Derr D, Xia Q, Wang Y (2019) OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res. 47:W52–W58 . https://doi.org/10.1093/nar/gkz333

Yokozeki K, Kubota K. March 1987. Method of producing L-carnitine. US patent 4650759 A.

Yuan S-F, Alper HS (2019) Metabolic engineering of microbial cell factories for production of nutraceuticals. Microb Cell Fact. 18:46 . https://doi.org/10.1186/s12934-019-1096-y

Yue S-J, Huang P, Li S, Jan M, Hu H-B, Wang W, Zhang X-H (2020) Enhanced production of 2-hydroxyphenazine from glycerol by a two-stage fermentation strategy in *Pseudomonas chlororaphis* GP72AN. J Agric Food Chem. 68:561–566 . https://doi.org/10.1021/acs.jafc.9b05033

Zhao LF, Xu YJ, Ma ZQ, Deng ZS, Shan CJ, Wei GH (2013) Colonization and plant growth promoting characterization of endophytic *Pseudomonas chlororaphis* strain Zong1 isolated from *Sophora alopecuroides* root nodules. Brazilian J Microbiol. 44:623–631 . https://doi.org/10.1590/S1517-83822013000200043

Zuñiga C, Li C-T, Huelsman T, Levering J, Zielinski DC, McConnell BO, Long CP, Knoshaug EP, Guarnieri MT, Antoniewicz MR, Betenbaugh MJ, Zengler K (2016) Genome-scale metabolic model for the green alga *Chlorella vulgaris* UTEX 395 accurately predicts phenotypes under autotrophic, heterotrophic, and mixotrophic growth conditions. Plant Physiol. pp.00593.2016 . https://doi.org/10.1104/pp.16.00593

Apéndice 1. Fuentes de carbono metabolizables *in silico* por *P. chlororaphis* DSM 50083

Rxn Intercambio	Fuente de carbono	Rxn Intercambio	Fuente de carbono	Rxn Intercambio	Fuente de carbono
EX_arg__L_e	Arginina	EX_3oxoadp_e	3-Oxoadipato	EX_but_e	Butirato
EX_orn_e	Ornitina	EX_xtsn_e	Xantosina	EX_lys__L_e	Lisina
EX_ser__L_e	Serina	EX_fru_e	Fructosa	EX_dca_e	Decanoato
EX_glyc3p_e	Gliceraldehído 3 fosfato	EX_lac__L_e	Lactato	EX_gly_e	Glicina
EX_glyc_e	Glicerol	EX_ttdcea_e	Tetradecenoato	EX_etha_e	Etanolamina
EX_asp__L_e	Aspartato	EX_ocdcea_e	Octadecenoacto	EX_acald_e	Acetaldehído
EX_tartr__D_e	Tartrato	EX_hdcea_e	Hexadecenoato	EX_octa_e	Octanoato
EX_glu__L_e	Glutamato	EX_ocdca_e	Octadecanoato	EX_ppoh_e	Propanol
EX_glyald_e	Glicerladehído	EX_hdca_e	Hexadecanoato	EX_ac_e	Acetato
EX_his__L_e	Histidina	EX_ttdca_e	Tetradecanoato	EX_skm_e	Shikimato
EX_thr__L_e	Treonina	EX_ddca_e	Dodecanoato	EX_ppa_e	Pripionato
EX_acon_C_e	Aconitato	EX_etoh_e	Etanol	EX_meoh_e	Metanol
EX_ala__L_e	Alanina	EX_val__L_e	Valina	EX_34dhbz_e	3,4-Dihidroxibenzoato
EX_idon__L_e	Idonato	EX_ile__L_e	Isoleucina	EX_pac_e	Fenilacetato
EX_glcn_e	Gluconato	EX_hxa_e	Hexanoato	EX_tyr__L_e	Tirosina
EX_lcts_e	Lactosa	EX_acac_e	Acetoacetato	EX_4hphac_e	4-Hidroxifenilacetato
EX_sbt__D_e	Sorbitol	EX_mal__L_e	Malato	EX_3hoxpac_e	3-Hidroxifenilacetato
EX_gsn_e	Guanosina	EX_lac__D_e	Lactato	EX_vanln_e	Vanilina
EX_galctn__D_e	Galactonato	EX_fum_e	Fumarato	EX_vanlt_e	Vanilato
EX_14glucan_e	1,4-alfa-Glucano	EX_ala__D_e	Alanina	EX_hexs_e	Hexanosulfonato

Apéndice 1. Fuentes de carbono metabolizables *in silico* por *P. chlororaphis* DSM 50083 (continuación)

Rxn Intercambio	Fuente de carbono	Rxn Intercambio	Fuente de carbono	Rxn Intercambio	Fuente de carbono
EX_icit_e	Isocitrato	EX_sucr_e	Sacarosa	EX_bz_e	Benzoato
EX_cit_e	Citrato	EX_malt_e	Maltosa	EX_phe__L_e	Fenilalanina
EX_ptrc_e	Putrecina	EX_arab__L_e	Arabinosa	EX_4hbz_e	4-Hidroxibenxoato
EX_pro__L_e	Prolina	EX_akg_e	Alfacetoglutarato	EX_butso3_e	Butanosulfonato
EX_glcr_e	Glucarato	EX_4abut_e	4-aminobutirato	EX_ura_e	Uracilo
EX_galctr__D_e	Galacterato	EX_oaa_e	Oxalacetato	EX_alltn_e	Alantonia
EX_progly_e	Prolinilgicia	EX_glc__D_e	Glucosa	EX_csn_e	Citosina
EX_glyc__R_e	Glicerato	EX_mal__D_e	Malato	EX_glyclt_e	Glicolato
EX_glcur_e	Glucoronato	EX_gal_e	Galactosa	EX_ade_e	Adenina
EX_succ_e	Succinato	EX_adn_e	Adenosina	EX_xan_e	Xantina
EX_ins_e	Inosina	EX_tartr__L_e	Tartrato	EX_gua_e	Guanina
EX_gln__L_e	Glutamina	EX_rib__D_e	Ribosa	EX_2pglyc_e	2-Fosfoglicolato
EX_urate_e	Urato				

*Rxn, Reacción

Apéndice 2. Medios de cultivo utilizados
1.Medio LB (Bertani 1951)

Bacto triptona 10 g/L

NaCl 10g/L

Extracto de levadura 5 g/L

2. Medio M9 (Maniatis et al. 1982)

Fuente de carbono 2 g/L

$Na_2HPO_4*7H_2O$ 12.8 g/L

KH_2PO_4 3 g/L

NH_4Cl 1 g/l

NaCl 0.5 g/L

$MgSO_4$ 2mM

Draft Genome Sequence of *Pseudomonas chlororaphis* ATCC 9446, a Nonpathogenic Bacterium with Bioremediation and Industrial Potential

Fabian Moreno-Avitia,[a] Luis Lozano,[b] Jose Utrilla,[b] Francisco Bolívar,[a] Adelfo Escalante[a]

Departamento de Ingeniería Celular y Biocátalisis, Instituto de Biotecnología, Universidad Nacional Autónoma de México (UNAM), Cuernavaca, Morelos, México[a]; Centro de Ciencias Genómicas, UNAM, Cuernavaca, Morelos, México[b]

ABSTRACT *Pseudomonas chlororaphis* strain ATCC 9446 is a biocontrol-related organism. We report here its draft genome sequence assembled into 35 contigs consisting of 6,783,030 bp. Genome annotation predicted a total of 6,200 genes, 6,128 coding sequences, 81 pseudogenes, 58 tRNAs, 4 noncoding RNAs (ncRNAs), and 41 frameshifted genes.

Different strains of *Pseudomonas chlororaphis* have been reported as biocontrol agents over a wide variety of organisms (1–6). *P. chlororaphis* strain ATCC 9446 is a member of a genus which has features that are interesting for different applications. Its potential capability to degrade triphenyltin, an organotin compound known to cause harmful effects on a variety of aquatic organisms, has been studied (7, 8). Furthermore, this strain was reported to produce 2-keto-L-gulonic acid, L-carnitine, and 6-aminopenicillanic acid (9–11).

The genome of strain ATCC 9446v was sequenced at the Sequencing and Polymorphism Identification Unit of the Instituto Nacional de Medicina Genómica in a NextSeq 500 Illumina sequencer with a 2×150 paired-end indexed configuration, achieving a genome coverage of 120-fold. Assembly was performed with SPAdes version 3.9.0 and IDBA_UD version 1.1.1 parallel, and the contigs obtained were merged and optimized using METASSEMBLER version 1.5 (12–14). Contigs were reordered using MAUVE version snapshot_2015_02_13 using the genome of *P. chlororaphis* 189 as a reference (15).

The draft genome of *P. chlororaphis* strain ATCC 9446 consists of 6,783,030 bp assembled in 35 contigs, with an average G+C content of 63%; the contigs length average was 15,918 bp (*N50*, 843,135 bp). The draft genome was annotated using the NCBI Prokaryotic Genome Annotation Pipeline, which predicted a total of 6,200 genes, 6,128 coding sequences, 81 pseudogenes, 58 tRNAs, 4 noncoding RNAs (ncRNAs), and 41 frameshifted genes (16). Annotation revealed genes related to toluene tolerance, alkanoates, alginate, phenazines, nonribosomal peptides, and polyketide synthesis, among many other genes involved in diverse metabolic pathways with biotechnological applications.

Accession number(s). This whole-genome shotgun project was deposited at DDBJ/EMBL/GenBank under the accession number NBAT00000000. The version described in this report is BioProject PRJNA380114 and BioSample SAMN06627644.

ACKNOWLEDGMENTS

REFERENCES

1. Schmidt-Eisenlohr H, Baron C. 2003. The competitiveness of *Pseudomonas chlororaphis* carrying pJP4 is reduced in the *Arabidopsis thaliana* rhizosphere. Appl Environ Microbiol 69:1827–1831. https://doi.org/10.1128/AEM.69.3.1827-1831.2003.

2. Kupferschmied P, Maurhofer M, Keel C. 2013. Promise for plant pest control: root-associated pseudomonads with insecticidal activities. Front Plant Sci 4:287. https://doi.org/10.3389/fpls.2013.00287.

3. Dimkpa CO, Zeng J, McLean JE, Britt DW, Zhan J, Anderson AJ. 2012. Production of indole-3-acetic acid via the indole-3-acetamide pathway in the plant-beneficial bacterium *Pseudomonas chlororaphis* O6 is inhibited by ZnO nanoparticles but enhanced by CuO nanoparticles. Appl Environ Microbiol 78:1404–1410. https://doi.org/10.1128/AEM.07424-11.

4. Shen T, Liu Q, Xie X, Xu Q, Chen N. 2012. Improved production of tryptophan in genetically engineered *Escherichia coli* with TktA and PpsA overexpression. J Biomed Biotechnol 2012:605219. https://doi.org/10.1155/2012/605219.

5. Loewen PC, Villenueva J, Fernando WGD, de Kievit T. 2014. Genome sequence of *Pseudomonas chlororaphis* strain PA23. Genome Announc 2(4):e00689-14. https://doi.org/10.1128/genomeA.00689-14.

6. Wang GW, Hu WT, Huang BK, Qin LP. 2011. *Illicium verum*: a review on its botany, traditional use, chemistry and pharmacology. J Ethnopharmacol 136:10–20. https://doi.org/10.1016/j.jep.2011.04.051.

7. Inoue H, Takimura O, Fuse H, Murakami K, Kamimura K, Yamaoka Y. 2000. Degradation of triphenyltin by a fluorescent pseudomonad. Appl Environ Microbiol 66:3492–3498. https://doi.org/10.1128/AEM.66.8.3492-3498.2000.

8. Inoue H, Takimura O, Kawaguchi K, Nitoda T, Fuse H, Murakami K, Yamaoka Y. 2003. Tin-carbon cleavage of organotin compounds by pyoverdine from *Pseudomonas chlororaphis*. Appl Environ Microbiol 69:878–883. https://doi.org/10.1128/AEM.69.2.878-883.2003.

9. Huang HT, English AR. March 1966. Production of 6-aminopenicillanic acid. US patent 3239427 A.

10. Sonoyama T, Kageyama B, Honjo T. November 1975. Process for producing 2-keto-L-gulonic acid. US patent 3922194 A.

11. Yokozeki K, Kubota K. March 1987. Method of producing L-carnitine. US patent 4650759 A.

12. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and Its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.

13. Peng Y, Zhang MM, Chen ZF, Hu K, Liu YC, Chen X, Liang H. 2013. Synthesis, characterization, and interaction with biomolecules of platinum(II) complexes with shikimic acid-based ligands. Bioinorg Chem Appl 2013:565032. https://doi.org/10.1155/2013/565032.

14. Wences AH, Schatz MC. 2015. Metassembler: merging and optimizing de novo genome assemblies. Genome Biol 16:207. https://doi.org/10.1186/s13059-015-0764-4.

15. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403. https://doi.org/10.1101/gr.2289704.

16. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Ciufo S, Li W. 2013. Prokaryotic genome annotation pipeline. *In* the NCBI Handbook, 2nd ed. National Center for Biotechnology Information, Bethesda, MD.

CrossMark

Analysis of differentially upregulated proteins in *ptsHIcrr*⁻ and *rppH*⁻ mutants in *Escherichia coli* during an adaptive laboratory evolution experiment

César Aguilar[1] · Gabriel Martínez-Batallar[2] · Noemí Flores[3] · Fabián Moreno-Avitia[3] · Sergio Encarnación[2] · Adelfo Escalante[3] · Francisco Bolívar[3,4]

Abstract

The previous deletion of the cytoplasmic components of the phosphotransferase system (PTS) in *Escherichia coli* JM101 resulted in the PTS⁻ derivative strain PB11 with severely impaired growth capability in glucose as the sole carbon source. Previous adaptive laboratory evolution (ALE) experiment led to select a fast-growing strain named PB12 from PB11. Comparative genome analysis of PB12 showed a chromosomal deletion, which result in the loss of several genes including *rppH* which codes for the RNA pyrophosphohydrolase RppH, involved in the preparation of hundreds of mRNAs for further degradation by RNase E. Previous inactivation of *rppH* in PB11 (PB11*rppH*⁻) improved significantly its growing capabilities and increased several mRNAs respect its parental strain PB11. These previous results led to propose to the PB11*rppH*⁻ mutant as an intermediate between PB11 and PB12 strains merged during the early ALE experiment. In this contribution, we report the metabolic response to the PTS⁻ and *rppH*⁻ mutations in the deep of a proteomic approach to understanding the relevance of *rppH*⁻ phenotype during an ALE experiment. Differentially upregulated proteins between the wild-type JM101/PB11, PB11/PB11*rppH*⁻, and PB11/PB12 comparisons led to identifying 45 proteins between strain comparisons. Downregulated or upregulated proteins in PB11*rppH*⁻ were found expressed at an intermediate level with respect to PB11 and PB12. Many of these proteins were found involved in non-previously metabolic traits reported in the study of the PTS⁻ strains, including glucose, amino acids, ribose transport; amino acid biosynthesis; NAD biosynthesis/salvage pathway, biosynthesis of Ac-CoA precursors; detoxification and degradation pathways; stress response; protein synthesis; and possible mutator activities between comparisons. No changes were found in the expression of galactose permease GalP, previously proposed as the primary glucose transporter in the absence of PTS selected by the PTS⁻ derivatives during the ALE experiment. This result suggests that the evolving PTS⁻ population selected other transporters such as LamB, MglB, and ManX instead of GalP for glucose uptake during the early ALE experiment. Analysis of the biological relevance of the metabolic traits developed by the studied strains provided valuable information to understand the relevance of the *rppH*⁻ mutation in the PTS⁻ background during an ALE experiment as a strategy for the selection of valuable phenotypes for metabolic engineering purposes.

Keywords *Escherichia coli* PTS⁻ · Short-term adaptive laboratory evolution · Proteomic analysis · *rppH* · MglB · GalP

César Aguilar and Gabriel Martínez-Batallar contributed equally to this work.

✉ Adelfo Escalante
  adelfo@ibt.unam.mx

1 Laboratorio Nacional de Genómica para la Biodiversidad, Unidad de Genómica Avanzada, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Irapuato, Guanajuato, México

2 Programa de Genómica Funcional de Procariontes, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México (UNAM), Cuernavaca, Morelos, México

3 Departamento de Ingeniería Celular y Biocatálisis, Instituto de Biotecnología, UNAM, Cuernavaca, Morelos, Mexico

4 Member of El Colegio Nacional, Ciudad de México, México

⚠ Springer

Introduction

In *Escherichia coli*, the phosphoenolpyruvate:carbohydrate phosphotransferase (PTS) is the primary system for the transport and phosphorylation of glucose. This system has a significant impact on the cellular carbon flux distribution and plays a crucial role in the global signaling that controls the preferential consumption of glucose over other carbon sources (Escalante et al. 2012; Carmona et al. 2015). In a minimal medium containing glucose as the sole carbon source, the PTS consumes 50% of the available phosphoenolpyruvate (PEP), whereas the remaining 50% is distributed differentially among other relevant pathways (Flores et al. 2002; Gosset 2005). Several metabolic engineering approaches for the overproduction of aromatic compounds in *E. coli* have implemented the deletion of the *ptsHIcrr* operon, which encodes the cytoplasmic components of the PTS. The absence of this operon results in higher amounts of PEP, which, through metabolic engineering, is diverted toward the synthesis of commercially relevant products (Rodriguez et al. 2014; Martínez et al. 2015). The previous inactivation of this glucose translocation system in the wild-type strain of *E. coli* JM101 resulted in a PTS⁻ derivative (strain PB11) that was severely impaired in its ability to transport glucose (PTS⁻ Glc⁻) (Flores et al. 2005, 2007; Aguilar et al. 2012). Therefore, the specific growth rate (μ) of this strain decreased to 0.13 h⁻¹ in M9 minimal medium with glucose as the sole carbon source, representing a decrease of more than 80% compared with the μ of the parental strain JM101 ($\mu = 0.75$ h⁻¹). The PTS⁻ strain PB11 was subjected to a short-term adaptive laboratory evolution (ALE) experiment for the selection of faster-growing derivatives in minimal M9 broth with glucose as the sole carbon source. A derived mutant strain with improved growth capabilities, named, PB12 (PTS⁻ Glc⁺), isolated after 120 h of cultivation had an increased growth, with a $\mu = 0.44$ h⁻¹, representing a ~330% increase respect to the μ of the strain PB11 (Flores et al. 2007).

Previous physiological and transcriptional characterizations of PB11 led to propose that in the absence of PTS, this strain selected for glucose transport, the galactose permease, GalP. Additionally, to this physiological adaptation, this strain developed a constitutive scavenging-stress condition, where gluconeogenic, anaplerotic, and carbon-recycling pathways were triggered in response to the inactivation of the cytoplasmic components of the PTS (Flores et al. 2005). This inactivation also resulted in the abolition of the catabolic repression machinery, thereby allowing the PTS⁻ strains to simultaneously consume different carbon sources (Flores et al. 2005; Martínez et al. 2008; Sigala et al. 2009; Sabido et al. 2014). Comparative genome sequencing of the evolved strain PB12 allowed for the identification of genetic changes undergone by PB11 during the ALE experiment (Aguilar et al. 2012). One of the relevant genetic changes in the strain PB12 was the loss of a 10,328-bp chromosomal fragment, which deleted the *ptsP*,

rppH, *ydgT*, *mutH*, *ygdQ*, *ygdR*, *tas*, *lplT*, *aas*, *omrA*, *omrB*, and *galR* genes. The study of the physiological relevance of the deletion of some of these genes on the growth of strain PB11 was previously analyzed (Aguilar et al. 2012). Among deleted genes, *rppH* codes for the RNA pyrophosphohydrolase RppH, which hydrolyzes the diphosphorylated 5′-ends of mRNAs to a monophosphorylated form, suitable for further degradation by RNase E (Deana et al. 2008; Luciano et al. 2017; Luciano et al. 2018). In vivo and in vitro activity of RppH was stimulated when a diaminopimelate epimerase (DapF) interact directly with RppH resulting in the promotion and stabilizing of this enzyme facilitating RNA binding and stimulating mRNA hydrolysis (Wang et al. 2018). Deletion of RppH was involved in stabilization of ~ 400 mRNAs (Himabindu and Anupama 2017). However, the deletion of RppH has been proposed to has a pleiotropic role, mainly by maintaining the cellular envelope integrity as its deletion resulted in sensitivity to antibiotics and stress such as osmotic and ethanol, resulting in an increased envelope permeability (Choi et al. 2017).

Our previous results showed that deletion of *rppH* in PB11 (PB11*rppH⁻*) resulted in a significant gain of fitness (2.6-fold) and an increment in its μ (0.34 h⁻¹) compared with the PB11 (Aguilar et al. 2012). As expected, loss of this gene in the PTS⁻ strain PB11 increased half-lives of several mRNAs coding for alternative glucose transporter systems, glycolytic, the tricarboxylic acid (TCA) cycle, gluconeogenic, glyoxylate, *pox* shunt, respiratory, and ppGpp metabolism (Aguilar et al. 2012). These previous results led to propose to the PB11*rppH⁻* mutant as an earlier event during the ALE experiment, and a possible intermediate between PB11 and PB12 strains.

In this contribution, we report the comparative analysis of the upregulated proteins in response to the deletion of the cytoplasmic components of PTS in the wild-type strain JM101, in response to deletion of *rppH* in PB11 (PB11*rppH⁻*), and in the fast-growing evolved strain PB12. As a result, 45 proteins were identified as differentially expressed in the analyzed PTS⁻ strains with respect to the wild-type strain JM101. Differentially upregulated proteins between JM101/PB11, PB11/PB11*rppH⁻*, and PB11*rppH⁻*/PB12 comparisons led to identifying several non-previously metabolic traits reported for the studied strains including glucose, amino acids, ribose transport; amino acid biosynthesis; NAD biosynthesis/salvage pathway, biosynthesis of Ac-CoA precursors; detoxification and degradation pathways; stress response; protein synthesis; and possible mutator activities between comparisons. No correlation was found between many differentially upregulated proteins with previous transcriptomic data reported. This result is particularly relevant for the galactose permease GalP (*galP*) as no changes were found in relative protein concentration of GalP, previously proposed as the primary glucose transporter selected by the PTS⁻ population during the ALE experiment (Flores et al. 2005; Aguilar et al. 2012). Differentially expressed

protein comparisons between strains provided valuable information in the deep of a proteomic approach to understand the role of the deletion of *rppH* in a PTS⁻ background during a short-term ALE experiment, on the metabolic solutions developed by the cell to improve glucose transport (probably by LamB, MglB, and ManX) during the early ALE experiment to develop a fast-growing derivative strain such as PB12. This knowledge could be beneficial for the further optimization of valuable PTS⁻ derivative strains for the production of relevant biotechnological products.

Materials and methods

Bacterial strains, growth conditions, and gene deletion method

Table 1 shows *E. coli* strains included in this study. The PTS⁻ derivatives PB11 and PB12, lacking the *galP* gene, were obtained by knockout-mediated inactivation by the Datsenko and Wanner method (Datsenko and Wanner 2000) using the oligonucleotide sequence shown in Table S1. PCR verified gene disruption. For expression of *glk*, this gene was fused to the *trc* promoter and cloned into the low-copy-number plasmid pCL1920 (Hernández-Montalvo et al. 2003), generating the pCLv1Glk plasmid.

To prepare the inocula, strains stored at − 72 °C in glycerol were inoculated and cultured into Luria (LB) broth for overnight. For growth, μ, and specific glucose uptake (qs) rate determinations, cells were cultured in LB broth and then used to inoculate flasks with M9 minimal medium containing 2 g/L glucose as the sole carbon source. During the growth phase, the cultures were used to inoculate fresh flasks containing the same, prewarmed medium (50 mL), and the flasks were incubated at 37 °C, shaking at 300 rpm; the initial optical density at 600 nm of the cultures was 0.1, as determined by using a photocolorimeter (Klett/Summerson device model 800–3, Bel-Art Products, Wayne, USA). All the μ values presented in Table 2 are the means ±

standard deviation of at least two independent cultures, each in duplicate. For the strains that had low growth rates, it was necessary to grow up to 10 independent cultures, in duplicate, to establish whether the variation was significant. Only for the wild-type strain JM101 (New England Biolabs, Inc. Ipswich, USA), two independent cultures, in duplicate, were performed since the differences observed were minimal, and the data agreed with previously published values.

Comparative analysis of upregulated proteins in response to the inactivation of PTS and *rppH*

E. coli proteins were obtained as previously reported (Salazar et al. 2010; Higareda-Almaraz et al. 2011; Martínez-Gómez et al. 2012), and briefly described: sonication obtained proteins at 24 kHz for 5 cycles consisting of 1 min on and 1 min off at 4 °C in the presence of protease inhibitor, and protein isolation was phenol extracted to limit proteolysis. To solubilize and obtain completely denatured and reduced proteins, pellets were dried and resuspended as previously reported (Encarnación et al. 2005). Samples were mixed with 7 M urea, 2 M thiourea, 4% 3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate (CHAPS), 2 mM tributyl phosphine (TBP), 2% ampholytes, and 60 mM dithiothreitol (DTT) prior to electrophoresis. Methods used for sample preparation, analytical and preparative two-dimensional (2D) polyacrylamide gel electrophoresis (PAGE), and image analysis have been described previously (Salazar et al. 2010; Higareda-Almaraz et al. 2011; Martínez-Gómez et al. 2012). pH gradients were determined by using a 2D SDS-PAGE standard (Sigma-Aldrich, San Luis Missouri, USA). The protein concentration of the supernatant was determined by the Bradford method: 500 μg of protein was loaded per gel. 2D gels were stained with colloidal Coomassie blue G-250 (Higareda-Almaraz et al. 2011).

In silico processing of gel images was made as previous reports (Salazar et al. 2010; Higareda-Almaraz et al. 2011; Martínez-Gómez et al. 2012; de Luna-Valdez et al. 2014): Images of 2D-gels of three biologically independent protein

Table 1 *E. coli* strains employed in this work

Strains	Relevant characteristics	Reference
JM101	[F′ *traD36 proAB*⁺ *lac*I�q *lac*ZΔM15/*supE thi* Δ(*lac-proAB*) *rpoS(33am)*]	Flores et al. 2005
PB11	JM101*ptsH-ptsI-crr*::kan	Flores et al. 2005
PB12	PB11ev.	Flores et al. 2005
PB11Δ*galP*	PB11*galP*::Cm	This work
PB11Δ*galP*/pCL*glk*	PB11*galP*::Cm/pCL*glk*	This work
PB12Δ*galP*	PB12*galP*::Cm	This work
PB12Δ*galP*/pCL*glk*	PB12*galP*::Cm/pCL*glk*	This work
Plasmid		
pCL*glk*	pCL1920 derivative harboring *glk* from *Escherichia coli* W3110	Hernández-Montalvo et al. 2003

Table 2 Specific growth (μ) and specific glucose uptake (qs) rates of the wild-type JM101 and PTS$^-$ derivative strains

Strain	μ[1] (h^{-1})	Relative value vs JM101 strain	qs[1] (g Glc/g DCW h)	Relative value vs JM101 strain
JM101	0.75 ± 0.04	100.00	2.05 ± 0.12	100.00
PB11	0.13	17.33	0.48 ± 0.01	23.41
PB11$\Delta galP$	0.15 ± 0.01	20.00	0.5	24.39
PB11$\Delta galP$/pCLglk	0.25 ± 0.01	33.33	0.84 ± 0.05	40.97
PB12	0.44 ± 0.02	58.66	1.25 ± 0.02	60.97
PB12$\Delta galP$	0.23 ± 0.01	30.66	0.67 ± 0.05	32.68
PB12$\Delta galP$/pCLglk	0.23 ± 0.02	30.66	0.72 ± 0.09	35.12

[1] Values correspond to the mean ± standard deviation of at least two independent cultures, each one in duplicate

extracts from analyzed strains (JM101, PB11, PB11$rppH^-$, and PB12) were generated at a resolution of 127×127 μm in a GS-800 densitometer. Digital images were analyzed and compared with PDQuest 8.0.1 software (Bio-Rad Laboratories, Hercules, USA), and the total number of spots was determined. The total number of spots detected for the triplicate experiments of strain JM101 was as follows: JM101_1, 825 spots; JM101_2, 815 spots; and JM101_3, 791 spots, with 766 matching spots between the triplicates. For strain PB11: PB11_1, 814 spots; PB11_2, 825 spots; and PB11_3, 836 spots, with 718 matching spots between triplicates. For strain PB11$rppH^-$: PB11$rppH^-$_1, 872 spots; PB11$rppH^-$_2, 835 spots; and PB11$rppH^-$_3, 803 spots, with 726 matching spots. For strain PB12: PB12_1, 837 spots; PB12_2, 780 spots; and PB12_3, 842 spots, with 727 matching spots between triplicates. The total number of matching spots detected in all replicates was 581. Following gel comparisons, the spots of interest were identified (e.g., those that differed between treatments), and the phenotypes tested for reproducibility (with a Student's t test value $p \leq 0.01$ using PDQuest 2D-image analysis software). Comparison between strains threw several differentially expressed proteins (spots): JM101 and PB11, 173 spots; JM101 and PB11$rppH^-$, 169 spots; and JM101 and PB12, 145 spots. From this comparison, 53 spots displaying an average of $\geq$ 1-fold increase or decrease in abundance (Margalef-Català et al. 2016) with a $p \leq 0.01$ were selected for further protein identification. Spots of interest were cut, alkylated, reduced, digested, and automatically transferred to a MALDI analysis target by a Proteineer SP II and SP robot using the SPcontrol 3.1.48.0 v software (Bruker Daltonics, Germany), with the aid of a DP Chemicals 96 gel digestion kit (Bruker Daltonics, Germany) and processed in a MALDI-TOF Autoflex (Bruker Daltonics, Germany) to obtain a mass fingerprint: 100 successful shots were fired in 20 steps; the peak resolution threshold was set at 1500; the signal/noise ratio of tolerance was 6, and contaminants were not excluded. The spectra were externally calibrated using a peptide calibration standard by flexAnalysis 1.2 v SD1 Patch 2 software (Bruker Daltonics). The mass range acquired was from 600 to 4000. The search engine

MASCOT (Perkins et al. 1999) was used to compare the fingerprints against the UniPROT (The UniProt Consortium 2010) release 2011–01 database with the following parameters: taxon, *Escherichia coli*; type of search, peptide mass fingerprint; enzyme, trypsin; fixed modifications, carbamidomethyl (C); variable modifications, oxidation (M); mass values, monoisotopic; protein mass, unrestricted; peptide mass tolerance, up to 300 ppm; peptide charge state, 1+; and maximum missed cleavages, 1. Only proteins with statistically significant MASCOT scores ($p < 0.05$) are shown (Table S2). Relative protein concentration to wild-type strain JM101 is expressed as the average of recorded density for the corresponding spots.

Results

Upregulated proteins identified in response to the PTS and *rppH* deletion

Protein extraction and two-dimensional gel electrophoresis of analyzed strains resulted in the identification of 45 proteins differentially expressed with respect to the parental strain JM101 (considering $\geq$ 1 average-fold increase in abundance with a $p < 0.05$) (Fig. 1). Differentially expressed proteins were grouped on the basis to the biological process in which they are involved according to the Gene Ontology (GO) Consortium classification (Ashburner et al. 2000; Keseler et al. 2017). Results of this analysis grouped proteins involved in transport, oxidation-reduction process/oxidation metabolic processes, carbohydrate metabolic processes, response to stress, translation/peptide biosynthetic processes, biosynthetic processes, other metabolic processes, binding activity, proteolysis, and regulation (Table 3).

Coding gene, b number, protein name, and relative protein concentration respect parental strain JM101 is shown in Table 4. The biological role of several relevant proteins was analyzed in the EcoCyc database (Keseler et al. 2017) and is shown in Figs. 2 and 3. The differential spots, from the
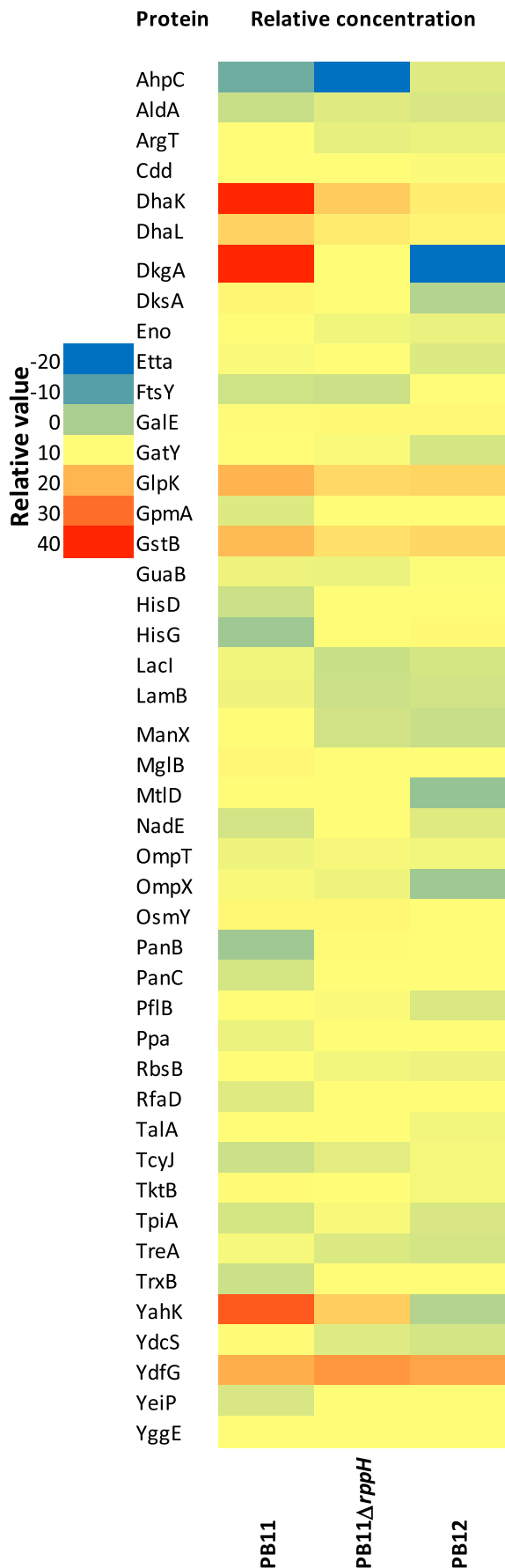
digitalized protein gels that were the most significantly changed among strains PB11, PB11*rppH⁻*, and PB12 are shown in Fig. 4, while the complete images of 2D protein separation are in Fig. S1.

Differentially upregulated proteins in response to the PTS inactivation

Analysis of upregulated proteins in the PTS⁻ derivative PB11 led to identifying 42 upregulated proteins with respect to the parental strain JM101 in response to the inactivation of the *ptsHIcrr* operon. These include the outer membrane OmpX and LamB; the inner membrane transporters MglB, and ManX, involved in carbohydrate transport; the amino acid transporters ArgT (lysine, arginine, and ornithine); and RbsB, involved in ribose transport. Other membrane-associated proteins found to be upregulated, including MtlD and YdcS, are involved in bacterial cell division (Table 4). MtlD is a lytic murein transglycosylase involved in cell division, while YdcS is a putative transport protein, probably associated with protein transit through the periplasmic space to the outer membrane (Keseler et al. 2017). The periplasmic proteins TreA, a periplasmic trehalase, and OsmY, a periplasmic chaperone, reported to be osmotically induced (Keseler et al. 2017), were also upregulated (Table 4).

Cytoplasmic proteins found upregulated in the strain PB11 are involved in several biosynthetic, catabolic, and general cellular processes. GpmA, Eno, TktB, and TalA are involved in the central carbon metabolism (CCM), lower glycolysis, and the pentose phosphate pathway (PPP), while GalE participates in the galactose degradation pathway (Fig. 2). PflB, a pyruvate formate-lyase, catalyzes the synthesis of Ac-CoA from PYR + CoA, whereas other upregulated cytoplasmic proteins (YeiP, EttA, Ppa, YggE, and DksA) are involved in essential process for cellular growth, such as protein synthesis and energy metabolism (Table 4 and Fig. 3) (Perederina et al. 2004; Keseler et al. 2017). The unregulated proteins Cdd and NadE (Table 4 and Fig. 3) are involved in pyrimidine ribonucleotide degradation and the NAD biosynthesis/salvage pathways, respectively (Keseler et al. 2017).

Other upregulated proteins detected in the strain PB11 respect JM101 were DkgA, a methylglyoxal reductase, and YahK, an NADPH-dependent aldehyde reductase. DkgA is involved in the degradation of methylglyoxal, a highly toxic compound produced in small amounts during glycolysis or as a byproduct of fatty acid and amino acid (threonine) metabolism. This compound is detoxified via several pathways, resulting in the

Table 3 Biological processes associated to differentially upregulated proteins identified

Biological processes[a]	Proteins	Total
Transport	ArgT, TcyJ, MglB, ManX, LamB, YdcS	6
Oxidation-reduction process/oxidation metabolic processes	YahK, YdfG, GstB, DkgA, TrxB, HisD, GuaB, AldA	8
Carbohydrate metabolic processes	GlpK, DhaL, DhaK, TalA, RfaD, RbsB, PflB, GpmA, GatY, GalE, Eno	11
Response to stress	AhpC, OsmY, YggE, TreA	4
Translation/peptide biosynthetic activity	YeiP, EttA	2
Biosynthetic processes	PanB, PanC, NadE, HisG	4
Other metabolic processes	TktB, TpiA, Ppa, MltD, Cdd, FtsY	6
Binding activity	DksA, OmpX	2
Proteolysis	OmpT	1
Regulation	LacI	1
Total differentially upregulated proteins		45

[a] According to the Gene Ontology Consortium classification (Ashburner et al. 2000; Keseler et al. 2017)

synthesis of pyruvate, which in turn is used in lower glycolysis (Cooper and Anderson 1970) or in the synthesis of (S)-propane-1,2-diol (1,2-propanediol), which is transported out of the cell. YhaK is involved in the degradation of (S)-lactaldehyde to 1,2-propanediol (Keseler et al. 2017) (Table 4 and Fig. 3). Another upregulated enzyme found to be involved in detoxification pathways was GstB, glutathione S-transferase B (Table 4 and Fig. 3) (Cooper 1984; Keseler et al. 2017).

Upregulated proteins in response to the inactivation of *rppH* in the strain PB11

Inactivation of *rppH* in PB11 (PB11*rppH*⁻) resulted in a decreased relative protein concentration of 24 proteins shown as upregulated in the strain PB11, including the OMPs OmpX and LamB, the transporter proteins MglB, ManX, ArgT, and RbsB. Among these proteins, the levels of MglB, ManX, and RbsB were found to be drastically reduced, by 76, 71, and 66%, respectively, with respect to PB11. Other membrane-associated, periplasmic, and stress-response proteins observed as being upregulated in PB11, including MtlD, YdcS, and TreA, exhibited decreased relative concentrations in PB11*rppH*⁻. The cytoplasmic proteins that exhibited reduced relative concentrations were Eno (lower glycolysis); GatY (gluconeogenesis); TktB and TalA (PPP); DhaK, DhaL, and GlpK (glycerol metabolism); PflB (synthesis of Ac-CoA); Cdd (pyrimidine degradation); YahK and DkgA (detoxification); YggE (protein synthesis); and DksA (transcript elongation) (Keseler et al. 2017) (Table 4 and Figs. 2 and 3). The levels of FtsY (signal recognition particle receptor), GpmA (glycolysis), OmpT (OMP protease), and OsmY (response to stress) remained unchanged in this strain (Table 4 and Figs. 1, 2, and 3).

Upregulated proteins in response to the inactivation of *rppH* in PB11 (14 proteins) were identified as involved in cystine/cysteine transport (TcyJ), oxidation-reduction process/oxidation

metabolic process (AldA, HisD, and YdfG), carbohydrate metabolic processes (GalE, RfaD), translation activity (YeiP), biosynthetic processes (NadE, PanC), and other metabolic processes (Ppa, TpiA, TrxB). Other proteins involved in biosynthetic processes (HisG and PanB) were found upregulated in the PB11*rppH*⁻ derivative but not in its parental strain PB11 (Table 4). Interestingly, downregulated or upregulated proteins in PB11*rppH*⁻ were found expressed at an intermediate level with respect to PB11 and PB12 (Table 4 and Figs. 1, 2, and 3).

Differentially upregulated proteins in the evolved strain PB12

The evolved strain PB12 showed the reduction in the relative concentration of 26 proteins, upregulation in 11 proteins, and no change in eight proteins, with respect to PB11*rppH*⁻. Downregulated proteins in the evolved strain PB12 were identified as involved in transport (MglB), oxidation/reduction metabolic processes (AldA, YdfG, DkgA, YhaK), carbohydrate metabolism (RfaD, Eno, GatY, PflB, RbsB, TalA, DhaK, DhaL), response to stress (OsmY, YggE), translation (EttA, YeiP), biosynthetic processes (NadE, PanB), other metabolic processes (Ppa, TpiA, Cdd, MtlD, TktB), and binding activity (OmpX, DksA) PB12 (Table 4 and Figs. 1, 2, and 3).

Upregulated proteins respect the PB11*rppH*⁻ derivative included proteins involved in transport (TcyJ), oxidation/reduction metabolic processes (GuaB, HisD, GstB), carbohydrate metabolism (GalE, GpmA, GlpK), response to stress (AhpC), translation (TstY), and biosynthetic processes (PanC, HisG). Finally, eight proteins were found with no-changes changes respect the *rppH*⁻ derivative: transport (ArgT, LamB, ManX, YdcS), oxidation/reduction metabolic processes (TrxB), response to stress (TreA), OMP protease (OmpT), and regulation (LacI) (Table 4 and Figs. 1, 2, and 3). AhpC (a component of the peroxidase component of alkyl hydroperoxide reductase possible involves in response to

Table 4 Relative protein concentration in the derivative strains PB11, PB11$rppH^-$, and PB12

Proteomic analysis

Gene	Accession ID[a]	Protein	Protein name[a]	Relative protein concentration		
				PB11	PB11 $\Delta rppH$	PB12
ahpC	b0605	AhpC	Alkyl hydroperoxide reductase, AhpC component	−3.6	−9.4	2.2
aldA	b1415	AldA	Aldehyde dehydrogenase A	1.1	2.3	1.9
argT	b2310	ArgT	Lysine/arginine/ornithine ABC transporter periplasmic binding protein	5.9	2.6	2.9
cdd	b2142	Cdd	Cytidine/deoxycytidine deaminase	10.0	5.7	3.7
eno	b2779	Eno	Enolase	5.2	3.1	2.8
ettA	b4391	Etta	Energy-dependent translational throttle protein EttA	3.6	7.5	2.1
ftsY	b3464	FtsY	Signal recognition particle receptor	1.3	1.3	3.8
galE	b0759	GalE	UDP-glucose 4-epimerase	18.0	19.2	24.4
gatY	b2096	GatY	Tagatose-1,6-bisphosphate aldolase 2 subunit GatY	8.7	3.6	1.8
gpmA	b0755	GpmA	2,3-Bisphosphoglycerate-dependent phosphoglycerate mutase	2.1	6.9	10.3
guaB	b2508	GuaB	IMP dehydrogenase	3.0	2.9	3.8
hisD	b2020	HisD	Histidinal dehydrogenase/histidinol dehydrogenase	1.2	3.9	7.6
hisG	b2019	HisG	ATP phosphoribosyltransferase	−1.1	8.6	18.8
lacI	b0345	LacI	LacI DNA-binding transcriptional repressor LacI	3.2	1.2	1.7
lamB	b4036	LamB	Maltose outer membrane porin/phage lambda receptor protein	3.1	1.2	1.5
manX	b1817	ManX	Mannose-specific PTS enzyme IIAB component	5.0	1.5	1.0
mglB	b2150	MglB	D-Galactose/methyl-galactoside ABC transporter periplasmic binding protein	28.9	7.2	5.5
mtlD	b3600	MtlD	Mannitol-1-phosphate 5-dehydrogenase	9.6	4.9	−1.6
nadE	b1740	NadE	NAD synthetase, NH$_3$-dependent	1.5	6.1	2.3
ompT	b0565	OmpT	Outer membrane protease VII (outer memebrane protein 3b)	3.0	3.5	3.2
ompX	b0814	OmpX	Outer membrane protein X	3.5	3.0	−1.1
osmY	b4376	OsmY	Periplasmic chaperone OsmY	28.0	27.5	6.8
panB	b0134	PanB	3-Methyl-2-oxobutanoate hydroxymethyltransferase	−1.1	12.0	5.0
panC	b0133	PanC	Pantothenate synthetase	1.7	3.9	4.7
pflB	b0903	PflB	Pyruvate formate-lyase (inactive)	11.0	3.7	2.0
ppa	b4226	Ppa	Inorganic pyrophosphatase	2.8	9.6	6.5
rbsB	b3751	RbsB	Ribose ABC transporter periplasmic binding protein	9.7	3.3	2.9
rfaD	b3619	RfaD	ADP-L-glycero-D-mannoheptose-6-epimerase	2.2	8.3	6.0
talA	b2464	TalA	Transaldolase A	11.2	7.5	3.3
tcyJ	b1920	TcyJ	Cystine/cysteine ABC transporter periplasmic binding protein	1.2	2.4	3.4
tktB	b2465	TktB	Transketolase II	13.2	6.8	3.4
tpiA	b3919	TpiA	Triosephosphate isomerase	1.7	3.6	1.8
treA	b1197	TreA	Periplasmic trehalase	3.4	2.0	1.6
trxB	b0888	TrxB	Thioredoxin reductase	1.2	7.5	7.0
ydcS	b1440	YdcS	Putative ABC transporter periplasmic binding protein/polyhydroxybutyrate synthase	13.7	2.2	1.6
yeiP	b2171	YeiP	Elongation factor P family protein	1.9	10.2	9.1
yggE	b2922	YggE	Uncharacterized membrane protein YggE	10.8	8.0	3.9
dkgA	**b3012**	**DkgA**	**Methylglyoxal reductase**	**53.4**	**25.9**	**0**
dhaK	**b1200**	**DhaK**	**Dihydroxyacetone kinase subunit K**	**869.4**	**200**	**67.3**
dhaL	**b1199**	**DhaL**	**Dihydroxyacetone kinase subunit L**	**173**	**74.8**	**40.6**
dksA	**b0145**	**DksA**	**RNA polymerase-binding transcription factor DksA**	**35.6**	**7.5**	**0**
glpK	**b3926**	**GlpK**	**Glycerol kinase**	**300.2**	**150.5**	**165.1**
gstB	**b0828**	**GstB**	**Glutathione S-transferase B**	**270.4**	**121.5**	**153.5**
ydfG	**b1539**	**YdfG**	**3-Hydroxy acid dehydrogenase**	**322**	**416.2**	**363.6**
yahK	**b3106**	**YahK**	**Bicupin-related protein**	**663.6**	**194.7**	**0**

Values are relative to the protein concentration in the reference strain JM101. Only absolute fold-change values ≥ 1-fold were considered as significantly different and were considered for discussion purposes. [a] Data retrieved from the EcoCyc database (Keseler et al. 2017). The data in the first section of the table correspond to differential regulated proteins detected in all the studied strains relative to the wild-type strain JM101. The bottom part of the table (bold type text) included those proteins not detected in the control strain JM101 (concentration = 0) but expressed and identified in the other strains included in this study. For these, protein concentration values shown are expressed as the average of recorded density for the corresponding spots
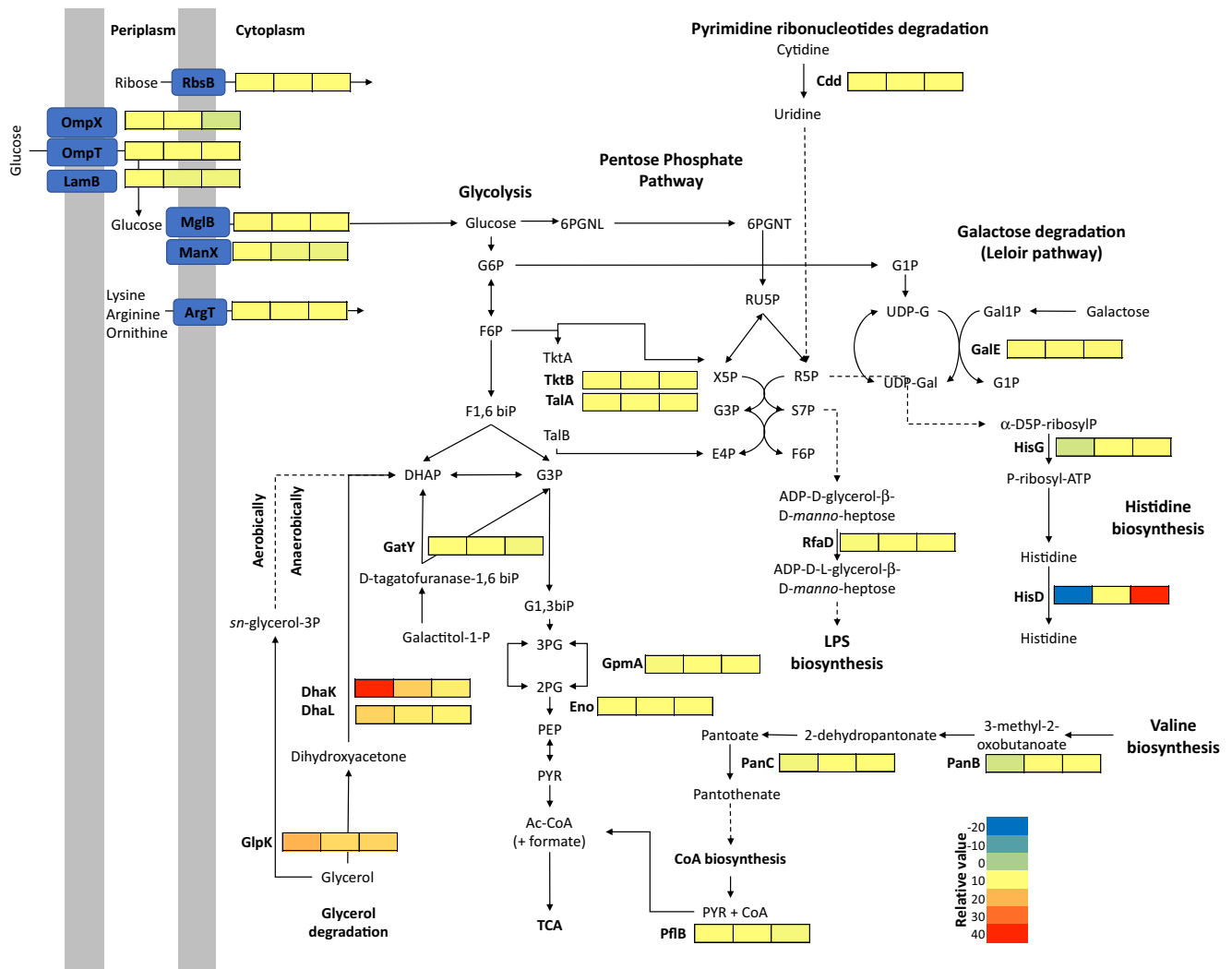
Fig. 2 Proposed metabolic traits in the strains PB11, PB11*rppH⁻*, and PB12: outer membrane proteins, transporters, proteins involved in central carbon metabolism, biosynthetic metabolism, and gluconeogenic processes. Relative protein concentration values were referred respect the parental strain JM101 and shown as a heat map. Values in squares represent protein concentration detected in the strain PB11 (left), PB11*rppH⁻*, and PB12 (right), respectively. Dotted arrows represent multiple reactions in the pathway. Continuous lines represent a unique enzymatic reaction. Metabolic pathways were adapted from the information of the EcoCyc database

stress (Keseler et al. 2017)) were found upregulated in the evolved strain PB12 but not in the PB11 and the *rppH⁻* derivative (Table 4).

Effect of the deletion of *galP* and cloning of an additional copy of *glk* on the μ and qs of the PTS⁻ strains PB11 and PB12

Previous results showed that the PTS⁻ strain PB11 selected the galactose permease (GalP) for glucose transport in the absence of PTS, as this strain showed increased transcript levels of *galP* with respect to the wild-type strain JM101 (Flores et al. 2005). To evaluate the relevance of GalP on the transport of glucose as the sole carbon source and the role of glucokinase (Glk) in

the phosphorylation of transported glucose in the PTS⁻ strains PB11 and PB12, the *galP* gene was inactivated, and *glk* cloned into the low-copy-number plasmid pCL1920, in these strains. Deletion of *galP* did not have an apparent adverse effect on the μ of the derivative strain PB11Δ*galP* compared to that of PB11, whereas the derivative PB12Δ*galP* showed μ, and qs values decreased by ~52 and ~54%, respectively, compared to those of PB12 (Table 2). On the other hand, the expression of *glk* in PB11Δ*galP*/pCL*glk* increased the μ and qs values by ~167 and ~168%, respectively, compared to those of strain PB11Δ*galP*, whereas in the strain PB12Δ*galP*, the expression of *glk* led to only a slightly increased qs value (~107%) with respect to the parental strain.
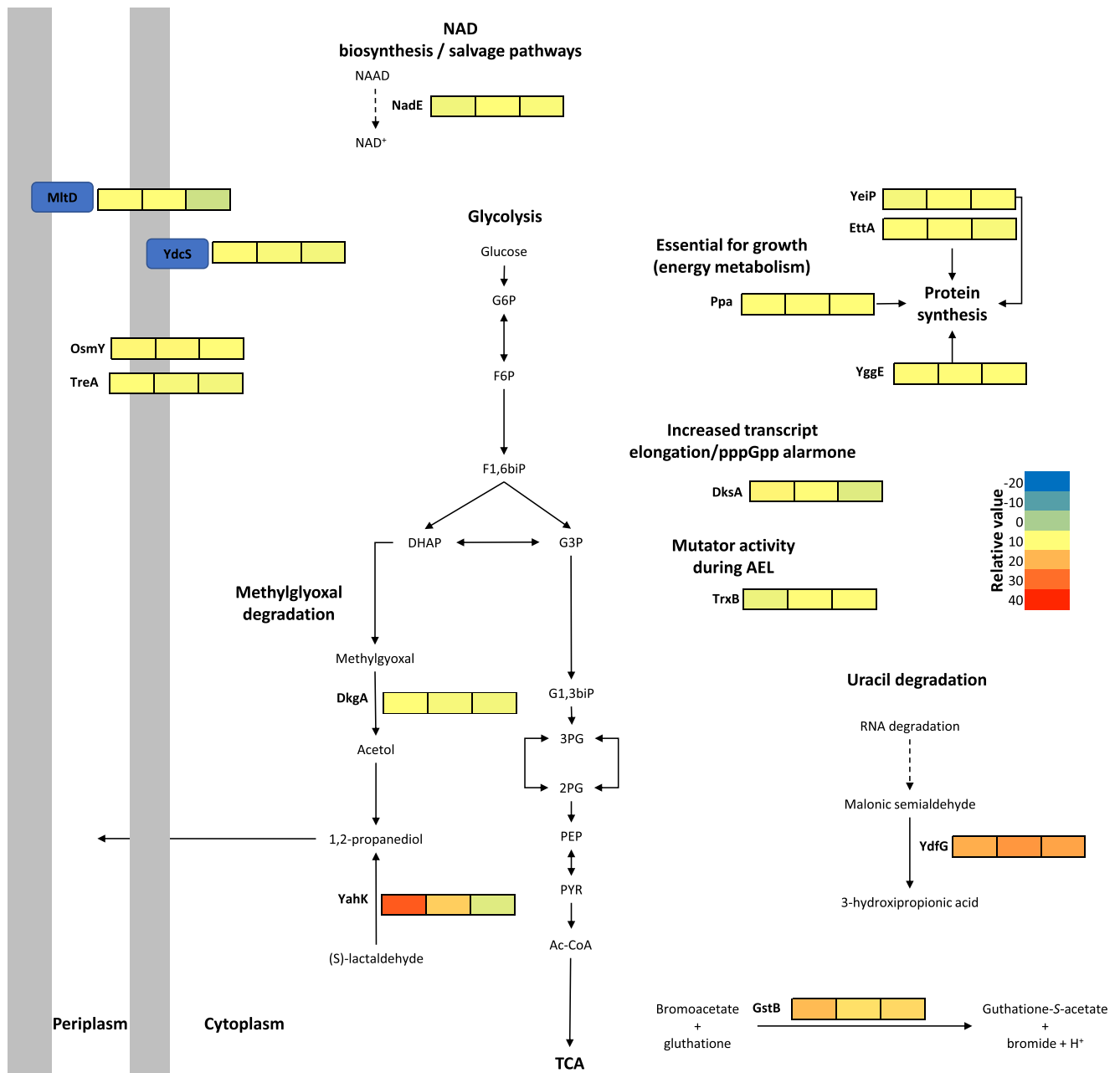
Fig. 3 Proposed metabolic traits in the strains PB11, PB11rppH⁻, and PB12: stress-response proteins, proteins involved in degradative pathways, and other relevant cellular processes. Relative protein concentration values were referred respect the parental strain JM101 and shown as a heat map. Values in squares represent protein concentration detected in the strain PB11 (left), PB11rppH⁻ (middle), and PB12 (right), respectively. Dotted arrows represent diverse reactions in the pathway. Continuous lines represent a unique enzymatic reaction. Metabolic pathways were adapted from the information of the EcoCyc database

Discussion

Analysis of upregulated proteins in response to the inactivation of the PTS in the wild-type strain JM101 (resulting in the PTS⁻ strain PB11); in response of the loss of *rppH* (PB11*rppH*⁻), as in the evolved strain PB12, allowed to identify non-previously reported proteins/enzymes involved in the metabolic traits developed by PB11 during a short-term ALE experiment. The discussion on the biological functions of upregulated proteins in the PTS⁻ derivatives in the following sections highlights the relevance of some of these proteins in glucose, amino acids, ribose transport; amino acid biosynthesis; NAD biosynthesis/ salvage pathway, biosynthesis of Ac-CoA precursors; detoxification and degradation pathways; stress response; protein synthesis; and possible mutator activities between JM101/PB11, PB11/PB11*rppH*⁻, and PB11*rppH*⁻/PB12 comparisons.
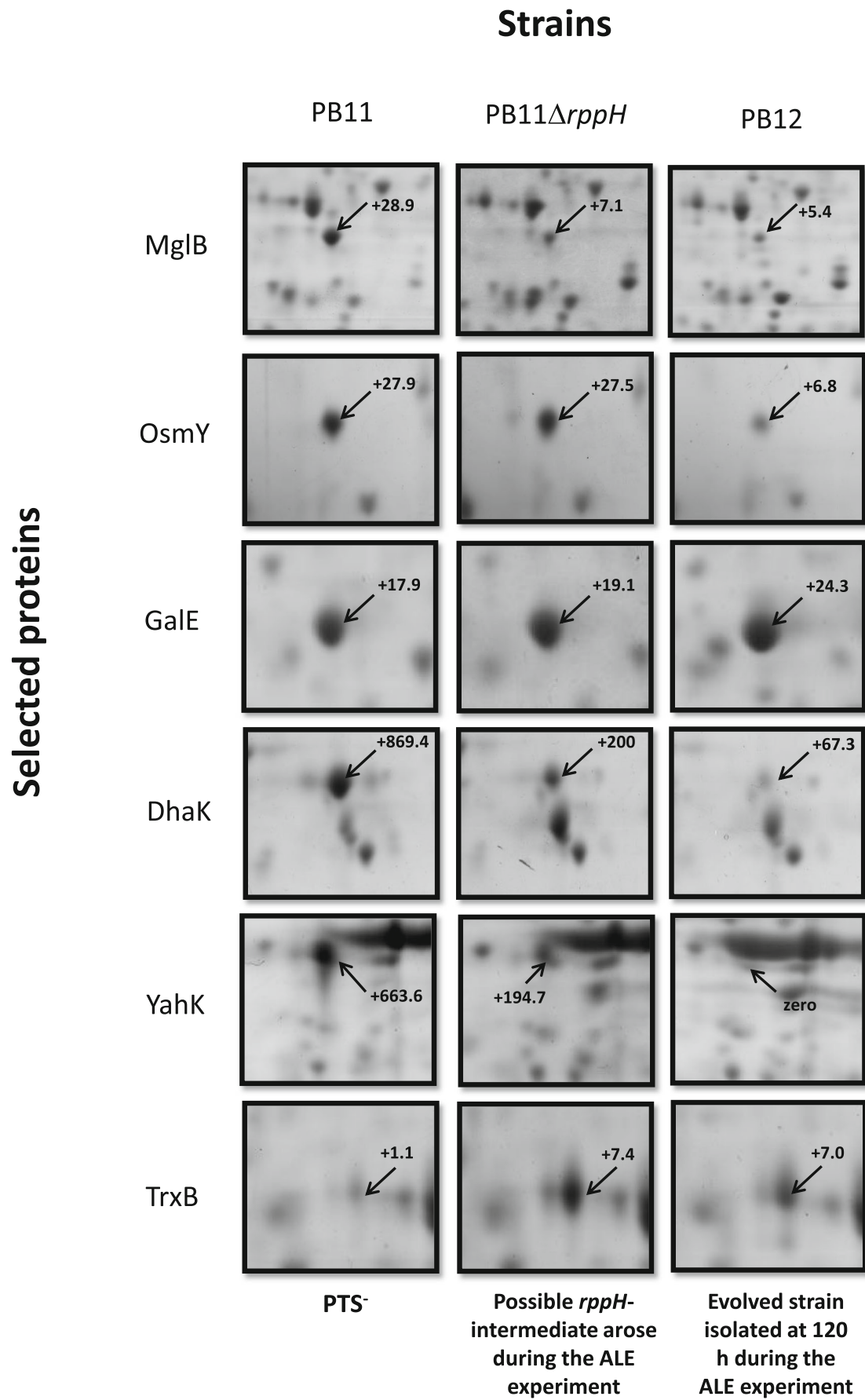
Strains

Fig. 4 Spots of some proteins with abrupt changes between PTS and $rppH^-$ derivatives detected by 2D gel electrophoresis. Effects of the deletion of $rppH$ over the relative concentration of MglB, OsmY, GalE, DhaK, YahK, and TrxB in strains PB11, PB11$rppH^-$, and PB12, with respect to the wild-type strain JM101

Upregulated proteins in response to deletion of the *ptsHIcrr* operon in the wild-type strain JM101 explains how the resulting PTS⁻ strain contends with the scavenging-stress conditions imposed during the early ALE experiment

As reported previously, inactivation of the *ptsHIcrr* operon in the wild-type strain JM101 affected the growing capabilities of the resultant PTS⁻ derivative PB11 drastically, by reducing its μ and qs by ~ 80 and ~ 76%, respectively (Flores et al. 2005, 2007; Aguilar et al. 2012). Upregulated proteins in PB11 in response to the inactivation of PTS suggest the selection of a PTS⁻ derivative population that gradually grew by transport and metabolize glucose efficiently. Under this condition, upregulated proteins involved in the transport of glucose acquired a relevant role, including LamB, MglB, and ManX. MglB, the periplasmic binding protein of the D-galactose/D-galactoside ABC transport system (MglBAC), was the one identified as having a highest relative concentration in PB11 strain with respect to the parental strain JM101 (~ 29-fold higher).

The Mgl system participates in glucose import under conditions of glucose starvation (Ferenci 2001) and LamB mediates sugar uptake by passive diffusion along a concentration gradient with a very high-affinity uptake of maltose and glucose under stress condition and glucose starvation (Death et al. 1993; Dumas et al. 2000). The higher relative concentration of these proteins in PB11 indicates a relevant role as the possible primary transporters used by this strain in the absence of the PTS. Interestingly, upregulation of LamB and MglB agrees with the previously observed upregulation of *lamB* and *mglB* reported for PB11 (Flores et al. 2005). However, GalP, the galactose permease proposed previously as the glucose transporter selected in PB11 due to the higher transcript of *galP* (Flores et al. 2005, 2007; Aguilar et al. 2012), showed no significant upregulation in PB11.

ManX, the subunit of the mannose PTS permease, was upregulated ~ 5-fold in PB11 with respect to JM101. This observation was intriguing since the common cytoplasmic phosphate (from PEP)-carrier proteins EI (*ptsI*), and Hpr (*ptsH*) were deleted in the strain PB11 (Flores et al. 1996), abolishing the activity of the remaining PTS systems in the cell. Interestingly, the upregulation of ManX has been previously reported to be a response to the inactivation of *ptsG* (Steinsiek and Bettenbrock 2012). These authors stated that inactivation of *ptsG* increased the expression of *mgl*, *malE*, and *manX* during the growth of *E. coli* in glucose-limiting conditions. However, these authors showed that inactivation of *mgl* demonstrated that the activity of this galactose ABC transporter (Mgl) is not essential, as

deletion of this gene does not affect the growth capability of derivatives of *E. coli* MG1655 (Steinsiek and Bettenbrock 2012). Based on this explanation, upregulation of ManX in strain PB11 is proposed as an additional glucose transport-associated response to the inactivation of the cytoplasmic components of the PTS.

Several authors have previously reported some of the upregulated proteins in PB11 or the genes coding for these proteins as being upregulated in response to various stress conditions. OmpX was reported as being overexpressed during osmotic, acidic, and alkaline stress (Nakashima et al. 1995; Stancik et al. 2002). OsmY, TreA, and YdcS were reported as being induced during growth under nitrogen limitation and being upregulated during adaptation to glucose-limiting conditions, respectively (Weissbach and Hurwitz 1959; Repoila and Gutierrez 1991; Hengge-Aronis et al. 1991; Yim and Villarejo 1992; Weber et al. 2005; Franchini 2006). The proposed role of these proteins is consistent with the response of strain PB11 to the nutritional-stress condition imposed by the deletion of PTS during the early ALE experiment.

A remarkable metabolic trait developed by strain PB11 (and observed in the evolved strain PB12) is the gluconeogenic capability. As previously reported, this condition allowed the co-utilization of glucose and acetate by strain PB11, resulting in the enhanced yield of aromatic compounds in engineered derivatives of this strain (Sigala et al. 2009; Sabido et al. 2014). Analysis of upregulated proteins showed the upregulation of GatY and Eno involved in gluconeogenic pathways in PB11, which provides evidence to explain this phenotypic trait. Relevant upregulated proteins include DhaK, and DhaL (components of the dihydroxyacetone kinase), which are involved in the degradation of glycerol to DHAP (Fig. 2). GlpK, a glycerol kinase, catalyzes the phosphorylation of glycerol to *sn*-glycerol-3P, which in turn is aerobically or anaerobically converted to DHAP (Keseler et al. 2017) (Fig. 2). YdfG, a protein involved in the uracil degradation pathway, which reduces pyrimidines to β-amino acids, CO_2, and ammonia, was also upregulated (Fig. 3) (Kim et al. 2010). This pathway, known as the Rut pathway, was upregulated in *E. coli* and allows the use of pyrimidines as nitrogen sources at 37 °C under nitrogen-limiting conditions. The last reaction of this pathway converts malonic semialdehyde to 3-hydroxypropinoic acid. RutH performs this reaction, but evidence suggests that the function of RutH overlaps with that of YdfG (Kim et al. 2010).

The upregulation of proteins of the PPP (TktB and TalA) and proteins involved in ribonucleotide degradation (Cdd) and transformation (RfaD) suggests an essential role in the non-oxidative branch of the PPP. The intermediate sedoheptulose-7-P is the precursor for the synthesis of lipopolysaccharides via ADP-L-glycero-β-D-manno-heptose biosynthesis, while D-ribose 5-phosphate indicates a possible association between the PPP and the biosynthesis of histidine (Fig. 2) (Keseler et al. 2017). Upregulation of TktB and TalA is consistent with the previously

observed increase in the transcript of their corresponding genes *tktB* and *talA* in PB11 respect the wild-type strain JM101 (Flores et al. 2005).

GalE, a UDP-glucose 4-epimerase, catalyzes the interconversion of UDP-galactose and UDP-glucose as part of galactose catabolism. The high relative concentration of GalE is consistent with the upregulation of several genes of the *gal* regulon, including *galP*, *galS*, *galE*, *galT*, and *galK*, as determined previously in this strain by a transcriptomic analysis (Flores et al. 2005; Aguilar et al. 2012). These results support that GalE synthesizes galactose as an autoinducer of the *gal* regulon. In turn, this autoinducer inactivates the repressors GalS and GalR, resulting in the induction of other genes of this regulon, including *galP*, as possible alternative glucose transporter in the absence of PTS, allowing the internalization of glucose via this galactose transporter (Ferenci 2001; Flores et al. 2005). However, as discussed above, LamB and MglB were the alternative transporters selected by strain PB11 to internalize glucose as GalP was not found upregulated in the present study.

Interestingly, upregulation of the enzymes involved in the degradation of glycerol and detoxification of methylglyoxal is consistent with a previous report of a pathway involved in the maintenance of the cellular membrane in PB11 (Aguilar et al. 2015). Specifically, this pathway involves the degradation of phosphatidylethanolamine (PtdEtn) to ethanolamine and glycerol-3P, which can be incorporated into both the de novo phospholipid biosynthesis pathway and the glycolysis pathway. Glycerol-3-P is then converted, either aerobically or anaerobically, to DHAP, which in turn can be converted to methylglyoxal or glyceraldehyde-3P by the triose phosphate isomerase (Keseler et al. 2017; McCloskey et al. 2018a). As enzymes involved in DHAP synthesis (DhaK, DhaL, and GlpK) and methylglyoxal detoxification (DkgA) were found to be upregulated in strain PB11, we tested the hypothesis that a significant amount of DHAP is converted to methylglyoxal and then detoxified to 1,2-propanediol in this strain. However, no accumulation of this compound was observed in the culture supernatants (determined by HPLC, data not shown) of strains PB11 and PB12. This result suggests that the cell can metabolize the methylglyoxal entirely to pyruvate, accomplishing the detoxification of this compound, and re-incorporation of this compound into glycolytic intermediates (Cooper 1984), suggesting the utilization of methylglyoxal bypass by PB11.

Upregulated proteins in response to the inactivation of *rppH* in the strain PB11

As shown previously, inactivation of *rppH* in PB11 increased the μ and qs values by 262 and 169%, respectively, compared to the parental strain PB11 (Aguilar et al. 2012). Interestingly, in counterpart, overexpression of *rppH* resulted in the strongest inhibition of cell growth in PTS$^+$ wild-type strain (Choi et al. 2017). Analysis of differentially regulated proteins in the derivative

PB11*rppH*$^-$ showed the downregulation of many proteins observed upregulated in strain PB11, including the OMPs OmpX and LamB and the transporter proteins MglB, ManX, ArgT, and RbsB. As discussed above, the drastic downregulation of the transporters MglB and ManX in the PB11*rppH*$^-$ derivative suggests that other sugar transporters replaced them for the efficient transport of glucose during the ALE experiment. Previous evidence showed that the inactivation of the *mglBAC* component in a PTS$^-$ mutant ($\Delta ptsG$) of *E. coli* MG1655 after the upregulation of the *mgl* operon in glucose-limiting conditions is not essential for growth (Steinsiek and Bettenbrock 2012). This result indicates a non-essential role for this glucose uptake system, suggesting that the activities of the remaining transporters are sufficient for glucose transport.

The observed reduction in the relative concentrations of other membrane-associated, periplasmic, stress-response, and cytoplasmic proteins, which were found to be upregulated in PB11, suggests that the pathways and cellular processes in which these proteins are involved play a relevant physiological role in the scavenging-stress response in strain PB11 but not in the *rppH*$^-$ derivative strain. The decrease in the concentrations of these proteins in the *rppH*$^-$ mutant suggests the presence of a possible global regulatory factor triggered by the deletion of this gene, which regulates the concentrations of proteins involved in various pathways during the transition between PB11 and a possible *rppH*$^-$ derivative during the ALE experiment. Proteins found to be upregulated or unchanged in the PB11*rppH*$^-$ mutant with respect to strain PB11 included those involved in catabolic and biosynthetic pathways and general cellular processes such as glycolysis (GpmA), galactose degradation (GalE) (Fig. 2), biosynthesis of pantothenate as a precursor for CoA (PanB, PanC), synthesis of precursors for lipopolysaccharides (RfaD), histidine biosynthesis (HisG, HisD), NAD biosynthesis/salvage pathways (NadE), uracil degradation pathway (YdfG), and protein synthesis and energy metabolism (YeiP, EttA, and Ppa) (Keseler et al. 2017) (Figs. 2 and 3). Among these proteins, upregulation of GpmA, GalE, and Ppa is also consistent with the previously observed upregulation of their corresponding genes; however, a remarkable increase in the transcript of *galP* was observed in the *rppH*$^-$ derivative respect its parental strain PB11 (Aguilar et al. 2012), but as stated above, GalP was not identified as differentially upregulated in none of the PTS$^-$ strains.

Remarkably, a protein identified as being upregulated in PB11*rppH*$^-$ with respect to PB11 is TrxB (increased by 625%) (Table 3 and Fig. 3), a thioredoxin reductase involved in the thioredoxin pathway. TrxB catalyzes the transfer of electrons from NADPH to thioredoxin (oxidized thioredoxin + NADPH + H$^+$ → a reduced thioredoxin + NADP$^+$). This enzyme was reported to be involved in various cellular processes in *E. coli* (Keseler et al. 2017). TrxB was identified as a member of a network that included more than 90 genes involved in the repair of double-stranded DNA breaks via homologous recombination in unstressed cells without having any mutagenic effect on the

cell. Interestingly, under conditions of stress, the σ^S transcriptional regulator or starvation response factor is activated, and this network switches on a mutagenic mode, promoting mutation and accelerating evolution (Al Mamun et al. 2012).

It is important to highlight that the improved capabilities observed in growth, qs (Aguilar et al. 2012), and changes in differentially upregulated proteins in the PB11$rppH^-$ derivative respect its parental strain PB11 were a response to the deletion of $rppH$. Importantly, the previously proposed hypothesis on the effect of the absence of RppH in the PB11$rppH^-$ derivative was associated to higher mRNA levels of several genes involved in glycolysis, TCA, and respiratory capabilities in this derivative strain (Aguilar et al. 2012). However, our results showed that some upregulated proteins in PB11, such as LamB, MglB, TktB, and TalA (consistent with the previously observed increasing in their corresponding transcripts (Flores et al. 2005; Flores et al. 2007; Aguilar et al. 2012)), were found downregulated in the $rppH^-$ derivative with respect to PB11. Lack of correlation between transcript and proteomic was proposed to relevant post-transcriptional mechanisms (Margalef-Català et al. 2016).

Recent reports about the mechanism of RppH on messenger RNA degradation in *E. coli* showed that this enzyme prefers diphosphorylated forms of mRNAs to the triphosphorylated counterpart. In fact, ~35–50% of *E. coli* mRNAs are diphosphorylated whereas the remainder is mainly monophosphorylated and surprisingly, little triphosphorylated 5′-end form was found (Luciano et al. 2017). A diphosphorylated form of representative RNAs increases to ~100% in RppH$^-$ mutants (Luciano et al. 2018). Conversion from triphosphorylated 5′-ends to the monophosphorylated form to induce 5′-end-dependent degradation by RNase E is a two-step process involving an unidentified enzyme responsible for the conversion from the triphosphorylated form to the diphosphorylated form, which is subsequently converted to the monophosphorylated form by RppH, concluding about mRNA stability, that absence of RppH not prevent the RNA degradation (Luciano et al. 2017).

Metabolic traits of the evolved derivative strain PB12

Decreased relative protein concentration of OMPs (OmpX, LamB), osmotic stress-response proteins (YdcS, TreA, and OsmY), proteins involved in cell division/protein secretion (MtlD), and proteins involved in glucose transport (MglB, ManX) were observed in the evolved strain PB12 respect the PB11$rppH^-$ derivative (Figs. 2 and 3). These results suggest possible substantial changes on the cell surface, in transport capabilities, and in the stress response of the evolved strain PB12 with respect to PB11 and PB11$rppH^-$. These

results agree with the previous report on the modified cell envelope in response to deletion of RppH (Choi et al. 2017). GalE showed a marked upregulation in this strain with respect to PB11 and PB11$rppH^-$ (Table 4 and Figs. 3 and 4). As stated above, the upregulation of GalE was proposed as result of the constitutive production of the autoinducer galactose in strain PB12 (as in PB11), leading to the induction of most of the genes in the *gal* regulon (including *galE*). Additionally, upregulation of GalE in PB12 is consistent with the increased transcript level of *galE* observed previously (Flores et al. 2005).

Other upregulated protein found in the evolved strain PB12 was PanC, which catalyzes the final reaction in the synthesis of pantothenate, the precursor of CoA. The relevance of this pathway is unclear in PB12 because, as shown in Fig. 2, the enzyme PflB, which is involved in the synthesis of Ac-CoA from CoA + PYR, was downregulated with respect to PB11 and PB11$rppH^-$. HisG and HisD, which participate in the biosynthesis of histidine, were upregulated in the evolved strain PB12. The biological relevance of the upregulation of these proteins is unclear; the proteinogenic role of histidine is, of course, of vital importance, but no other significance for these proteins was reported for *E. coli*. Other Gram-negative bacteria, including *Salmonella enterica*, possess the Hut pathway (histidine utilization pathway), which leads to the synthesis of glutamate; however, an orthologous pathway is absent in *E. coli* (Bender 2012).

The increased relative protein concentration of TrxB in strain PB12 with respect to PB11 is relevant (Table 4 and Figs. 3 and 4). Upregulation of this protein, together with the deletion of *mutH*, suggests a correlation with the high mutation rate in PB12 compared to JM101 and PB11, as determined previously by the Luria-Delbrück fluctuation test (Aguilar et al. 2015). This behavior agrees with the suggested mutator phenotype exhibited by the PB11$rppH^-$ derivative as a possible metabolic intermediate during the ALE experiment performed to select for faster-growing derivatives. As discussed above, it was proposed that during the first 75 h of the short-term ALE experiment, the PB11 strain lost a > 10-kb chromosomal region, leading to the deletion of 12 genes, including *rppH*, *mutH*, and *galR* (Aguilar et al. 2012) (Fig. S2). The deletion of *mutH*, coding for the MutH endonuclease, which is part of the MutHLS complex involved in the DNA mismatch repair process, was proposed to be the event responsible for the development of further mutations in this strain during a short-term ALE experiment (Aguilar et al. 2015). The proposed role of TrxB as a member of a network promoting an accelerated evolution phenotype under conditions of stress provides evidence that the loss of *mutH* and, possibly, the upregulation of TrxB drives the mutator phenotype in strain PB11 during the ALE experiment, resulting in the selection of the evolved strain PB12. However, further evidence is necessary to corroborate the proposed role of TrxB in the derivative PTS$^-$ strains.

Effect of inactivation of GalP and upregulation of Glk on the transport and phosphorylation of glucose in the PTS⁻ strains

MglABC and GalP are the primary routes for galactose transport in *E. coli* K12 (Keseler et al. 2017). These systems were reported as upregulated in several strains of *E. coli* and act as promiscuous glucose transporters when PTS⁻ derivatives are cultivated with glucose as the sole carbon source (Ferenci 2001; Flores et al. 2005, 2007; Steinsiek and Bettenbrock 2012). As described above, upregulation of LamB and MglB in PB11 suggested that this strain uses these proteins preferentially to transport glucose in the absence of the PTS, but previous transcriptomic analysis has suggested a relevant role for GalP in glucose transport in the PTS⁻ strains PB11, PB11*rppH⁻*, and PB12 (Flores et al. 2005; Aguilar et al. 2012). Deletion of *galP* allowed the determination of the relevance of the permease GalP in glucose transport in the PTS⁻ strains PB11 and PB12. As the elimination of this permease does not have an apparent adverse effect on the μ of the PB11Δ*galP* mutant (Table 2), this result corroborates the proposed role of MglB as the principal glucose transporter in this strain. However, in the PB12Δ*galP* derivative, deletion of this gene reduced the μ value of the strain by 54% compared to that of PB12 (Table 2). This result showed that glucose transport in the evolved strain PB12 is mainly dependent on GalP instead of MglB, which is consistent with the 81% decrease observed in the relative concentration of MglB in PB12 with respect to that in PB11 (Fig. 2). However, our results showed no significant changes in the relative concentration of GalP in both PB11 and PB12. As stated above, changes in relative concentration of MglB observed between PB12 and PB11 (Table 3 and Figs. 2 and 4) are also consistent with the results of a previous report on the elimination of *mglBAC* in a *ptsG⁻* mutant of *E. coli* MG1655 after the upregulation of this operon without an adverse impact on glucose transport (Steinsiek and Bettenbrock 2012). As GalP depends on a sugar-proton symport mechanism to transport glucose, while MglABC does not (Keseler et al. 2017), differences in the roles of Mgl/GalP associated with the upregulation of OmpX (induced by acidic conditions) (Stancik et al. 2002) between strains PB11, PB11*rppH⁻*, and PB12 suggest differences in the acidic conditions of the periplasm. However, the physiological relevance of this proposed effect on the growth recovery capability of these strains remains to be determined.

Finally, cloning of *glk* into a low copy number plasmid (pCL1920) followed by transformation into the PB11Δ*galP* mutant led to a 67 and 75% increase in μ and qs values, respectively, in the derivative strain PB11Δ*galP*/pCL*glk* with respect to the parent strain PB11Δ*galP*. These results suggest an impaired glucose phosphorylation capability of Glk in this strain, limiting the growth of the strain. However, the expression of *glk* in the PB12Δ*galP*/pCL*glk* strain did not have any effect on the μ but resulted in a slight increase in the qs value (~ 107%) with respect to the parent strain PB12Δ*galP*. A previous transcriptomic

analysis showed *glk* to be upregulated in PB12 but not in PB11 (Flores et al. 2005). This evidence indicates that the acquired overexpression of *glk* by PB12 is enough to phosphorylate the transported glucose, minimizing the contribution to phosphorylation by the cloned *glk*.

As concluding remarks, characterization of the differentially upregulated proteins in response to the PTS deletion (PB11 and PB12) and the *rppH* deletion (PB11*rppH⁻*) led the identification of proteins involved in glucose transport, catabolic and biosynthetic pathways, and other relevant cellular processes, many of them non-previously reported even in other transcriptomic or proteomic studies. Remarkably, the analysis of differentially expressed proteins in response to the PTS⁻ and *rppH⁻* mutations showed a lower correlation with previously reported transcriptomic results for these strains, particularly for some glycolytic, TCA, gluconeogenic, glyoxylate, *pox* shunt, respiratory, and ppGpp metabolism. Most evident is the lack of correlation between the previously increased transcript of *galP* coding for the galactose permease GalP proposed as the principal alternative transporter selected by the PTS⁻ strains during the ALE experiment (Aguilar et al. 2012; Flores et al. 2005, 2007).

It is relevant to understand the changes both in genotype and phenotype in response to the inactivation of *rppH*, as the evolved strain PB12 has been used successfully for aromatic compounds production purposes (Rodriguez et al. 2014; Martínez et al. 2015). The inactivation of *rppH* is an early event during a short-term ALE experiment proposed to play a crucial role, by improving the growing capabilities of the *rppH⁻* mutant in glucose significantly. As new evidence showed that absence of RppH does not prevent mRNA degradation, we suggest that deletion of RppH could play a role as a central global effector by modulating the concentration of proteins involved in the metabolic traits selected by PB11 during the ALE experiment. Additionally, upregulation of the levels of the thioredoxin reductase TrxB also suggests a crucial possible role associated with potentially increased mutability in the PTS⁻ derivatives during the ALE experiment. Overall, ALE experiment provides cellular solutions for the selection of strains with improved glucose import capabilities and growth solutions in the phenotype *ptsHIcrr⁻* for the production of aromatic compounds (McCloskey et al. 2018b).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Aguilar C, Escalante A, Flores N, de Anda R, Riveros-McKay F, Gosset G, Morett E, Bolívar F (2012) Genetic changes during a laboratory adaptive evolution process that allowed fast growth in glucose to an *Escherichia coli* strain lacking the major glucose transport system. BMC Genomics 13:385. https://doi.org/10.1186/1471-2164-13-385

Aguilar C, Flores N, Riveros-McKay F, Sahonero-Canavesi D, Carmona SB, Geiger O, Escalante A, Bolívar F (2015) Deletion of the 2-acyl-glycerophosphoethanolamine cycle improve glucose metabolism in *Escherichia coli* strains employed for overproduction of aromatic compounds. Microb Cell Factories 14:1–18. https://doi.org/10.1186/s12934-015-0382-6

Al Mamun AAM, Lombardo M-J, Shee C, Lisewski AM, Gonzalez C, Lin D, Nehring RB, Saint-Ruf C, Gibson JL, Frisch RL, Lichtarge O, Hastings PJ, Rosenberg SM (2012) Identity and function of a large gene network underlying mutagenic repair of DNA breaks. Science 338:1344–1348. https://doi.org/10.1126/science.1226683

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25:25–29. https://doi.org/10.1038/75556

Bender RA (2012) Regulation of the histidine utilization (hut) system in bacteria. Microbiol Mol Biol Rev 76:565–584. https://doi.org/10.1128/MMBR.00014-12

Carmona SB, Moreno F, Bolívar F, Gosset G, Escalante A (2015) Inactivation of the PTS as a strategy to engineer the production of aromatic metabolites in *Escherichia coli*. J Mol Microbiol Biotechnol 25:195–208. https://doi.org/10.1159/000380854

Choi U, Park Y-H, Kim Y-R, Seok Y-J, Lee C-R (2017) Effect of the RNA pyrophosphohydrolase RppH on envelope integrity in *Escherichia coli*. FEMS Microbiol Lett 364:1–7. https://doi.org/10.1093/femsle/fnx152

Cooper RA (1984) Metabolism of methylglyoxal in microorganisms. Annu Rev Microbiol 38:49–68. https://doi.org/10.1146/annurev.mi.38.100184.000405

Cooper RA, Anderson A (1970) The formation and catabolism of methylglyoxal during glycolysis in *Escherichia coli*. FEBS Lett 11:273–276. https://doi.org/10.1016/0014-5793(70)80546-4

Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97:6640–6645. https://doi.org/10.1073/pnas.120163297

de Luna-Valdez LA, Martínez-Batallar AG, Hernández-Ortiz M, Encarnación-Guevara S, Ramos-Vega M, López-Bucio JS, León P, Guevara-García AA (2014) Proteomic analysis of chloroplast biogenesis (*clb*) mutants uncovers novel proteins potentially involved in the development of *Arabidopsis thaliana* chloroplasts. J Proteome 111:148–164. https://doi.org/10.1016/j.jprot.2014.07.003

Deana A, Celesnik H, Belasco JG (2008) The bacterial enzyme RppH triggers messenger RNA degradation by 5′ pyrophosphate removal. Nature 451:355–358. https://doi.org/10.1038/nature06475

Death A, Notley L, Ferenci T (1993) Derepression of LamB protein facilitates outer membrane permeation of carbohydrates into *Escherichia coli* under conditions of nutrient stress. J Bacteriol 175:1475–1483

Dumas F, Koebnik R, Winterhalter M, Van Gelder P (2000) Sugar transport through maltoporin of *Escherichia coli*: role of polar tracks. J Biol Chem 275:19747–19751. https://doi.org/10.1074/jbc.M000268200

Escalante A, Salinas Cervantes A, Gosset G, Bolívar F (2012) Current knowledge of the *Escherichia coli* phosphoenolpyruvate–carbohydrate phosphotransferase system: peculiarities of regulation and impact on growth and product formation. Appl Microbiol Biotechnol 94:1483–1494. https://doi.org/10.1007/s00253-012-4101-5

Encarnación S, Hernández M, Martínez-Batallar G, Contreras S, Vargas M del C, Mora J (2005) Comparative proteomics using 2-D gel electrophoresis and mass spectrometry as tools to dissect stimulons and regulons in bacteria with sequenced or partially sequenced genomes. Biol Proced Online 7:117–135. https://doi.org/10.1251/bpo110

Ferenci T (2001) Hungry bacteria–definition and properties of a nutritional state. Environ Microbiol 3:605–611 https://doi-org.biblioteca.ibt.unam.mx:8080/10.1046/j.1462-2920.2001.00238.x

Flores N, Xiao J, Berry A, Bolivar F, Valle F (1996) Pathway engineering for the production of aromatic compounds in *Escherichia coli*. Nat Biotechnol 14:620–623. https://doi.org/10.1038/nbt0596-620

Flores S, Gosset G, Flores N, de Graaf AA, Bolívar F (2002) Analysis of carbon metabolism in *Escherichia coli* strains with an inactive phosphotransferase system by ^{13}C labeling and NMR spectroscopy. Metab Eng 4:124–137. https://doi.org/10.1006/mben.2001.0209

Flores N, Flores S, Escalante A, de Anda R, Leal L, Malpica R, Georgellis D, Gosset G, Bolívar F (2005) Adaptation for fast growth on glucose by differential expression of central carbon metabolism and *gal* regulon genes in an *Escherichia coli* strain lacking the phosphoenolpyruvate:carbohydrate phosphotransferase system. Metab Eng 7:70–87. https://doi.org/10.1016/j.ymben.2004.10.002

Flores N, Leal L, Sigala JC, de Anda R, Escalante A, Martínez A, Ramírez OT, Gosset G, Bolivar F (2007) Growth recovery on glucose under aerobic conditions of an *Escherichia coli* strain carrying a phosphoenolpyruvate:carbohydrate phosphotransferase system deletion by inactivating *arcA* and overexpressing the genes coding for glucokinase and galactose permease. J Mol Microbiol Biotechnol 13:105–116. https://doi.org/10.1159/000103602

Franchini AG (2006) Global gene expression in *Escherichia coli* K-12 during short-term and long-term adaptation to glucose-limited continuous culture conditions. Microbiology 152:2111–2127. https://doi.org/10.1099/mic.0.28939-0

Gosset G (2005) Improvement of *Escherichia coli* production strains by modification of the phosphoenolpyruvate: sugar phosphotransferase system. Microb Cell Factories 4:14. https://doi.org/10.1186/1475-2859-4-14

Hengge-Aronis R, Klein W, Lange R, Rimmele M, Boos W (1991) Trehalose synthesis genes are controlled by the putative sigma factor encoded by *rpoS* and are involved in stationary-phase thermotolerance in *Escherichia coli*. J Bacteriol 173:7918–7924. https://doi.org/10.1128/jb.173.24.7918-7924.1991

Hernández-Montalvo V, Martínez A, Hernández-Chavez G, Bolivar F, Valle F, Gosset G (2003) Expression of *galP* and *glk* in a *Escherichia coli* PTS mutant restores glucose transport and increases glycolytic flux to fermentation products. Biotechnol Bioeng 83:687–694. https://doi.org/10.1002/bit.10702

Higareda-Almaraz JC, del Rocío E-GM, Hernández-Ortiz M, Resendis-Antonio O, Encarnación-Guevara S (2011) Proteomic patterns of cervical cancer cell lines, a network perspective. BMC Syst Biol 5:96. https://doi.org/10.1186/1752-0509-5-96

Himabindu P, Anupama K (2017) Decreased expression of stable RNA can alleviate the lethality associated with RNase E deficiency in *Escherichia coli*. J Bacteriol 199:e00724–e00716. https://doi.org/10.1128/JB.00724-16

Keseler IM, Mackie A, Santos-Zavaleta A, Billington R, Bonavides-Martínez C, Caspi R, Fulcher C, Gama-Castro S, Kothari A, Krummenacker M, Latendresse M, Muñiz-Rascado L, Ong Q, Paley S, Peralta-Gil M, Subhraveti P, Velázquez-Ramírez DA, Weaver D, Collado-Vides J, Paulsen I, Karp PD (2017) The EcoCyc database:

reflecting new knowledge about *Escherichia coli* K-12. Nucleic Acids Res 45:D543–D550. https://doi.org/10.1093/nar/gkw1003

Kim K-S, Pelton JG, Inwood WB, Andersen U, Kustu S, Wemmer DE (2010) The Rut pathway for pyrimidine degradation: novel chemistry and toxicity problems. J Bacteriol 192:4089–4102. https://doi.org/10.1128/JB.00201-10

Luciano DJ, Vasilyev N, Richards J, Serganov A, Belasco JG (2017) A novel RNA phosphorylation state enables 5′ end-dependent degradation in *Escherichia coli*. Mol Cell 67:44–54.e6. https://doi.org/10.1016/j.molcel.2017.05.035

Luciano DJ, Vasilyev N, Richards J, Serganov A, Belasco JG (2018) Importance of a diphosphorylated intermediate for RppH-dependent RNA degradation. RNA Biol 0:1–4. https://doi.org/10.1080/15476286.2018.1460995

Margalef-Català M, Araque I, Bordons A, Reguant C, Bautista-Gallego J (2016) Transcriptomic and proteomic analysis of *Oenococcus oeni* adaptation to wine stress conditions. Front Microbiol 7:1554. https://doi.org/10.3389/fmicb.2016.01554

Martínez K, de AR, Hernández G, Escalante A, Gosset G, Ramírez OT, Bolívar FG (2008) Coutilization of glucose and glycerol enhances the production of aromatic compounds in an *Escherichia coli* strain lacking the phosphoenolpyruvate: carbohydrate phosphotransferase system. Microb Cell Factories 7(1):1. https://doi.org/10.1186/1475-2859-7-1

Martínez JA, Bolívar F, Escalante A (2015) Shikimic acid production in *Escherichia coli*: from classical metabolic engineering strategies to omics applied to improve its production. Front Bioeng Biotechnol 3:1–16. https://doi.org/10.3389/fbioe.2015.00145

Martínez-Gómez K, Flores N, Castañeda HM, Martínez-Batallar G, Hernández-Chávez G, Ramírez OT, Gosset G, Encarnación S, Bolivar F (2012) New insights into *Escherichia coli* metabolism: carbon scavenging, acetate metabolism and carbon recycling responses during growth on glycerol. Microb Cell Factories 11:46. https://doi.org/10.1186/1475-2859-11-46

McCloskey D, Xu S, Sandberg TE, Brunk E, Hefner Y, Szubin R, Feist AM, Palsson BO (2018a) Adaptation to the coupling of glycolysis to toxic methylglyoxal production in *tpiA* deletion strains of *Escherichia coli* requires synchronized and counterintuitive genetic changes. Metab Eng 48:82–93. https://doi.org/10.1016/j.ymben.2018.05.012

McCloskey D, Xu S, Sandberg TE, Brunk E, Hefner Y, Szubin R, Feist AM, Palsson BO (2018b) Adaptive laboratory evolution resolves energy depletion to maintain high aromatic metabolite phenotypes in *Escherichia coli* strains lacking the phosphotransferase system. Metab Eng 48:233–242. https://doi.org/10.1016/j.ymben.2018.06.005

Nakashima K, Horikoshi K, Mizuno T (1995) Effect of hydrostatic pressure on the synthesis of outer membrane proteins in *Escherichia coli*. Biosci Biotechnol Biochem 59:130–132. https://doi.org/10.1271/bbb.59.130

Perederina A, Svetlov V, Vassylyeva MN, Tahirov TH, Yokoyama S, Artsimovitch I, Vassylyev DG (2004) Regulation through the secondary channel—structural framework for ppGpp-DksA synergism during transcription. Cell 118:297–309. https://doi.org/10.1016/j.cell.2004.06.030

Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass

spectrometry data. Electrophoresis 20:3551–3567. https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2

Repoila F, Gutierrez C (1991) Osmotic induction of the periplasmic trehalase in *Escherichia coli* K12: characterization of the *treA* gene promoter. Mol Microbiol 5:747–755 https://doi-org.biblioteca.ibt.unam.mx:8080/10.1111/j.1365-2958.1991.tb00745.x

Rodriguez A, Martinez JA, Flores N, Escalante A, Gosset G, Bolivar F (2014) Engineering *Escherichia coli* to overproduce aromatic amino acids and derived compounds. Microb Cell Factories 13(126):126. https://doi.org/10.1186/s12934-014-0126-z

Sabido A, Sigala JC, Hernández-Chávez G, Flores N, Gosset G, Bolívar F (2014) Physiological and transcriptional characterization of *Escherichia coli* strains lacking interconversion of phosphoenolpyruvate and pyruvate when glucose and acetate are coutilized. Biotechnol Bioeng 111:1150–1160. https://doi.org/10.1002/bit.25177

Salazar E, Díaz-Mejía JJ, Moreno-Hagelsieb G, Martínez-Batallar G, Mora Y, Mora J, Encarnación S (2010) Characterization of the NifA-RpoN regulon in *Rhizobium etli* in free life and in symbiosis with *Phaseolus vulgaris*. Appl Environ Microbiol 76:4510–4520. https://doi.org/10.1128/AEM.02007-09

Sigala JC, Flores S, Flores N, Aguilar C, de Anda R, Gosset G, Bolívar F (2009) Acetate metabolism in *Escherichia coli* strains lacking phosphoenolpyruvate: carbohydrate phosphotransferase system; evidence of carbon recycling strategies and futile cycles. J Mol Microbiol Biotechnol 16:224–235. https://doi.org/10.1159/000151219

Stancik LM, Stancik DM, Schmidt B, Barnhart DM, Yoncheva YN, Slonczewski JL (2002) pH-dependent expression of periplasmic proteins and amino acid catabolism in *Escherichia coli*. J Bacteriol 184:4246–4258. https://doi.org/10.1128/JB.184.15.4246-4258.2002

Steinsiek S, Bettenbrock K (2012) Glucose transport in *Escherichia coli* mutant strains with defects in sugar transport systems. J Bacteriol 194:5897–5908. https://doi.org/10.1128/JB.01502-12

The UniProt Consortium (2010) The universal protein resource (UniProt) in 2010. Nucleic Acids Res 38:D142–D148. https://doi.org/10.1093/nar/gkp84

Wang Q, Zhang D, Guan Z, Li D, Pei K, Liu J, Zou T, Yin P (2018) DapF stabilizes the substrate-favoring conformation of RppH to stimulate its RNA-pyrophosphohydrolase activity in *Escherichia coli*. Nucleic Acids Res 46:6880–6892. https://doi.org/10.1093/nar/gky528

Weber H, Polen T, Heuveling J, Wendisch VF, Hengge R (2005) Genome-wide analysis of the general stress response network in *Escherichia coli*: SigmaS-dependent genes, promoters, and sigma factor selectivity. J Bacteriol 187:1591–1603. https://doi.org/10.1128/JB.187.5.1591-1603.2005

Weissbach A, Hurwitz J (1959) The formation of 2-keto-3-deoxyheptonic acid in extracts of *Escherichia coli* B. J Biol Chem 234:705–709

Yim HH, Villarejo M (1992) *osmY*, a new hyperosmotically inducible gene, encodes a periplasmic protein in *Escherichia coli*. J Bacteriol 174:3637–3644. https://doi.org/10.1128/jb.174.11.3637-3644.1992

BMC Systems Biology

RESEARCH ARTICLE

Open Access

CrossMark

Metabolic modeling and response surface analysis of an Escherichia coli strain engineered for shikimic acid production

Juan A. Martínez[1] ⬤, Alberto Rodriguez[1], Fabian Moreno[1], Noemí Flores[1], Alvaro R. Lara[2], Octavio T. Ramírez[3], Guillermo Gosset[1] and Francisco Bolivar[1]*

Abstract

Background: Classic metabolic engineering strategies often induce significant flux imbalances to microbial metabolism, causing undesirable outcomes such as suboptimal conversion of substrates to products. Several mathematical frameworks have been developed to understand the physiological and metabolic state of production strains and to identify genetic modification targets for improved bioproduct formation. In this work, a modeling approach was applied to describe the physiological behavior and the metabolic fluxes of a shikimic acid overproducing *Escherichia coli* strain lacking the major glucose transport system, grown on complex media.

Results: The obtained flux distributions indicate the presence of high fluxes through the pentose phosphate and Entner-Doudoroff pathways, which could limit the availability of erythrose-4-phosphate for shikimic acid production even with high flux redirection through the pentose phosphate pathway. In addition, highly active glyoxylate shunt fluxes and a pyruvate/acetate cycle are indicators of overflow glycolytic metabolism in the tested conditions. The analysis of the combined physiological and flux response surfaces, enabled zone allocation for different physiological outputs within variant substrate conditions. This information was then used for an improved fed-batch process designed to preserve the metabolic conditions that were found to enhance shikimic acid productivity. This resulted in a 40% increase in the shikimic acid titer (60 g/L) and 70% increase in volumetric productivity (2.45 gSA/L*h), while preserving yields, compared to the batch process.

Conclusions: The combination of dynamic metabolic modeling and experimental parameter response surfaces was a successful approach to understand and predict the behavior of a shikimic acid producing strain under variable substrate concentrations. Response surfaces were useful for allocating different physiological behavior zones with different preferential product outcomes. Both model sets provided information that could be applied to enhance shikimic acid production on an engineered shikimic acid overproducing *Escherichia coli* strain.

Keywords: Metabolic modeling, Central carbon metabolism, Response surface analysis, Cybernetic modeling, Shikimic acid

Background

The aromatic amino acid pathway (AAAP) branches from the central carbon metabolism (CCM) by the aldolic condensation of erythrose-4-phosphate (E4P) and phosphoenolpyruvate (PEP), being present in bacteria and plants. The AAAP is responsible for the production of aromatic amino acids and aromatic vitamins. As a consequence, it is an essential and highly regulated pathway [1, 2]. AAAP intermediates and final compounds play important roles in the pharmaceutical and food industries, either as raw materials, additives or final products [3–9]. Among them, shikimic acid (SA) can be used as an enantiomeric precursor to produce valuable biological molecules such as antipyretics, antioxidants, anticoagulants, antithrombotics, anti-inflammatories, analgesic agents, antibacterial, hormonal

*Correspondence: bolivar@ibt.unam.mx
Francisco Bolivar is a member from El Colegio Nacional, México
[1]Departamento de Ingeniería Celular y Biocatálisis, Instituto de Biotecnología, Universidad Nacional Autónoma de México (UNAM), Avenida Universidad 2001, Colonia Chamilpa, 62210 Cuernavaca, Morelos, México
Full list of author information is available at the end of the article

Martínez et al. BMC Systems Biology   (2018) 12:102

Page 2 of 26

or antiviral compounds [8, 9]. SA was at first produced from the seed of the Chinese star anise plant *Illicium verum*, employing classic extraction processes with yields of only 30 mg/Kg approximately [10–12]. For this reason, over the past years, many studies concerning SA production have focused on recovery technologies, chemical synthesis methods and biotechnological production using different microorganisms [9, 13, 14]. The latter resulted in many genetically engineered strains that produce SA at laboratory and industrial scales with relatively high yields (between 40–50% mol/mol), but still far from the theoretical maximum (86% mol/mol) [2, 9, 13–15].

Although classic metabolic engineering (ME) allows flux redirection in a biochemical network into valuable compounds by genetic manipulation, it often induces significant flux imbalances to the CCM that may cause undesirable outcomes. These imbalances can disrupt precursor availability and energy balances, causing the accumulation of pathway intermediates and unwanted byproducts, reducing strain fitness and product yields [16]. These imbalances derive from alterations to the complex connectivity of biological information networks (genome, transcriptome, proteome, and metabolome) [17, 18]. Therefore, there is an increasing interest into a more global and detailed understanding of the metabolic and regulatory network changes imposed by different genetic modifications or process conditions in various production systems. In recent years, mathematical models, advances on informatics and the availability of big and more precise *omics* data sets have proved useful to resolve and clarify the complex network interactions and system characteristics [19–22].

To mathematically model metabolism, a metabolic network must be assembled with sufficient detail and curated from genomic data to be represented as a matrix of equations, including all available stoichiometric, thermodynamic and kinetic data. Given the complexity of microorganisms, the parameter sets required to describe the networks for genome-scale models are quite large and require informatically-intensive modeling approaches [23]. Most of the constructed metabolic models use mass balances and assume pseudo-steady state conditions to solve the highly undetermined linear equation systems and render a convex space, which contains all the possible solutions for the system. This solution space then must be narrowed with experimental data and some other assumptions to acquire a meaningful and useful solution [18, 19, 21, 24]. Different approaches have been developed to find the most meaningful solution, such as mechanism-based, interaction-based and the constraint-based methodologies. The latter, are the most commonly used for their capability to render useful flux distributions, even with relatively small amounts of information [20, 23]. Nevertheless, a challenging ground for models still exists

for high-throughput data acquisition and interpretation when non-defined cultivation media and dynamic processes are used. The challenges and achievements within this field can be consulted elsewhere [20, 21, 25–31].

Regarding SA production with *E. coli*, few modeling studies could be found in the literature. Chen et al. (2011) [32] used a constraint-based analysis with flux balance analysis (FBA), assumed no growth and used SA as the objective function, to design modifications for the overproduction of AAAP intermediates. The model identified *aroF*, *tktA*, *ppsA* and *glf* genes as candidates for overexpression. As well, suggested the inactivation of *ldhA* and *ackA* genes to avoid carbon waste through lactate (LA) and acetate (AC) fluxes. These genes and nodes are in accordance to other reports on AAAP intermediate production [2, 5, 7, 14]. Nevertheless, this model also identified the non-evident *zwf* gene as critical for redirection of the carbon flux into E4P on the AAAP. Its overexpression resulted in an increase of 47% molar conversion of glucose (GLC) to aromatic intermediates [32, 33]. Similarly, Ahn et al. (2008) [34] constructed a model for maximizing SA production from GLC highlighting the importance of CCM genes like *tktA* and *zwf*, although growth or maintenance requirements were not considered. Rizk and Liao (2009) [35] used ensemble modeling, a mechanism-based approach, to identify *tktA* as the first-rate controlling step, founding that the *ppsA* gene can only augment production of aromatic intermediates when *tktA* is simultaneously overexpressed. There still are several challenges that must be addressed regarding model construction and implementation. For example, models are often limited by specific assumptions, defined conditions and are performed primarily under stationary constraints. Importantly, the assumption of stationary state provides only limited information on the dynamic properties of the system or network regulation. These limitations can result in some contradictions to real cell behavior under changing conditions, given by the existence of complex regulatory mechanisms modifying metabolic fluxes. New models and tools accounting for more complex solutions and on dynamic conditions, would result in a better understanding of cell behavior and produce new insights for strain and bioprocess design. On the other hand, for *E. coli* strains constructed for SA production, most of the work done has been focused on testing and improving expression platforms, genetic backgrounds, including the use of strains lacking the main phosphoglucotransferase transport (PTS), which lack catabolite repression and can redirect part of the carbon flux in to the production of aromatic compounds [2, 5, 7, 9, 33, 36], and culture strategies using traditional engineering approaches. Only few studies have focused on metabolic modeling to better understand and engineer SA overproduction at a more global level. Even less has

been done on modeling production strains under complex media or on dynamic conditions, which are critical considerations for further process improvement. Here, a dynamic modeling approach of the physiological behavior and the dynamic metabolic flux distributions for an engineered *E. coli* strain is presented. The results were useful for strain behavior characterization and SA productivity enhancement on variable complex media compositions.

Results

Physiological characterization, parametrization and modeling of strain AR36 on variant substrate conditions

Figure 1 depicts the results from the 9 experimental design fermentations with the central point done by triplicate along with constructed physiological models (see "Methods" section). Central experimental condition (100:30 GLC:YE g/L) average parameters and deviations are summarized in Table 1. The standard deviations show relatively low values in accordance to experiments using yeast extract (YE) from three different batches. The largest standard deviations corresponded to final SA produced ($[SA]_f$), final consumed GLC ($\Delta[GLC]$) and the exponential consumption rate (q_s^{exp}). Nevertheless, the averaged model depicts a fair agreement with data as can be observed in Fig. 1. The observed behavior and statistical data proved that the logistic models were suitable to describe and parametrize the consumption of GLC and production of SA in strain AR36, within the boundaries of the experimental design. Statistical validation and accuracy of the models are presented in Additional file 1.

With the parameters obtained, three-dimensional response surfaces were constructed (Fig. 2 and "Methods" section). Maximum biomass (X_{max}) response surface (Fig. 2a) shows only small increases with higher GLC concentrations at similar amounts of YE. $\Delta[GLC]$ response surface (Fig. 2b) depicts that GLC consumption increases proportionally with higher starting GLC and YE concentrations. This is especially observed under high GLC concentrations, where at least ≈40 g/L of YE are required for complete exhaustion of more than ≈110 g/L of GLC. Regarding final SA concentration ($[SA]_f$), surface morphology is similar to the consumed GLC surface, but exhibits a maximum critical point at 110:40 g/L GLC:YE initial concentrations (Fig. 2c). For kinetic parameters, GLC consumption rate at exponential phase (q_{glc}^{exp}) shows a saddle type behavior on its response surface (Fig. 2e). This morphology is characterized by the existence of a maximum critical point for GLC and a simultaneous minimum for YE, found at 96 g/L and 37 g/L concentrations, respectively. These results suggest that cellular responses to GLC concentrations lower than ≈75 g/L (increasing consumption) or higher than ≈100 g/L (decreasing consumption), may be occurring in strain AR36. Surfaces

also showed that q_{glc}^{exp} highest values are found at lower concentrations of YE and GLC and the lowest rates under high concentrations of initial GLC. For the SA exponential production rate (q_{sa}^{exp}) surface, a tendency to increase towards lower initial [YE] was found, with an up to 50% decrease when more than 40 g/L of YE are utilized (Fig. 2e and f). The growth rate (μ_{max}) displays a minimum critical point on 105:21 g/L GLC:YE initial concentrations (Fig. 2k) with the highest values found towards lower [GLC] in combination with higher [YE]. Finally, the AC production rate (q_{ac}^{exp}) shows a tendency to present higher values as [YE] and [GLC] increase (Fig. 2i) and could be responsible for reducing biomass and SA production rates as the AC highest rates were found above ≈40 g/L [YE] and ≈110 g/L GLC. In summary, all the specific rates at exponential phase suggest an allocation of rate maximization zones or quadrants on the experimental design as follows: at high [GLC] and high [YE] concentrations AC production is predominant, at low [GLC] and high [YE] concentrations biomass production is predominant, at high [GLC] but low [YE] concentrations SA production is predominant and finally at lower concentrations of both substrates a more balanced growth and production of all final products is to be found (Fig. 2).

At the stationary phase, a reduced metabolic activity on all consumption and production rates was observed. q_{glc}^{sta} surface (Fig. 2g) tends to have larger values on higher initial concentration of substrate sources (GLC and YE). The SA stationary production rate (q_{sa}^{sta}) surface (Fig. 2h) reveals a tendency to increase towards low GLC with high YE initial concentrations, showing an opposite behavior than q_{ac}^{sta} (Fig. 2j). Their surface analysis helps to allocate predominant stationary phase output zones as follows. A SA production zone found above an imaginary diagonal line cutting the experimental design area from low initial concentrations of both substrate sources to high initial concentrations and a predominantly AC production zone found below this imaginary diagonal. It should be also noted that zone preferences on stationary phase are found on opposite sides respective to the allocated ones on the exponential phase. More so, SA specific production rates observed at higher initial [YE] and lower initial [GLC] conditions seem to have smaller variations between phases and AC specific production rates seem to vary less on low initial [YE] high initial [GLC] conditions.

The descriptive viability of the constructed response surfaces was validated by performing fermentations using three conditions not included in the experimental design (75:20, 80:40 and 115:45 GLC:YE initial conditions). Figure 3 shows the results for the logistic growth model and the consumption/production integrated models rendered with the surface calculated parameters: X_{max}, μ, q_{glc}^{exp}, q_{sa}^{exp}, q_{glc}^{sta}, q_{sa}^{sta} and SA_{final} parameters. As

Fig. 1 Physiological model approximation to each point of the experimental design
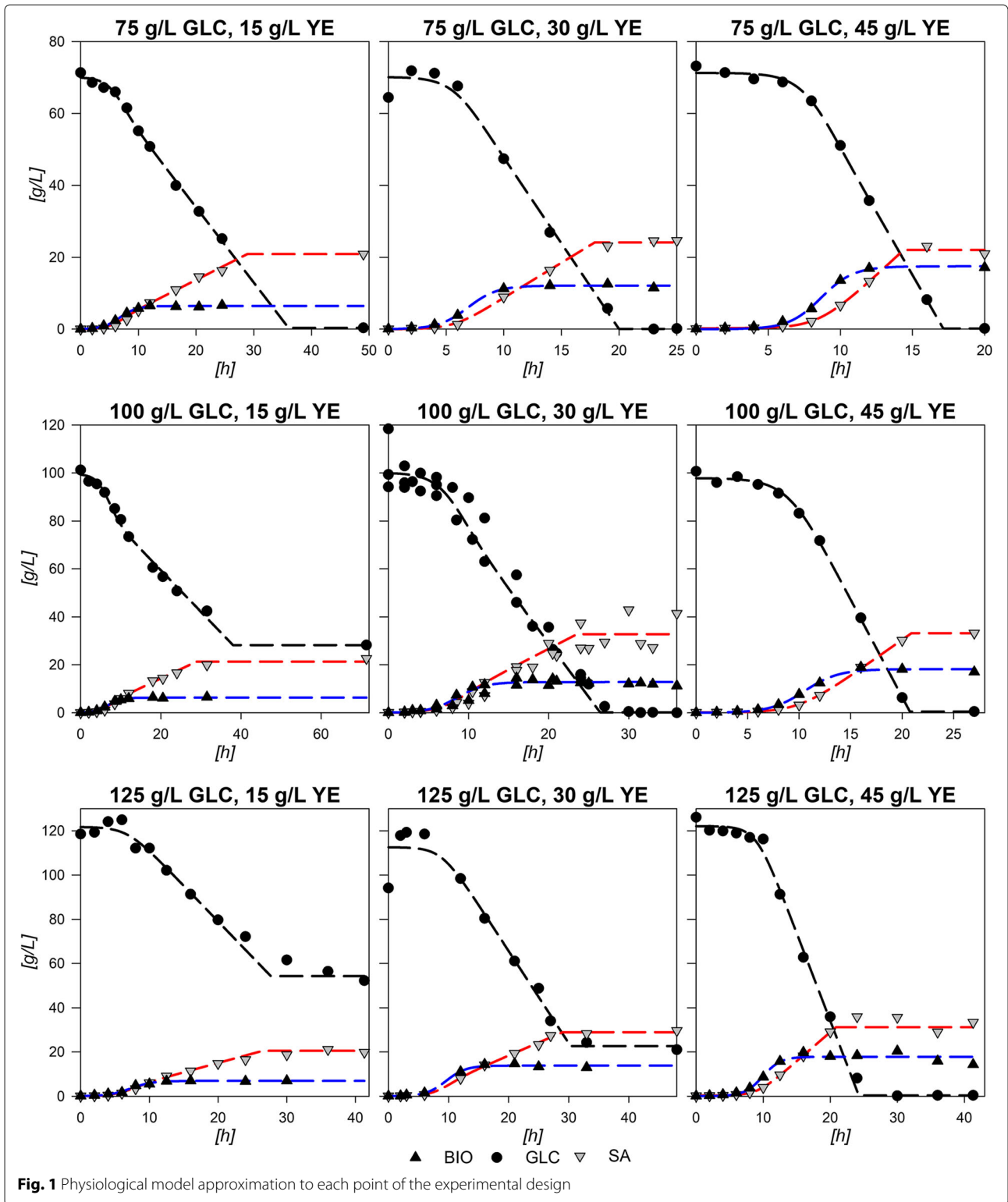
it can be seen, all models follow the experimental data with good agreement. The largest observable deviation is on the maximum SA achievable on the 80:40 GLC:YE

experiment, probably due to the contribution from YE to SA production. In addition, mathematical assessment of the validation was performed by a set of descriptive

Table 1 Average parameters and estimated standard deviations from the central point in the experimental design

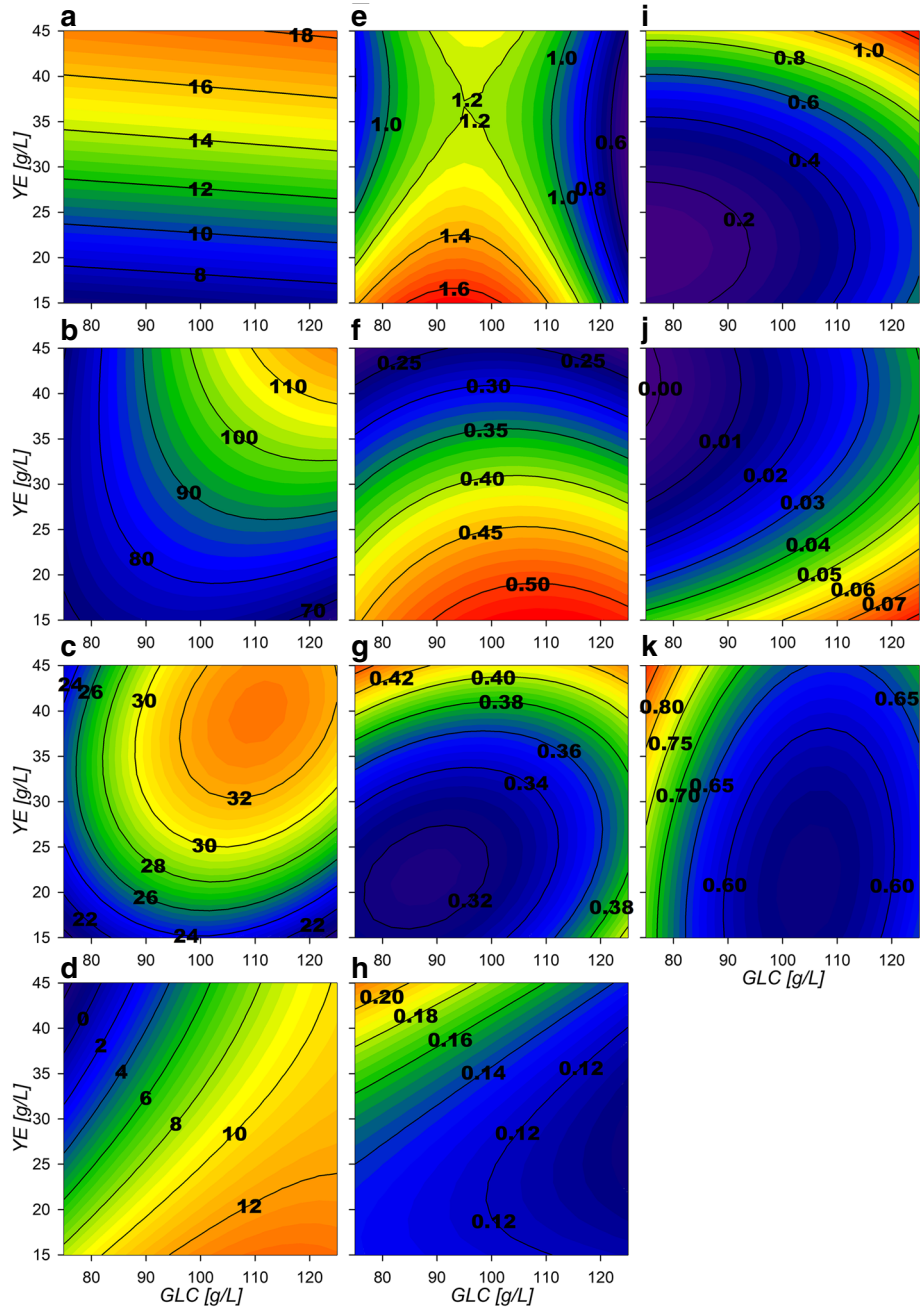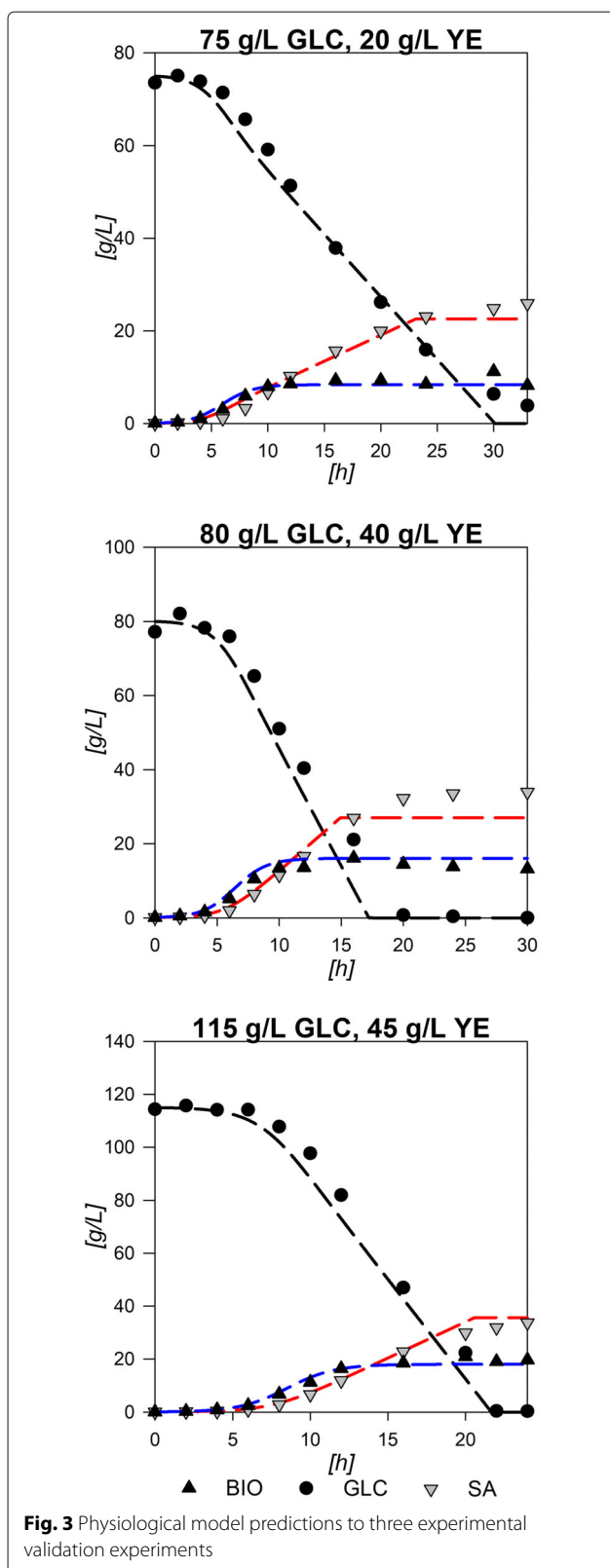|	X_{max}	$\Delta[GLC]$	$[SA]_f$	$[AC]_f$	μ	q_{glc}^{exp}	q_{sa}^{exp}	q_{glc}^{sta}	q_{sa}^{sta}	Y_{ps}	Y_{px}	q_{ac}^{exp}	q_{ac}^{sta}
	[g/L]	[g/L]	[g/L]	[g/L]	[h^{-1}]	[g/Lh]	[g/Lh]	[g/Lh]	[g/Lh]	[g/g]	[g/g]	[g/Lh]	[g/Lh]
Mean	12.80	99.82	32.80	9.24	0.58	1.24	0.35	0.45	0.13	0.36	0.76	0.52	0.50
σ^2	0.77	4.77	8.13	0.95	0.12	0.48	0.09	0.19	0.08	0.03	0.16	0.19	0.19

Fig. 2 Response surface contour plots for the model estimated parameters. **a** Max biomass [g/L], **b** $\Delta[GLC]$ [g/L], **c** Final SA [g/L], **d** Final AC [g/L], **e** q_{glc}^{exp} [g/Lh], **f** q_{sa}^{exp} [g/Lh], **g** q_{glc}^{sta} [g/Lh], **h** q_{sa}^{sta} [g/Lh], **i** q_{ac}^{exp} [g/Lh], **j** q_{ac}^{sta} [g/Lh], **k** μ_{max} [h^{-1}]

Martínez *et al. BMC Systems Biology*   (2018) 12:102

Page 6 of 26

Fig. 3 Physiological model predictions to three experimental validation experiments

and inferential statistical comparisons between the modeled data and the experimental results. Error percentage was obtained by ratio of quadratic sums and presented values from 0.14 to 0.91 for biomass comparisons, from 0.04 to 0.08 for GLC and from 0.04 to 0.26 for SA, suggesting a relatively small deviation between experimental and modeled data along the fermentation. R^2 values were found between 0.96 and 0.99 for all three curves with percentile deviations from the expected slope (SPD) values lower than 1% and *p*-values below 0.05. These statistical values and the depicted models from Fig. 3 show that the models constructed by the surface predicted parameters can render comprehensively good representations for biomass, GLC and SA for initial conditions within the range of the experimental design. Surface predicted parameters were also validated by comparison with the ones calculated directly from experimental data. Table 2 shows the experimental and modeled parameters for all validation experiments and the average error calculated. The individual experimental error between predicted and calculated parameters can be found on Additional file 1. [AC]$_0$ presented the highest error, probably because in experiments with high initial [GLC] and [YE] no AC was produced on stationary phase and surfaces constructed with the polynomial equation cannot properly render these behavior values. q_{glc}^{exp} and $Y_{p/s}$ also had relatively high errors above 10%. This may be related to the contribution of YE since it does not only contains the aromatic amino acids needed for growth, but also other amino acids and some carbohydrates that could contribute to some of the previously discussed effects. However, the two-tailed *t*-student test for the comparison of experimental and modeled parameters validated all parameters as similar, with *p*-values over 0.05. This means that the predicted values can be used within reason to compare and study the behavior of AR36 under the limits of the experimental design and that the constructed surfaces can be used to obtain further insights on cell behavior. Parameter data and statistical values for all experiments can be found on Additional file 1.

Dynamic Flux Metabolic Modeling of AR36 strain on variant substrate conditions

To get further insight into these different output zoned behaviors, dynamic flux models were constructed. It is evident that AR36 strain regulation showed no linear borders and contributions between predominant outcomes. Since data on internal fluxes, constraints on regulation or other kinetic data were not available, a cybernetic modeling approach was used (See "Methods" section). The simplified metabolic network used for the metabolic models is depicted on Fig. 4, names of reactions will be referred onward as indicated in this figure. The complete description of the reactions can be found on Additional

Table 2 Experimental vs Response surface predictions and statistical values calculated from fermentations used for model validation

initial GLC:YE g/L		75:20		80:40		115:45		%Error
		Exp.	Model	Exp.	Model	Exp.	Model	Average
Max biomass	[g/L]	9.30	8.40	14.30	16.03	19.76	18.05	7.18
Consumed GLC	[g/L]	70.03	70.70	80.11	78.19	115.76	113.84	1.18
Final SA	[g/L]	24.62	22.13	33.19	26.62	31.84	32.96	7.88
Final AC	[g/L]	5.25	6.45	0.00	0.85	4.90	8.73	35.70
μ_{max}	[h^{-1}]	0.62	0.73	0.78	0.77	0.55	0.65	8.56
q_{glc}^{exp}	[g/Lh]	0.89	1.17	0.93	0.97	0.73	0.99	17.03
q_{sa}^{exp}	[g/Lh]	0.36	0.42	0.32	0.28	0.24	0.24	7.35
q_{glc}^{sta}	[g/Lh]	0.33	0.32	0.38	0.39	0.41	0.42	2.07
q_{sa}^{sta}	[g/Lh]	0.15	0.13	0.18	0.18	0.15	0.15	2.88
Y_{ps}	[g/g]	0.41	0.36	0.35	0.29	0.34	0.23	14.49
Y_{px}	[g/g]	0.59	0.59	0.42	0.38	0.44	0.38	5.68

file 2. Calculations over this network resulted in dynamic models which followed the extracellular experimental data points with good agreement in all cases, as shown in Fig. 5. It should be noted that in this case, even the behavior of AC could be accurately described. The main characteristics of the common AC profile for fermentations start with an AC production section until approximately the middle of the exponential growth phase, only to be completely consumed in almost all fermentations towards the end of growth. A second AC production section starts at the stationary phase on all experiments except for the ones with 75:30 and 75:45 g/L GLC:YE initial conditions. For the models, values between 0.15 to 2.28% error were found for biomass approximations, from 0.56 to 2.3% error for GLC, from 0.25 to 4.19% error for SA profiles and 0.35 to 10.53% error for AC models in comparison to experimental data. All R^2 from Pearson linear regressions were found to be above 0.9 and their significance *p*-values were all found to be below 0.05. Regarding SPD, the highest values were found for GLC and AC profiles. Specifically, a 21% deviation was found for GLC in the 100:15 condition, where the model presents higher GLC consumption at the last part of the fermentation compared to the experimental values. As it can be seen on Fig. 5, on this particular condition model almost exhaust GLC but experiment presents a final GLC value of 156 mM, which means that model over estimates GLC consumption on this particular condition. In comparison, in all other cases, models tend to underestimate the consumption rate on the last part of the fermentations with values ranging from 1.07 to 19% SPD, where the highest deviations corresponded to fermentations with greater initial YE concentrations. On that regard, on 75:15 g/L the previously observed underestimation of consumption rates at late stationary phase for the other experimental design conditions could mean that GLC may be exhausted

on a time prior to the model estimations. Regarding AC, SPD deviations ranged from 1.31 to 9% in all cases, except for 100:30 and 100:45 where values where 21 and 42% respectively. For the 100:45 condition, this overestimation is due to the error in the AC peak found on mid exponential phase and to the lack of AC production in stationary phase. These large deviations can be explained by taking into account that YE contribution was simplified to only consider it as a biomass precursor and to provide simultaneously glutamate (GLU), alanine (ALA) and aromatic amino acids (taken as one individual metabolite). Nevertheless, the mathematical values along with the observed model behaviors depicted on Fig. 5 suggest that the models constructed are viable approximations to the observed strain behavior under the experimental conditions. All statistical data on the dynamic flux models are available in Additional file 1.

Calculated fluxes were normalized against GLC consumption derived fluxes for their analysis and surface construction on three different fermentation stages: initial exponential (IEx), mid exponential (MEx) and mid stationary (MST). IEx and MEx presented highly similar behaviors, so their description is similar and only IEx surfaces were addressed. However all surfaces and contour plots for all reactions and time sets can be found in Additional file 3.

Central carbon metabolism flux distribution behavior during growth

Selected CCM genes related to IEx flux response surfaces are presented in Figs. 6 and 7. Glycolytic surfaces under growth conditions show the same morphology from glucose-6-phosphate (G6P) to PEP reactions, a saddle critical point with greater relative fluxes at low GLC initial conditions. Pgi (Fig. 6a), Pfk and Fba flux surfaces describe the same morphological behavior as GalP and
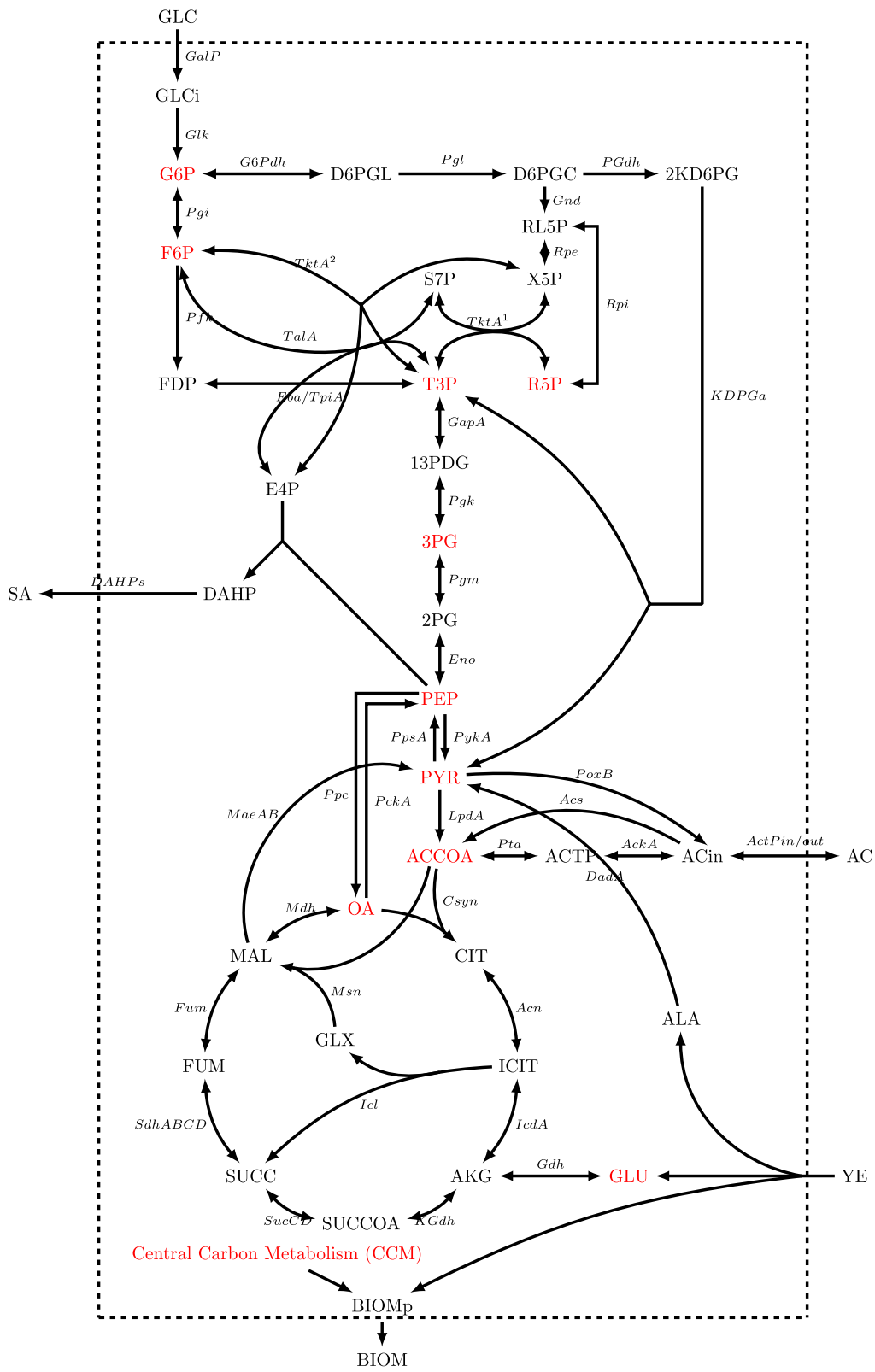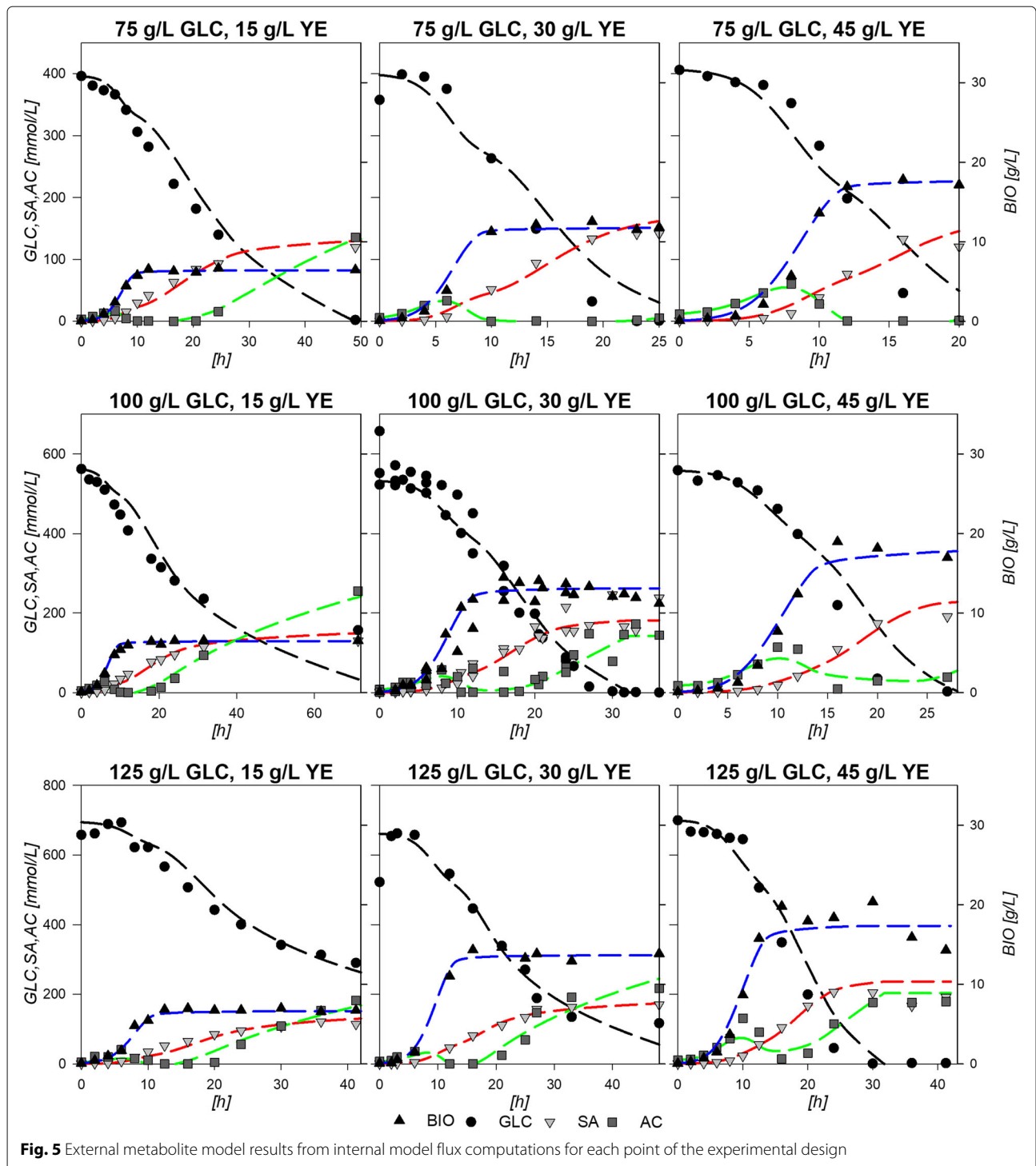
Fig. 4 Central carbon metabolism constructed metabolic network for dynamic metabolic flux models. Metabolites on red refer to the CCM intermediaries used to produce biomass precursor (BIOMp)

Fig. 5 External metabolite model results from internal model flux computations for each point of the experimental design

Glk flux surfaces, but have smaller values than expected, accounting for only ≈6 to 25% of the flux relative to Glk (Additional file 3). This means that the majority of the flux is predicted to enter the oxidative reactions of the pentose phosphate pathway (PPP) by G6Pdh coded

by *zwf* and the 6-phosphogluconolactonase (Pgl). G6Pdh presents relative flux values from ≈75 to 95% (Fig. 6e) and its morphology presents the inverse features than Pgi, which presents greater relative flux values at higher [GLC] and [YE] initial conditions. High relative flux values of
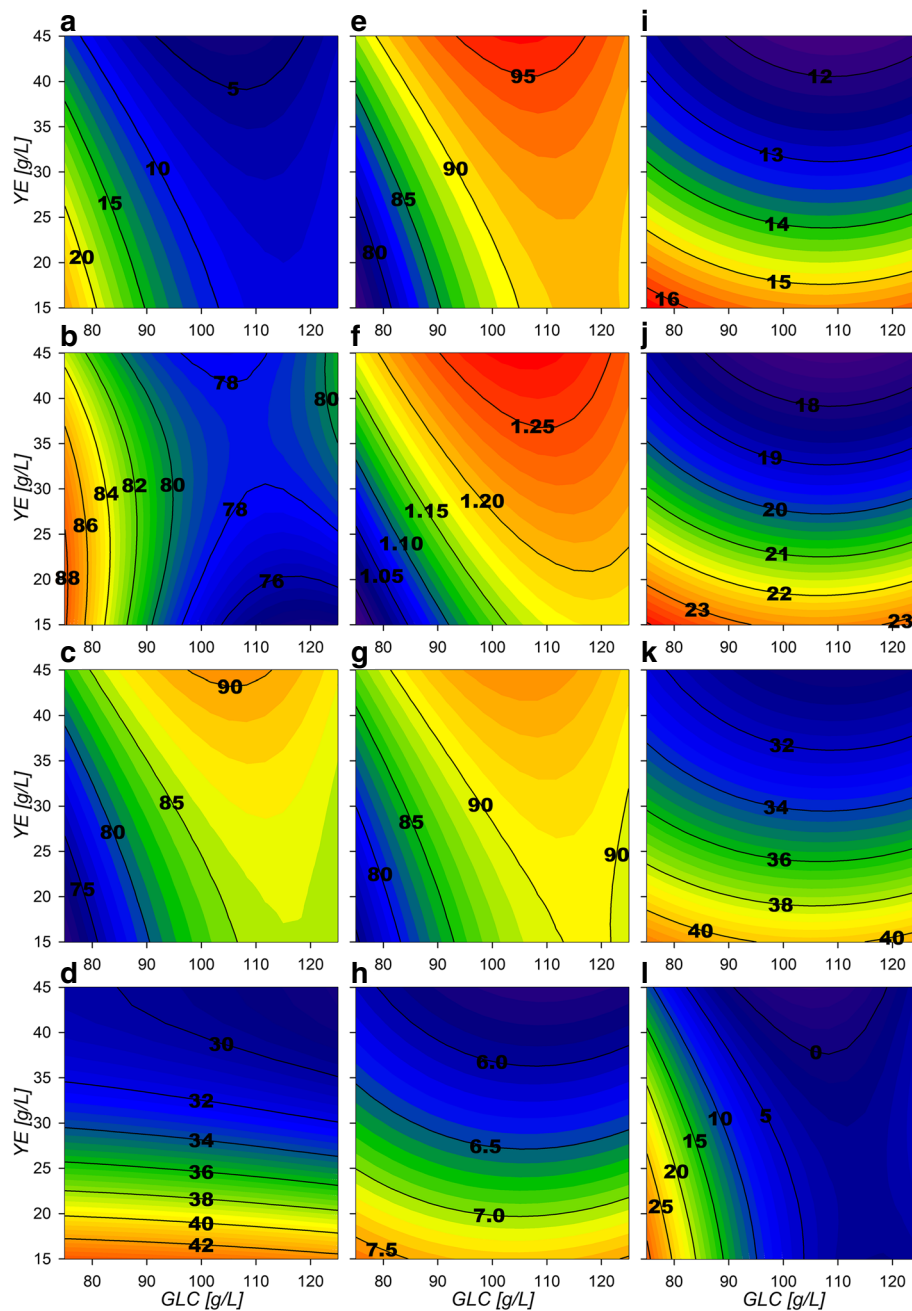
Fig. 6 Response surface contour plots for the estimated internal fluxes at IEx (% Flux relative to GLC consumption). **a** Pgi, **b** GapA, **c** PykA, **d** LpdA, **e** G6Pdh, **f** Gnd, **g** PGdh, **h** TktA 1, **i** TktA 2, **j** DAHPs, **k** PckA, **l** Ppc

75 to 88% were found for reactions from glyceraldehyde-3-phosphate dehydrogenase-A (GapA) (Fig. 6b) and the following Embden-Meyerhoff-Parnas pathway (EMP) reactions towards PEP. Their surfaces have a similar morphological behavior as the upstream glycolytic fluxes but with higher values, suggesting that even with small Pgi flux distributions, high total conversion rates of glucose to PEP can still be present. Also, the flux distributions calculated on the 6-phosphogluconate dehydrogenase(Gnd)/phosphogluconate dehydratase (PGdh) node, showed relative flux values from ≈75 to 94% going through the Entner-Doudoroff pathway (EDP) (Fig. 6f and g respectively). This suggests that most of the carbon flux going through to the oxidative PPP is redirected towards glycerol-3-phosphate (G3P) and pyruvate (PYR).
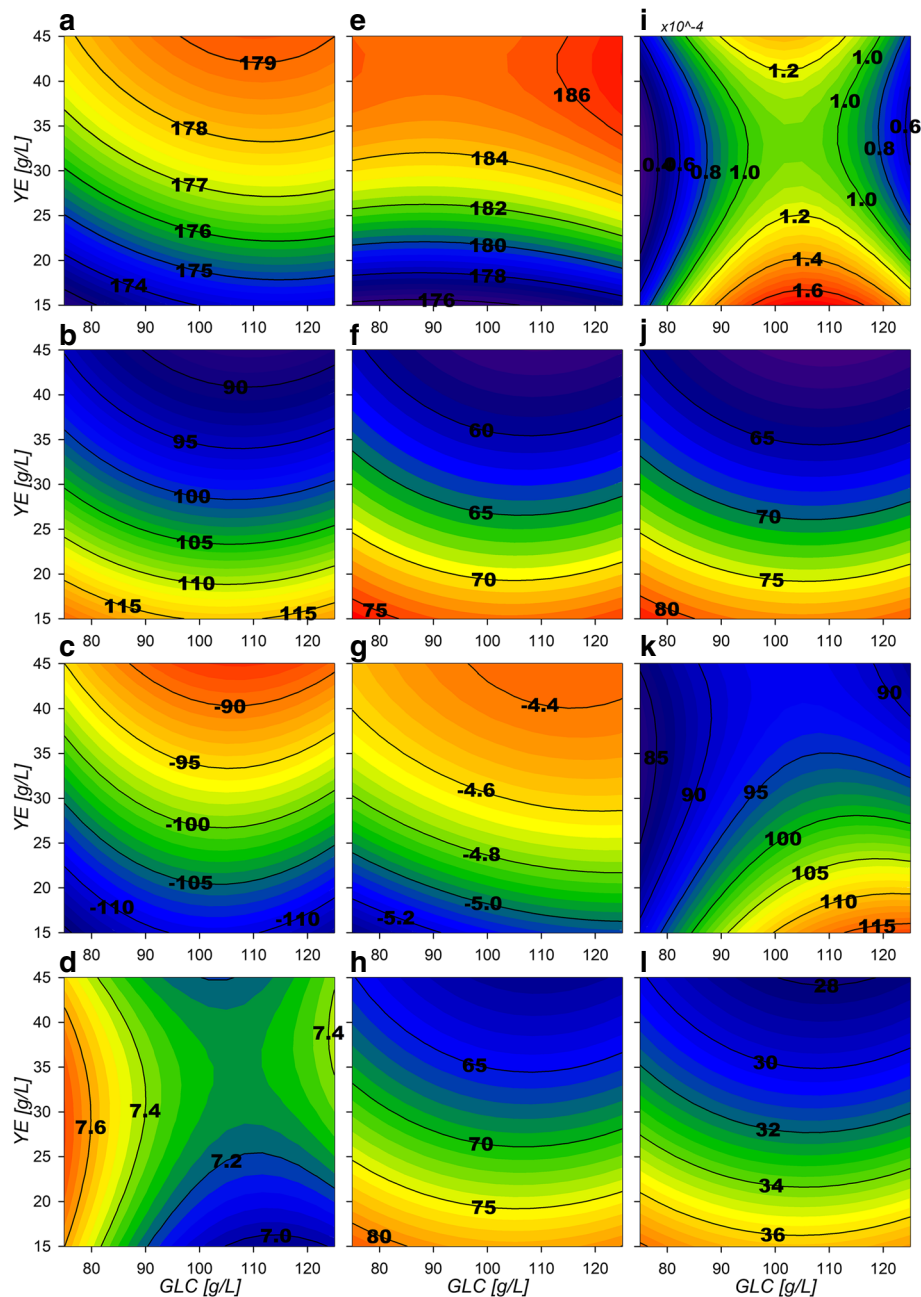
Fig. 7 Response surface contour plots for the estimated internal fluxes at IEx pt2 (% Flux relative to GLC consumption). **a** ActPout, **b** ActPin, **c** AckA, **d** Acs, **e** PoxB, **f** Csyn, **g** IcdA, **h** Icl, **i** KGdh, **j** SdhABCD, **k** Mdh, **l** MaeB

Pgl, Gnd, PGdh and 2-Keto-3-deoxy-6-phosphogluconate aldolase (KDPGa) flux surfaces show the same morphology as G6Pdh, a marked tendency of maximization towards higher initial [YE] (Fig. 6 e–g). In these conditions, the maximum biomass production zone was found that is in agreement to the previously observed high oxidative PPP flux distribution. They also show the

inverse morphology than the EMP fluxes surfaces, as they are expected to compete for carbon skeletons. Regarding the non-oxidative reactions of the PPP, transketolase I (TktA) and transaldolase (Tal) flux surfaces (Fig. 6h) display low relative flux values (≈6 to 8%). Their surface morphology shows a tendency to increase towards lower initial [YE]. TktA surface representing the fraction

of carbon being redirected from EMP towards the non-oxidative branch of PPP, shows values ranging from ≈ 12 to 16% (Fig. 6i). These non-oxidative branch reactions are responsible for the E4P formation and present combined flux values from ≈ 18 to 24%, which matches the predicted flux being redirected toward SA production by the 2-dehydro-3-deoxyphosphoheptonate aldolase (DAHPs) during growth (Fig. 6j). An interesting consequence is that up to ≈72 to 90% of flux modeled is going through the pyruvate kinase II (PykA) related reaction (Figure 6c), meaning that the majority of PEP is probably being converted to PYR. Phosphoenolpyruvate carboxylase (Ppc) surface shows the same morphology described by the glycolytic genes (Fig. 6l), indicating that on low [GLC] and low [YE] conditions, glycolytic metabolism is favored. Meanwhile, Phosphoenolpyruvate carboxykinase (PckA) (Fig. 6k) presents a tendency to increase flux towards low initial [YE] conditions. PckA and Ppc fluxes presented values ranging from ≈31 to 41% and ≈0 to 30%, respectively. Their simultaneous flux, suggests the existence of an ATP consuming futile cycle.

The high inflow to PYR is probably caused by the high EDP and PykA relative fluxes and increased even further by a high malic enzyme carbon reincorporation from the Tricarboxylic Acid Cycle (TCA), accounting for ≈28–38% from the NADPH dependent enzyme (MaeB) (Fig. 7l) and ≈0-30% from the NADH dependent (MaeA). In the model, PYR can also be produced from YE-derived ALA conversion by alanine D-amino acid dehydrogenase (DadA) reaction. On the other hand, for PYR conversion to acetyl coenzyme-A (ACCOA), the reaction was attributed to pyruvate dehydrogenase (LpdA). This reaction showed relatively small values, from ≈28–42% of relative flux (Fig. 6d) compared to pyruvate oxidase (PoxB), which presented fluxes towards AC calculated to be between ≈176–186% during growth phase (Fig. 7e). It is noticeable that for AC production, no constraint was imposed for flux preference on either acetate kinase (AckA), acetyl-CoA synthetase (Acs) and PoxB reactions and the model renders consumption over the reversible (AckA) since it is energetically favorable compared to Acs (Fig. 7c–e). Surfaces for extracellular AC export and import fluxes for AR36 (ActPout and ActPin) show greater export rates with higher initial YE concentrations. This could be attributed to the introduction of carbon to the CCM through ALA and GLU consumption (Fig. 7a and b), but can also be extended to other YE-derived amino acids catabolized through TCA not included on the model. On the other hand, the import of AC presents a maximization tendency towards low initial [YE] with relative flux values between ≈ 90 to 115%. To clarify the node distribution around PoxB, an AR36Δ*poxB* strain was constructed and cultured under

high [GLC] and high [YE], conditions that maximize AC production according to the response surfaces. Interestingly, the initial AC concentration peak observed in all previous experiments was not detected in this case with the mutant strain (Additional file 4). Furthermore, the final AC concentration was significantly lower compared to AR36 on similar fermentation conditions. This suggest that PoxB could be indeed the main contributor to AC production in the AR36 PTS⁻ strain [2, 33, 36–38]. The AR36Δ*poxB* cultures also showed lower growth rates (0.21 h⁻¹) and lower exponential GLC consumption rates (0.61 g/gh) (Additional file 4). This may indicate that its inactivation could be causing PYR accumulation and less ATP generation via the electron-transfer chain [39].

Regarding TCA behavior, the *glta* coded citrate synthase (Csyn) reaction presents relative flux values from ≈56 to 75% (Fig. 7f). As expected aconitase (Acn) reaction (Additional file 3) presents the same behavior as the Csyn reaction and both present the inverse morphological features compared to the AC producing surfaces. Conversely, the following reaction by isocitrate dehydrogenase (IcdA) seems to not be sending carbon flux down TCA. On the contrary, its reversible reaction is found, transporting the small excess of GLU derived from YE consumption towards isocitrate (ICIT) (Fig. 7g). The isocitrate lyase (Icl) and malate synthase (Msn) from this pathway, having relative fluxes values accounting from ≈60 to 82% of relative flux (Fig. 7h), and present the same surface morphology as the TCA carbon uptake Csyn flux surface. On the other hand, 2-ketoglutarate dehydrogenase (KGdh) and succinyl-CoA synthetase (SucCD) complexes seem to be catalyzing very small amounts of flux towards succinate (SUC) (Fig. 7i). Calculations for the glutamate dehydrogenase (Gdh) show relative fluxes between ≈3.1 to 3.7%, suggesting only small input by [YE] components into TCA and apparently processed mainly by IcdA. This means that SUC, is mostly produced by the glyoxylate shunt pathway (GSP) and subsequently catalyzed to malate (MAL) by the succinate dehydrogenase complex (SdhABCD) (Fig. 7j) and the fumarase (Fum). Their surfaces share the morphological characteristics of the Csyn and the GSP surfaces. In consequence, the malate dehydrogenase (Mdh) reaction exhibits higher relative flux values, from ≈85 to 1125% (Fig. 7k) as it also assimilates ACCOA carbon derived from the Msn reaction on GSP.

Surface morphologies suggest the allocation of different predominant extracellular production zones at the exponential growth phase. A SA production predominance flux zone is found at low initial [YE] and as the initial [GLC] diminishes a more balanced production towards biomass and SA is found. This follows up to the predominant region for biomass found at low initial [GLC]

conditions. Finally, a clear AC predominant production zone is found at high initial [GLC] and [YE].

Central carbon metabolism flux distribution behavior after the growth phase

As in IEx and MEx phases, a high flux distribution towards the PPP was found for the MST phase, presenting about 97–108% relative flux through G6Pdh and Pgl (Fig. 8e).

Both oxidative PPP reactions presented a marked minimization morphology towards low initial [GLC] and high [YE] initial concentrations and depicting decreasing ring like border lines. In contrast to the exponential phase, Gnd reaction shows values of relative flux from ≈40 to 100% (Fig. 8f), whereas its competing PGdh accounts for ≈0–70% of relative flux towards EDP. Both morphologies present inverse behavioral surface features, as observed
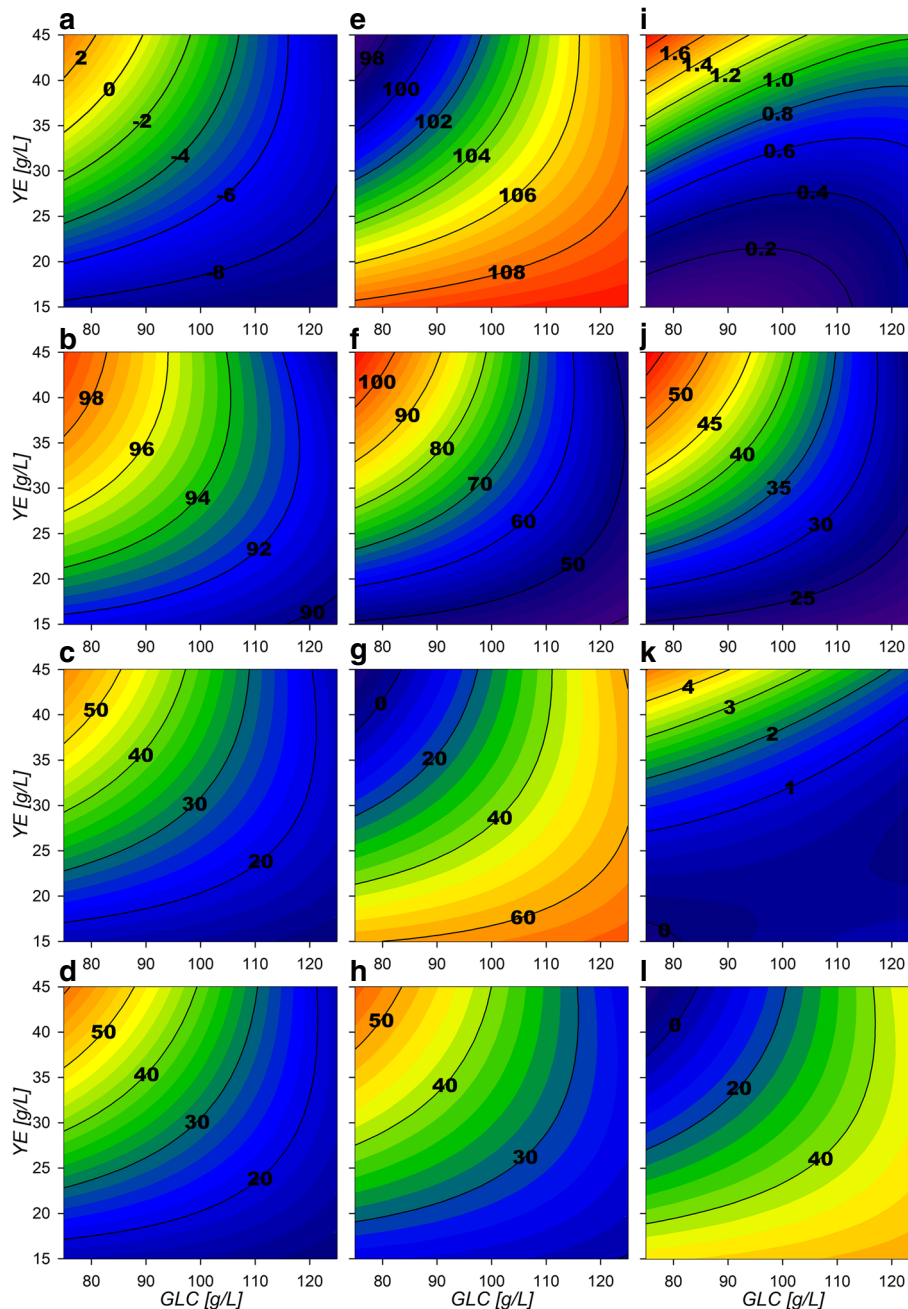

Fig. 8 Response surface contour plots for the estimated internal fluxes at MSt (% Flux relative to GLC consumption). **a** Pgi, **b** GapA, **c** PykA, **d** LpdA, **e** G6Pdh, **f** Gnd, **g** PGdh, **h** TktA 1, **i** TktA 2, **j** DAHPs, **k** PckA, **l** Ppc

Martínez *et al. BMC Systems Biology* (2018) 12:102

Page 14 of 26

on Fig. 8e and f. Interestingly, PGdh surface has the same max region on its morphology as G6Pdh surface, suggesting that excess flux could be still being processed by the EDP. Gnd surface morphology depicts a ringed type tendency with greater values towards initially low [GLC] and high [YE] experimental conditions (up-left corner of the experimental design). As expected, TktA reaction towards the non-oxidative PPP (Fig. 8h) and Tal reactions, exhibit the same behavior as the flux through Gnd surface and become greater contributors to the production of E4P and F6P. The latter enters the glycolytic EMP and it is mostly redirected down the glycolytic pathway through Pfk and Fba reactions with a relative fluxes of ≈10–55%. They also show a ringed surface morphology that tends to maximize towards low [GLC], high [YE] initial conditions. In fact, this morphology was observed through all following glycolytic reactions towards PEP formation and present values from ≈90 to 99% (Fig. 8b). Interestingly, the model renders a small (≈0–9%) unexpected flux distribution of carbon through Pgi (Fig. 8a), redirecting F6P to G6P, just to be consumed again through G6Pdf on almost all the experimental design area. Consequently, Pgi presents only a small flux in the standard G6P to F6P direction at low [GLC] and high [YE] initial conditions corner, representing only ≈0–2%.

On the PEP node reactions, the phosphoenolpyruvate synthase (PpsA) reaction is non-existent, suggesting no gluconeogenic flux from PYR towards PEP is obtained in any condition. To the same extent, PckA reaction catalyzing carbon flux from oxaloacetate (OA) towards PEP is quite low with relative fluxes between ≈0–5% and only being present under high [YE] initial conditions (Fig. 8k). On the other hand, its counterpart reaction Ppc shows ≈0–50% relative flux values, depicting inverse surface morphology features compared to the PckA surface, maximizing towards low initial [YE] conditions and towards higher [GLC] and also with a ringed behavior (Fig. 8l). Subsequently, PEP consumption by PykA and DAHPs presented the same ringed maximization tendency towards low initial [GLC] and high [YE] conditions as observed on the glycolytic surfaces. In contrast to the observations made under growth conditions, both of these fluxes have an equilibrated flux distribution along their surfaces with values between ranging between ≈15–55% (Fig. 8c and j), probably because of the higher E4P production on the PPP. These reactions compete with the AC production reactions, in specific with PoxB, which exhibits higher relative flux values as higher initial [GLC] and lower initial [YE] conditions are set on fermentation (Fig. 9e). Therefore, presenting the inverse surface morphological behavior compared to PykA and DAHPs surfaces. This may be explained as on higher initial concentrations of [YE] more biomass is produced and therefore more [GLC] is consumed by the start of stationary

phase, which means that less [GLC] is expected at this time and therefore, metabolic overflow is expected to be lower. The export modeled transport reaction follows PoxB flux surface behavior (Fig. 9a) as it is observed to be again the main AC producing reaction. On the other hand, import reaction presents almost ≈0 flux values on low initial [YE], with consumption of extracellular AC only towards the low [GLC] with high [YE] initial conditions corner (Fig. 9b). These results suggest that the futile carbon cycling on the AC pathways is found on this stage only under high [YE] conditions [2, 33, 36–38, 40].

Regarding TCA and GSP, their surfaces exhibit ringed type maximization or minimization morphologies towards the upper left corner of the experimental design. Specifically, Csyn and Acn reactions present ≈54 to 70% relative fluxes with the minimization morphology behavior towards low [GLC] and high [YE] initial experimental conditions corner (Fig. 9f). The GSP fluxes follow the same morphological behavior along the experimental design (Fig. 9h).

In contrast to the growth phase, IcdA, KGdh and SucCD reactions present flux directions towards GLC oxidation on all the experimental area, with relative fluxes between ≈10 to 55% (Fig. 9g, i, j), and with its surface morphology maximizing towards low [GLC], high [YE] initial conditions. SdhABCD and Fum follow the same behavior of the Csyn surface. In addition, as higher fluxes are pulled through the TCA, higher is the MaeB reaction flux which competes with Mdh (Fig. 9l and k respectively). Their surface morphology suggest carbon skeleton recycling from PYR, flowing through AC pathways and into TCA to PYR again. This behavior is found under high [GLC] substrate conditions with higher metabolic flux saturation zones.

The dynamic cybernetic model on this phase showed two predominant production zones. The observations made by the dynamic flux models allocate a SA production zone above an imaginary diagonal line, cutting the experimental design area from low to high initial substrate concentration, and a predominantly AC production zone was found below the same imaginary diagonal, in accordance to the physiological models surface allocations.

Bioprocess design for SA productivity enhancement on strain AR36

To assess the utility of the previously described models and considerations towards SA production enhancement, a fed-batch fermentation process was performed with initial conditions of 80 g/L and 40 g/L initial [GLC] and [YE] . Surfaces revealed higher growth rates have been found under high [YE] and low [GLC] conditions (Fig. 2k). Under these conditions there are also zones with lower final AC production and mid range SA production, and high biomass production (Fig. 2c, d and a) for the final metabolic outputs. Although maximum SA titer was

found near 110:40 GLC:YE condition in batch mode, this also results in higher AC production and lower consumption rates and yields on the stationary phase which result on incomplete substrate exhaustion (Fig. 2c, d, h and g). In contrast, higher SA production, higher GLC consumption and lower AC production rates on stationary phase are found (Fig. 2h, g and j) near 80:40 GLC:YE conditions. Therefore, with these initial conditions, biomass n with

high rates are expected on the exponential phase without compromising stationary phase SA production and GLC consumption capabilities. This is supported also by the flux surface analysis, on the 80:40 GLC:YE initial conditions on stationary phase where relative fluxes are found to enhance SA acid production and GLC consumption, marked by the maximizing tendency for the reactions Pgi, GapA, PykA, Gnd,TktA, DAHPs and PckA (Fig. 8a, b, c, f,
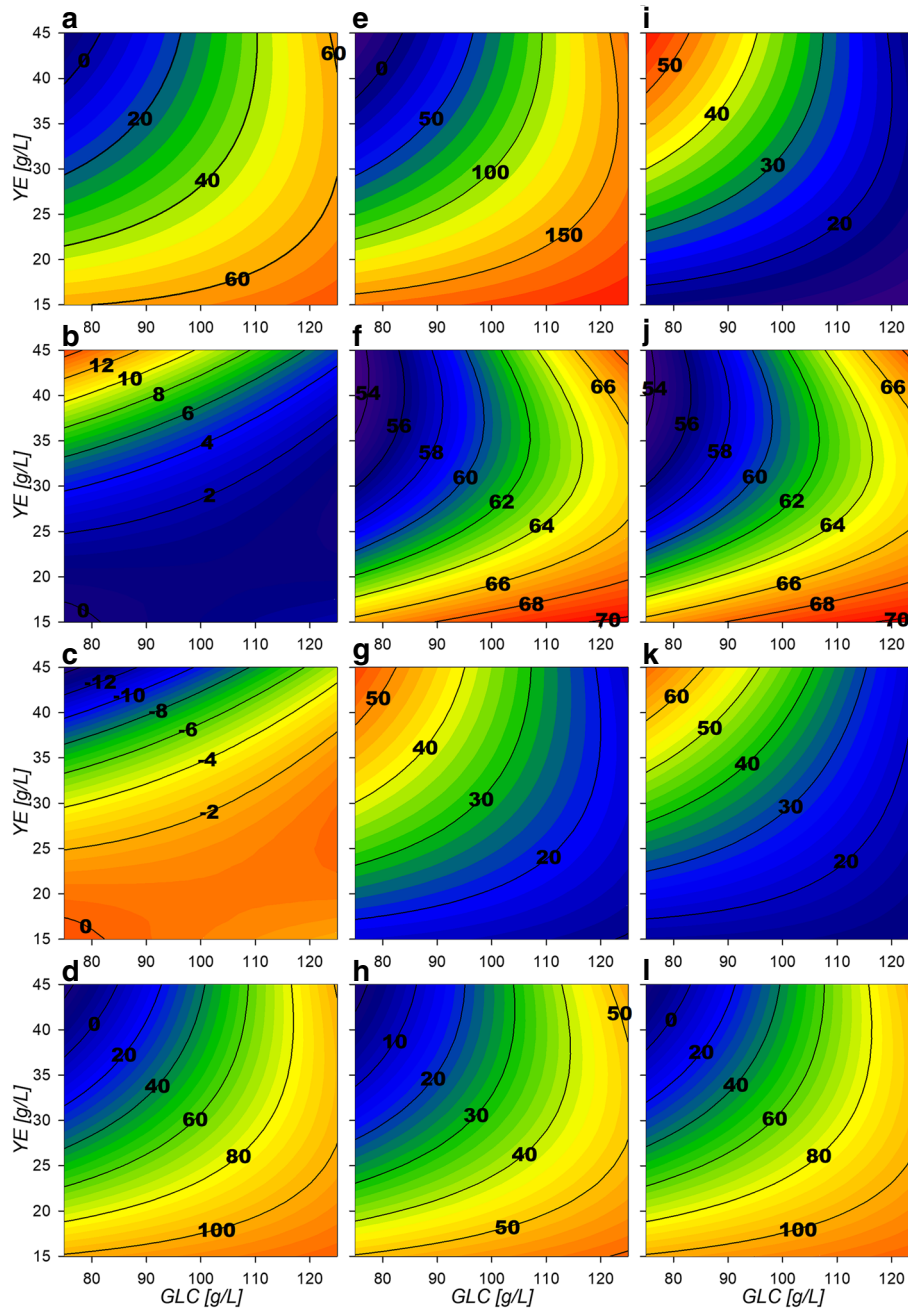

Fig. 9 Response surface contour plots for the estimated internal fluxes at MSt pt2 (% Flux relative to GLC consumption). **a** ActPout, **b** ActPin, **c** AckA, **d** Acs, **e** PoxB, **f** Csyn, **g** IcdA, **h** Icl, **i** KGdh, **j** SdhABCD, **k** Mdh, **l** MaeB

h, i, j and k). Also, in these selected conditions, lower values for relative fluxes were found for the reactions G6Pdh, PGdh and Ppc (Fig. 8e, g and l). This suggests that under the selected conditions less flux is sent towards the EDP and more is redirected to E4P through Gnd and TktA on the stationary phase. In fact the data indicate an $\approx$ 50% carbon flux redirection towards SA and the rest through lower glycolytic reactions by PykA and LpdA. Furthermore, consumption of AC fluxes will be maximized under this phase as seen on the relative flux surfaces for ActPin and AckA and lower AC production by PoxB (Fig. 9b, c and e). Therefore, higher fluxes are expected for GLC consumption and SA production, along with low AC production for the stationary phase. These conditions where then chosen even with the trade off with the exponential phase which presents higher GLC consumption flux rates for GalP (Fig. 10a) with high biomass production fluxes (Fig. 10b) and where SA production is not maximized. SA production on stationary phase presents higher values for DAHPs relative fluxes at lower initial [GLC]

concentrations (Fig. 6j). Nevertheless, flux rates for SA production reactions (DAHPs) presented medium range values within the experimental region (Fig. 10c).

As mentioned, fed-batch operation was designed to favor the biomass preferential production during growth phase and then use the SA production preferential zone during no-growth conditions. The hypothesis was that this would help to stabilize the flux distributions described on the modeled surfaces and therefore maintain yields with higher process productivity as more cells would be present. Upon ending the feed, a stationary or non-growth phase would in theory be expected to show similar physiological and flux distribution behavior as described by the stationary modeled response surfaces and in this way enhance SA production with controlling AC production at high yields and process substrate conversion. This was achieved by designing a pseudo-exponential feeding profile with concentrated solutions of GLC and YE that considered substrate addition from the beginning of the fermentation to maintain the initial
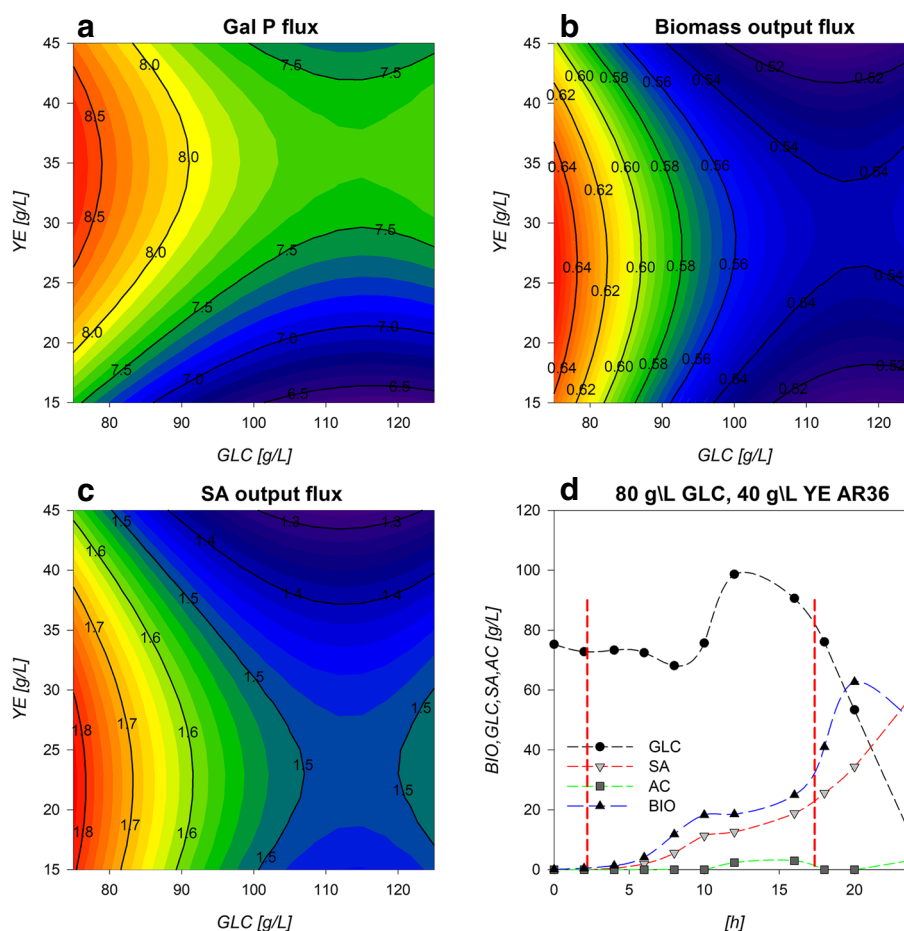

Fig. 10 **a** GalP flux surface [mM/h], **b** Biomass flux surface [mM/h], **c** DAHPs flux surface [mM/h], **d** Fed-batch reactor fermentation maintaining initial operation concentration parameters. 80 g/L GLC and 40 g/L YE

concentration conditions as long as possible. It is important to notice that all models consider YE as a unique metabolite so balancing flux for this complex substrate was just an approximation.

Figure 10d shows the fermentation profiles for the fed-batch optimized SA production process. Feed was performed from hour 3 (since calculated feed was previously too small for peristaltic operation) to hour 18 and controlled every 15 min manually, to match calculated growth and consumption parameters. GLC concentration was maintained near ≈75–80 g/L during the first 8–10 h of fermentation where an increase of GLC was observed up to 100 g/L concentration at hour 12. During this process, we also found a lower growth rate that could be responsible for the GLC accumulation, which was attributed to the manual control of the feed rate (that would lead to a significant overestimation of the feed over the time) and to the simultaneous feeding of similar amounts of YE and GLC (since individual component consumption cannot be calculated as it is taken as a simplified metabolite on models). Therefore, imbalances on feed may cause the exhaustion of crucial metabolic intermediates. Then the overestimated feed fluxes during the observed stall may cause an extracellular re-accumulation of these limiting components resulting in the second growth phase seen after 16 h and up to the 20th h of culture with a lower growth rate. Even though, it is still not clear the reason for this particular stall and further improvements on fed-batch operations could further give us insight on the strain behavior and enhance SA production. The first growth rate registered 0.90 h^{-1}, which is in range of the ones predicted by models (0.8–0.85 h^{-1}), while the second growth rate is only about 0.18 h^{-1}, suggesting limitation by an unknown substrate. Regarding the SA production rate, it also responded to this 8–16 h stall. Despite this, high GLC consumption rates on stationary phase were maintained and GLC was completely exhausted after only 24 h. A total of 180.5 g of GLC were consumed and 59.1 g/L of SA were produced. Although this titer is the highest obtained with this strain, it is 30% below the maximum titer reported on *E. coli* by Chandran et al. [5]. Moreover, the AC concentration was never found to be above 5 g/L, proving that process design was successful to limit the AC production even on this atypically high substrate conditions and compared to the batch culture where >15 g/L AC were accumulated. Furthermore, the process presented a global volumetric production rate of 2.45 g SA/L*h representing a 70% increase from the 1.43 gSA/L*h reported by Rodriguez et al. (2013)[2] and which is 20% higher than the previously reported industrial *E. coli* strains (2.04 gSA/L*h Chandran et al.[5]). Crucially, yields calculated by linear regression approximation were 0.40 for $Y_{p/s}$ and 0.744 for $Y_{p/x}$, which means that both yields were maintained along fermentation relative

to batch conditions and suggest that carbon distribution along metabolic nodes relevant for SA production were maintained within reason. These yields are also in accordance with previous works with strains lacking the PTS and are still among the highest reported on *E. coli* [2, 14, 33, 36, 38, 41, 42].

Discussion

Response surfaces showed the capacity to characterize correctly the physiological behavior of the AR36 strain. The observed increase on X_{max} mainly by [YE] and low increment by [GLC] is related to the fact that it is the only source of aromatic amino acids (Fig. 2a). This indicates that in all the experimental design area, YE can be taken as the limiting substrate for biomass production. This can also be observed on the fermentation profiles shown in Fig. 1, where the stationary phase of fermentations initiates always before limiting GLC concentrations. SA production is expected to follow the GLC consumption as it is the main source of carbon redirection to PEP and E4P, and this trend could be observed on Fig. 2b and c. However the maximum for SA production is found before maximizing consumption. This small difference can be explained by the Final AC ([AC]$_f$), which tends to increase at higher initial concentrations of GLC along with a maximization tendency at concomitant smaller initial concentrations of YE (Fig. 2d). High [AC] can hinder the H$^+$ balance across the membrane and considering that AR36 uses the *galP* coded galactose-proton symporter for GLC transport, consumption could be compromised [2]. Also, higher [AC] makes ATP production costlier, and in consequence, also the ATP-dependent phosphorylation of GLC by glucokinase towards the glycolytic metabolism [2]. Also, AC production is commonly related to metabolic overflow and considering that AR36 strain lacks the *pykF* gene, higher [GLC] and increasing intracellular [PEP] and [PYR] could be causing the observed higher AC production and lower GLC consumption [36, 38, 40, 42–44]. Therefore, the lower values for GLC consumption and SA production can be explained by the high AC concentrations produced by the strain within this experimental design region (high initial GLC, low initial YE). It is therefore possible to allocate a virtual SA vs AC critical line near 100 g/L GLC concentration, where below this line SA production may be favored and above this line AC production becomes relevant. It is important to notice that the very high GLC concentrations used in all experiments would typically result in higher AC production and slow growth in *E. coli* [45, 46]. Nevertheless, it is known that the AR36 strain can grow and maintain SA production at high GLC concentrations, with relatively low AC production, as a result of the high constitutive expression of SA biosynthetic genes and the lack of carbon repression present in strains lacking PTS [44].

Regarding specific rates a diminution of the GLC consumption rate under increasing [YE] was also found. This could be related to YE components competing for transport energy or amino acid allocation for their transporters. This may be possible since as mentioned, AR36 does not present catabolic repression (as a consequence of *crrHI* operon deletion) and is capable to transport simultaneously various carbon sources in the presence of GLC after synthesis induction of alternative transporters by cell carbon scavenging signals [36, 40, 42, 44]. More over, at higher [YE], PYR consumption reactions could be kinetically saturated since greater alanine (ALA) YE derived concentrations could enter the CCM to PYR saturating this metabolite pool and reducing GLC consumption. In contrast, at the stationary phase q_{glc}^{sta} presented higher values on higher initial concentration of both substrate sources. This behavior is expected for GLC, as higher concentrations of this substrate remain on the stationary phase and could be triggering higher consumption rates. However, it remains unclear why higher concentrations of initial YE could cause greater GLC stationary consumption rates. One possibility is that, higher initial [YE] could result in more oxaloacetate present on this phase increasing TCA activity and GLC consumption. This observation correlates with the behavior found for Δ[GLC] where only at high [YE], high [GLC] initial concentrations can be exhausted.

Regarding SA and AC production profiles it seems clear that as they are the main metabolic outputs for this strain they present the inverse maximization zones on experimental design area on both fermentation stages (Fig. 2f, h, i and j). This can be explained by the potential competition for carbon flux. These characteristics along with the biomass and the GLC related surfaces allowed to describe different output behavioral zones as described on the results section. However it is interesting that on the surfaces, it is possible to find similar consumption rates on opposing sides of the experimental design area with greatly different physiological outputs. Furthermore, the response surfaces morphologies have far from linear contours along the different physiological characteristics found on them. This could suggest the existence of metabolic state multiplicity similar to pseudo-stationary ones described by Namjoshi et al. on continuous bioreactors [47].The difference on response surface behaviors should derive from the dynamic properties, which produce different outcomes depending on the extracellular and intracellular metabolite concentrations and to the non-linearity associated with metabolic regulation [36, 42, 43, 47].

For the reasons described above and as the underlying characteristics of the systems were difficult to address only with external behavioral response surfaces, dynamic flux models were constructed. Cybernetic modeling was used mainly because interaction between cellular auto-regulated and inter-regulated subsystems (DNA, RNA, ENZYME) cannot be mechanistically described but some systematical characteristics can be approximately modeled [31, 48, 49]. Also, it is important to note that our media contains a non fully described compound substrate as YE, which was simplified as describe to only few metabolites resulting from its consumption by AR36. This simplification could not only impacts the growth phase, but also means that consumption of other YE components are not fully considered and may modify yields and rates. However, it is also noticeable that this simplification resulted enough to reasonably describe some behavioral characteristics on the fluxes found for AR36, which in consequence described the physiological outputs with reasonable accuracy. The latter observed responses are the results of a matrix made out from the network of the central carbon metabolism to which mathematical reduction and yield analysis from the previously determined parameters resulted on 6 elementary modes (EMs) modes for the exponential phase and 3 EMs for the stationary phase, and their combination across time renders the output described. This means that model is in essence the same for all 9 experiments and the approximation was performed on to the parameters that regulate their combination across time. Results suggest that with this EMs an estimated description of metabolic behavior can be made for the various initial conditions explored. The usage of the experimental design and surface rendering for the relative fluxes helps depict the metabolic behavior of the strain even with the errors described previously on individual points. Flux surfaces where then correlated to the physiological characterization as well as for behaviors known for this strain on the literature as discussed below.

An unusual flux distribution redirecting most of the GLC derived carbon through PPP was found. This may be possible on strain AR36 as it has the *zwf* gene over-expressed by a strong promoter on a high copy number plasmid [2]. Furthermore, FBA models by Chen et al. (2011)[32] have established that G6Pdh is rate limiting for PPP flux. Therefore, the high expression of *zwf* on AR36 strain could in fact be causing this low Pgi flux distribution. Despite this, a high total glycolytic flux is still found as described by Rodriguez et al. 2017 [44] as it was found that the operon-containing plasmid augmented the GLC consumption rate. The latter is in agreement to the high relative flux values of 75 to 88% towards PEP and PYR. On the modeled results this was possible since the high PPP flux was mainly redirected trough EDP to G3P and PYR. In this regard, it has been reported that on *pykF* mutants (such as AR36), fluxes through PPP are increased up to 79% by *pgi*, *pfkA* and *tpiA* down-regulation and *zwf*, *gnd* and *edd* concomitant

up-regulations [50, 51]. For AR36, a low intracellular level of fructose-1,6-bisphosphate (FDP) was previously found by comparative metabolomics and explained as a consequence of TktA activity [42, 44]. But with the results here obtained it can be suggested that it could also be influenced by a high flux deviation into the PPP by G6Pdh . Although *pgi* mutants have been reported to have lower growth rates caused by NADPH accumulation redox imbalance, it also has been found that overexpression of NADPH-consuming pathways can recover the growth rate [51]. Therefore, in AR36 the high production of SA, requiring NADPH by *aroE* coded shikimate dehydrogenase, could be alleviating the NADPH imbalance and promoting higher growth rates in the presence of high PPP flux distributions. In this sense, the production of SA in this strain could act as an important driver of its own synthesis when the *zwf* gene is overexpressed alongside the SA biosynthetic genes during growth phase. In contrast, on the stationary phase although a high PPP flux distribution was also found, low EDP flux was also found while high glycolytic PEP producing reactions were maintained. However this could also be possible since higher fluxes through Tal were found and that could cause higher FDP concentrations compared to the growth phase, signaling the up-regulation of the downstream glycolytic genes [43].

These results imply that E4P is the limiting substrate for SA production even with *zwf* overexpression and a high flux redirection towards the PPP, as previously suggested [42–44]. The modeled flux ratio analysis suggest that sole overexpression of *zwf* is not sufficient for alleviating E4P limitation. Therefore, *edd* and/or *eda* genes could be attractive deletion targets to avoid undesired partitioning of PPP fluxes, along with exerting better control under *zwf* and *gnd* overexpression to obtain higher but controlled flux distributions towards E4P and SA. Following on, PpsA presented a near-zero flux within all experiments. Considering the PEP overabundance to E4P, it can be deduced that even when *ppsA* has been previously used as a target to enhance SA production [14, 36, 38, 52, 53], the overexpression of this gene in AR36 may not further increase SA production. Furthermore, it is possible that overexpression of *ppsA* on this genetic background could hinder growth and GLC consumption by reducing carbon flux towards TCA and other PEP derived pathways.

Also, an unusually high PoxB flux was found in this strain as consequence of the increased influx towards PYR. This is supported by the previous findings in other related PTS-deficient strains lacking carbon catabolism repression, where PoxB has been proposed to be the main AC producing enzyme [33, 36, 38]. Furthermore, an *arcA/arcB* mutation has been found for this strain lineage that could be making PoxB available for expression

on earlier fermentation phases [33, 34, 36, 38, 40]. Other studies have also proposed PoxB as the main AC synthesizing enzyme under higher growth rates on accelerostats on other *E. coli* strains [54].

Regarding the AckA vs Acs AC consumption flux distribution, a high up-regulation of *acs* and *poxB* genes has been observed to occur as a response to PTS inactivation on this strain, suggesting that carbon cycling on AC occurs through Acs [36, 38, 40]. Therefore, it is probable that the model depicted the incorrect or inverse distribution around these reactions during growth. Moreover, the AR36 lineage strains does not show the expected PTS mutant low cAMP concentrations, probably due to AC cycling through Acs restoring cAMP along with adelynate cylase, which in turn has been also found to be up-regulated on these PTS⁻ strains during growth [36–38, 40]. It is interesting to find that the EMs used for stationary phase rendered Acs as the principal reaction responsible for redirecting AC to ACCOA (Fig. 9c and d), probably due to less ATP demand on this phase. Furthermore, the combination of the export and import AC surfaces strongly correlates with the q_b^{exp} approximated AC production rate on the physiological analysis presented before (Fig. 2i), where high extracellular AC production is found towards high initial [GLC] and towards high initial [YE]. Since both fluxes are present, it can be proposed that an AC production/consumption futile cycle could potentially be relieving part of the metabolic overflow on the CCM [33, 36, 37], as has previously been reported to help adjusting imbalances between glycolysis and the TCA activities [43].

TCA activity on its part, showed the unexpected IcdA reverse reaction. This could be attributed to the YE derived GLU entering the TCA through α-ketoglutarate (AKG) by Gdh, but since this compound is used also for biomass precursor formation, only if it is consumed on excess it will enter the CCM. This effect could also be increased by other YE derived amino acids entering CCM. On that regard, reports on complex media have shown a tendency to favor extracellular amino acid consumption and catabolism through AKG with the concomitant up-regulation of biomass producing pathways [55]. Also, the high GSP and malic enzyme activities suggest that this strain counteracts metabolic saturation by trying to relieve the PYR saturation by assimilating more ACCOA through lower CCM pathways. This may be the result of a selective pressure to the high osmotic pressure on this media to recover high substrate consumption rates and consume the highly concentrated substrates faster. In fact, high osmotic stress conditions have been found to increment the GSP activity and to reduce the *icdA/aceA* coded enzymes ratio, favoring the production of biomass building blocks [56]. Therefore, the high anaplerotic reactions (APR) fluxes along with GSP found also may contribute

to maintain high biomass and SA production since this could help to minimize toxic AC production by ACCOA fast consumption ,along with a response to osmotic pressure [43, 56]. This behavior may be expected as with the higher [GLC] more stress upregulating GSP could be used to relieve carbon flux from the PYR and ACCOA nodes [33, 36–38, 40, 43]. Furthermore, the excess carbon arriving to MAL then seems to be redirected to PYR and PEP by the previously detailed MaeB, MaeA and PckA reactions. MaeB presents greater fluxes possibly due to consuming the excess NADPH produced by the high PPP flux conditions [51].

Overall, the dynamic cybernetic model approach seems to unveil behaviors that are in accordance to the physiological observations and to the knowledge available for this laboratory evolved strain lacking the mayor GLC transport and therefore catabolite repression. The behavior of the calculated fluxes surfaces during the growth phase is in agreement with the results obtained with the physiological surface analysis. Particularly, on many of the IE and ME surfaces, critical surface saddle points between 110–115 g/L of initial [GLC] and between 35–40 g/L of initial [YE] have been found. It is interesting to notice that 9 EMs were enough to describe all the patterns conformed by the surfaces by only modifying the parameters which alter their combination across time. These changes, although cannot be used to describe regulatory mechanisms, unveil relevant systems characteristics and interestingly also suggest the existence of metabolic state multiplicity derived from changing extracellular conditions [47].

Conclusion

In this report we describe a modeling approach for a PTS⁻ laboratory evolved *E. coli* engineered strain for SA overproduction [2, 33, 36, 38, 40] to study and characterize its physiological and metabolic responses to variant complex substrate concentration. The constructed models were able to describe in good agreement the individual experimental fermentations performed with this strain. Three-dimensional response surfaces were constructed with polynomial equations allowing to morphologically describe the cell output behavior under the experimental conditions. It was found that the production strain responds differently to initial substrate concentrations, allocating resources in different ways. This was inferred since regulation along variations from complex media substrate conditions did not affect linearly the performance of the strain, but showed refined nonlinear borders between predominant outcomes. For these reasons a dynamic cybernetic model was constructed and their flux distributions studied and compared to the physiological models. The constructed dynamic metabolic model was able to follow the extracellular experimental behaviors

and three-dimensional response surfaces for relative flux distributions were used to unveil insights into the strain metabolism.

Flux distributions helped to explain the previously observed low intracellular level of fructose-1, 6-bisphosphate (FDP) reports by unveiling a high PPP flux during all fermentation processes. MaeB high relative flux, potentially helps alleviate the NADPH redox imbalance caused by NADPH-dependent SA production and contributes to growth rate recovery [2]. Flux distributions allocated AC production, GSP and APR with high fluxes to contend with the metabolic stress produced by the concentrated substrates on media helping to relieve PYR and ACCOA overflow. PoxB was found to be the predominant AC production enzymatic reaction on this strain under high substrate conditions. An AR36Δ*poxB* strain was constructed and cultured showing the loss of growth phase AC peak and along with lower growth, GLC consumption rates probably by greater PYR accumulation on this derivative [33, 36–38, 44]. Model analysis also found that *edd* and/or *eda* could be targeted for deletion, along with a better control under *zwf* overexpression and *gnd* expression, to obtain better flux distributions towards E4P and SA production. Also, *ppsA* overexpression and other modifications involving higher PEP accumulation may not improve SA production until E4P limitation is resolved on this strain. Dynamic models were found to be in accordance to the physiological observations and the knowledge available for AR36 a PTS⁻ strain lacking catabolite repression, and were useful to allocate preferential metabolism output zones within the experimental design area that correlated in good agreement with the zones observed during the physiological model characterization.

Finally, to assess the utility for SA production enhancement with all the previously described models, a fed-batch fermentation regime was designed. An unusual operation was employed to maintain initial media conditions which would in theory help maintain metabolic an physiological conditions. The fed-batch fermentation resulted in a 40% titer and 70% volumetric productivity increases while preserving product and biomass yields. Process presented yield values among the highest yields reported and presented the highest productivity reported on *E. coli* AR36. Although reports concerning other strains have shown higher titers [57], this report was centered on the mathematical approach to further extend *E. coli* production capabilities. On that matter, the model implemented in this report is the first approximation to render flux distributions for this *E. coli* PTS⁻ strain under high-substrate production conditions and one of the first approaches towards modeling *E. coli* metabolism in complex media containing high concentrations of GLC and YE.

Methods

Strain, cultivation and analytical procedures

E. coli AR36 strain constructed by Rodríguez et al. (2013) [2] was used for all experiments and calculations. AR36 is an *E. coli* PB12 laboratory evolved derivative lacking the phosphoenolpryruvate:carbohydrate phosphotransferase system (PTS) [33, 36, 38, 40]. AR36 carries additional inactivations in *aroK*, *aroL*, *pykF* and *lacI* genes, and contains a high copy number plasmid with the strong *trc* promoter controlling transcription of a six-gene operon composed of genes: *aroB*, *tktA*, *aroG*, *aroE*, *aroD* and *zwf* [2]. This strain is an aromatic amino acid auxotroph and therefore it must be cultured on supplemented media. Yeast extract (YE) (BD Bacto) and GLC (Fermentas) were used as nitrogen and carbon sources [2]. All cultures were performed on 0.5 *L* working volume bioreactors with AD 1010 controllers (Applikon). Bioreactors were operated as batch processes at 37°*C* and 1 *vvm* aeration. Dissolved oxygen tension (DOT) was maintained above 20% by an agitation cascade control between 500 and 1200 *rpm*. pH was maintained at 7 by means of NH_4OH and H_3PO_4 addition. Other media compounds, salts, buffer and antibiotics were used as previously described [2].

Physiological behavior characterization was performed with a central composite design experimental matrix with 3 levels for each substrate source. Experimental condition levels were: 75, 100 and 125 *g/L* for GLC and 15, 30 and 45 *g/L* for YE. Nine experiments were conducted with the central point 100 *g/L* GLC and 30 *g/L* YE, tested by triplicate to approximate the experimental design standard deviation. Fermentations were sampled every 2 h during the first 12 h, and every 4 to 6 h after this point. Each sample was used to determine biomass, GLC, SA and AC. Biomass was determined by optical density measurements at 600 *nm* with a DU700 Beckman spectrophotometer. GLC, SA and AC were determined by HPLC with a Waters equipment(600E quaternary pump, 717 automatic injector, 2410 refraction index an 966 photodiode array detectors) and an aminex HPX-87H column (300 x 7.8 *mm*; 9 *μm*), using 5 *mM* H_2SO_4 as mobile phase at 50°*C*; either UV or refractive index detectors were used for qualitative and quantitative determination. All measured parameters were volumetrically corrected for the acid or base added by pH control pumps.

Calculation of fermentation parameters

For fermentation data parametrization and analysis, a set of modeling approaches was constructed. The maximum growth rate μ_{max} and maximum biomass X_{max} were obtained by adjusting a logistic growth model to experimental data. Since the fermentation processes use complex media, calculation of yields and production/

consumption rates by classical calculations were difficult to address. Therefore, to provide a more accurate parametrization, GLC consumption and SA production integrated models were constructed. The integrated model equations used were:

$$X_{(t)} = \frac{X_0 e^{\mu_{max}t}}{1 - \left(\frac{X_0}{X_{max}} \left(1 - e^{\mu_{max}t} \right) \right)} \tag{1}$$

$$S_{(t)} = S_{(t-1)} - \left[\left(q_{glc}^{exp} X_{(t)}(\Delta t) \right) \left(1 - \frac{X_{(t)}}{X_{max}} \right) \right]$$
$$- \left[\left(q_{glc}^{sta} X_{(t)}(\Delta t) \right) \left(\frac{X_{(t)}}{X_{max}} \right) \right] \tag{2}$$

$$P_{(t)} = P_{(t-1)} + \left[\left(q_{sa}^{exp} X_{(t)}(\Delta t) \right) \left(1 - \frac{X_{(t)}}{X_{max}} \right) \right]$$
$$+ \left[\left(q_{sa}^{sta} X_{(t)}(\Delta t) \right) \left(\frac{X_{(t)}}{X_{max}} \right) \right] \tag{3}$$

where S refers to substrate, in this case GLC, and P refers to product, SA on this experimental design. $X_{(t)}$ is the biomass calculated at time t by the logistic growth model and X_{max} is the maximum biomass parameter. q_{glc}^{exp}, and q_{sa}^{exp} are the specific exponential rates for GLC consumption and SA production, respectively. q_{glc}^{sta}, and q_{sa}^{sta} are the specific stationary rates for GLC consumption and SA production, respectively. The participation of each exponential or stationary rates across time is regulated by the terminus describing the biomass and maximum biomass ratio correlated to the biomass logistic model. The production and consumption rate parameters were approximated by the sum of the square error (SSE) minimization against experimental data using MATLAB programming. Product/substrate and product/biomass yields were estimated from the obtained specific rates.

Models constructed were tested for their experimental data approximation by an error estimation calculated by the relation between the sum of the square error (SSE) and the sum of the square of the experimental points (SSEP). Model approximation was then mathematically described by the linear regression between experimental points and model points. A percentile deviation from the expected slope (1 for experimental and model equality) descriptive indicator constructed from regression (SDP) as well as Pearson regression coefficient (describing dispersion) and the regression significance proved by *p*-value statistics were used to qualify the acceptance of models as descriptors for the experimentally observed behavior.

AC presented a dynamic behavior (simultaneous production and consumption) in all fermentations that could not be described by any of the previously described equations. Nevertheless, initial exponential $\left(q_b^{exp} \right)$ and initial stationary $\left(q_b^{sta} \right)$ approximated production rates were calculated with the following equations:

$$q_b^{exp} = \mu_{max} \times Y_{ac/x}^{exp} \tag{4}$$

$$q_b^{sta} = Q_b^{sta}/X_{max} \tag{5}$$

where yield was calculated by linear regression for AC vs biomass and volumetric rate was calculated by linear regression for AC vs time on the first experimental data points for each phase.

Parameters describing the physiological behavior were then used to construct individual three-dimensional surfaces. A second-order bivariate polynomial equation was used for surface construction and its approximation was addressed and qualified by regression coefficients, *p*-values and square sums of error and percentile error. Surfaces were validated by prediction of parameters for three fermentations not contained on the set of the experimental design (75:20, 80:40 and 115:45 GLC:YE initial conditions). Surface calculated parameters were introduced to the logistic biomass, consumption and production models and compared to the experimental data sets. Surface calculated parameters were also compared to the ones calculated directly from experimental data by error percentage, estimated by the ratio between standard deviation between each calculation and experimental parameter. A two-tailed *t*-student test using the experimental design standard deviation, calculated from central point, was used to determine if the experimental parameters and surface predicted parameters were significantly different.

Dynamic metabolic flux model construction

Metabolic flux distribution was constructed with the use of a dynamic cybernetic model approach developed by Ramkrishna et al. [31, 48, 49], which has proved to be useful to address dynamic changes on fluxes when information on mechanistic details of regulatory processes is scarce or suboptimal [58]. The cybernetic modeling introduces regulation by the use of two vectors $u \equiv [u_1, u_2, u_3, ..., u_m]$ and $v \equiv [v_1, v_2, v_3, ..., v_m]$ referring to them as cybernetic variables, associated with fractional allocations of resources for enzyme synthesis and activity, respectively [31, 48, 49]. These variables are calculated along the fermentation and modify the participation of each elementary mode obtained from the stoichiometric matrix analysis. These elementary modes (EMs) are sets of non-decomposable pathways consisting of minimal sets of reactions that describe all the cellular metabolic routes. A subset of EMs must then be extracted to describe metabolic behavior on a parametrically achievable scale. Therefore, elementary mode analysis (EMA) must be performed to find the minimal set of EMs that can describe the behavior expected from the specific constraints imposed by either the strain or experimental conditions. Cybernetic models then calculate flux rates for each EMs described as sets of Michaelis-Menten type equations where a relative enzyme concentration and

biomass conform the maximum rate, modified at each time by the cybernetic variable *v*. The relative enzyme concentration is calculated by another Michaelis-Menten type equation that considers a maximum enzymatic production rate and a disappearance rate, regulated by the cybernetic variable *u*. The cybernetic variables are regulated by an objective function, evaluating the outputs at any given time *t* between all EMs, increasing priority on the next time step $t + \Delta t$ to the better performing EM by a matching law strategy. In this way, the cybernetic models can take into account dynamic regulation with respect to a specific cell metabolic objective, such as growth rate maximization or carbon uptake maximization, even with little information on the mechanistic particularities to its function, allowing for dynamic flux distribution modeling [31, 48, 49].

In this work, a CCM network was constructed from 60 reactions, 44 internal metabolites and 6 external metabolites, accounting for the Embden-Meyerhoff-Parnas pathway (EMP), the Pentose Phosphate pathway (PPP), Tricarboxylic Acid Cycle (TCA), Pyruvate Metabolism, Anaplerotic Reactions, respiration and energetic reactions, YE components uptake reactions, SA biosynthesis reactions and biomass generation reactions. External metabolites defined were AC, GLC, SA, YE, biomass, and maintenance. YE consumption was introduced to the network reaction as a metabolite and its consumption derived into biomass precursor (BIOMp), aromatic amino acids (taken as a unique metabolite), alanine (ALA), and glutamic acid (GLU). The stoichiometric values for YE conversion to these metabolites were estimated from the average composition described by the manufacturer, where BIOMp was taken as the rest of amino acids that account to produce proteins contained on biomass. EMs computation was made with efmtool protocol [59] embedded on MATLAB [60]. For EMA two EMs families were constructed, the first family contained the exponentially preferred EMs by only selecting the ones that contained simultaneous GLC and YE consumption and constrained to produce SA. The second family of EMs was selected from the ones containing GLC consumption and simultaneous production of SA and constrained to not consume YE, which will be preferred on the stationary phase of fermentations. Yield analysis reduction by convex hull volume was performed as described as by Song et. al. [60] to find the minimal subset of EMs. Experimental design central point yields were used for this analysis. Yield analysis around the Phosphoglucose isomerase (Pgi)/ Glucose 6-phosphate-1-dehydrogenase (G6Pdh) node was studied with values stated as: non-constrained, 0.25, 0.5, 0.75 and 0.90 Pgi/GalP yield or flux normalized to GLC uptake. This generated 5 flux distributions sets. All model sets were evaluated by Pearson linear regression coefficients between experimental points

and model points, slope deviation and their significance was proved by p-value statistics and error. The set with better behavior to experimental data was used for further analysis. Reduced EMs reaction rates were described by sets of Michaelis-Menten equations modified to couple families to each fermentation phase (exponential or stationary) by a terminus similar to the physiological consumption/production models described before, and with an added terminus to represent AC inhibition. Model rate equations were constructed as follow:

$$r_i^M = \left(\frac{k_{max,i}^M [\,GLC\,]_{(t)}}{K_{m,i}^M + [\,GLC\,]_{(t)}} \right) \left(1 + \frac{[\,AC\,]_{(t)}}{K_I^{ac}} \right)^{-1} \left(1 - \frac{X_{(t)}}{X_{max}} \right) \tag{6}$$

$$r_i^G = \left(\frac{k_{max,i}^G [\,GLC\,]_{(t)}}{K_{m,i}^G + [\,GLC\,]_{(t)}} \right) \left(1 + \frac{[\,AC\,]_{(t)}}{K_I^{ac}} \right)^{-1} \left(\frac{X_{(t)}}{X_{max}} \right) \tag{7}$$

where indexes M and G refer to the exponential (mixed consumption) family and stationary (glucose only consumption) family, respectively. Index i refers to each ith EM of each family, K_{max} and K_m are the Michaelis-Menten parameters, K_I^{ac} is the inhibition coefficient and X_{max} refers to the maximum biomass calculated from the previously calculated logistic growth model for each experimental point. $[\,GLC\,]_{(t)}$, $[\,AC\,]_{(t)}$, $X_{(t)}$ refer to the GLC, AC and biomass concentrations at each time t in $mmol/L$ for the first two and g/L for the biomass. Initial relative enzyme concentration ratio was set to 0.95 for the first EM, 0.5 for the EMs remaining of the exponential family, and 0.1 for the EMs of the stationary family. These values were set in this way as the first EM was the one that comprised AC production and was inferred to be the initially preferred one, due to the observed rapid increase in extracellular AC acid at the initial phases for all fermentations. In addition, EMs of the second family were chosen to be smaller as they are expected to be more relevant at later stages of fermentation. All other cybernetic model parameters for enzyme production and decay rates were set as described by Ramkrishna et al. [31, 48, 58, 60]. Flux rate equation parameters K_{max} and K_m were approximated with a genetic algorithm (Additional file 2). Briefly, the Matlab algorithm started with assigning K_{max} and Km initial values of 1 and 10, respectively for every EM. Then, by perturbation of one parameter at a time by a random numeric factor, 18 parameter sets were obtained. Subsequently, the sets were used for 200 step SSE driven nonlinear numeric minimization algorithms to generate new daughter K_{max} and K_m parameter model sets. From these daughter models, the set with the lowest SSE was extracted and crossed with the second lowest SSE set by acquiring the value of its perturbed parameter (K_{max} or K_m). This inter-crossed set passed onto the next generation where another round of individual parameter perturbations was made. The algorithm was cycled until SSE was found constant (less than 20 cycles in all cases). Finally, these parameters were subjected to a final SSE nonlinear numeric minimization to model the flux rates of each EM and the final metabolic dynamic flux model for each fermentation.

Flux distributions were used to construct three-dimensional behavioral surfaces with the second order two variable polynomial equation at three fermentation times: initial exponential (IEx), mid exponential (MEx) and mid stationary (MSt). Calculation of MEx time was made by obtaining the maximum point of the second derivative vs. time for each biomass model, IEx was set as the mid time between $t = 0$ and MEx time, and MSt as the middle point between the end of the fermentation and the initial time of stationary phase. Constructed surfaces were statistically qualified and used to analyze and study the behavior of strain metabolism and SA production.

Bioprocess design for SA productivity enhancement on strain AR36

To assess the utility of the metabolic models developed, a fermentation process was designed with the information acquired by physiological models and flux distribution surfaces to optimize SA productivity. Process was designed to maintain constant the initial GLC:YE conditions. This would in theory, maintain constant the cybernetic variables (which are a representation of the regulation parameters) and therefore the internal flux distributions accordingly. Initial conditions around 80:40 g/L GLC:YE were used. A pseudo-exponential flux was operated and regulated by a peristaltic pump manually set every 15 min to calculated exponential feeding needs, with measured GLC and calculated growth and consumption rates. Feed consisted on two simultaneously added solutions, one containing mineral media with GLC 400 g/L and the second containing phosphate buffer solution with YE at 400 g/L. After 12 h, the feed was stopped and the fermentation was allowed to enter stationary phase to consume the remaining GLC. Oxygen was added when needed to maintain dissolved oxygen tension (DOT) over 20% along with an agitation cascade. Other media compounds, salts, buffer and antibiotics were used as described by Rodriguez et. al. [2] and added through the feeding solutions to avoid dilution.

Additional files

Martínez *et al. BMC Systems Biology*   (2018) 12:102

Page 24 of 26

Additional file 2: Metabolic network definition for dynamic model. Description: Definition of reactions, internal and external metabolites comprehending the metabolic network. Matlab program section for the genetic algorithm used for parameter approximation. (PDF 173 kb)

Additional file 3: Response surfaces and contour plots for all fluxes. Description: Response surfaces and contour plots for all reactions detailed on the metabolic network for initial exponential, mid exponential and mid stationary fermentation stages. (PDF 25,629 kb)

Additional file 4: Strains, plasmids and oligonucleotides used on this work, AR36 poxB⁻ figures. Description: Table for all plasmids and oligonucleotides used. AR36ΔpoxB strain construction. Fermentation profiles for AR36ΔpoxB and AR36 on highly concentrated substrate media. (PDF 215 kb)

Availability of data and materials
The data generated or analyzed during this study are included in this published article [and its supplementary information files] or available from the corresponding author on reasonable request.

Authors' contributions
JAM, NF, OTR, ARL, GG and FB contributed to the conceptual creation, methods design, interpretation and critical evaluation and discussion of the results presented in this report. Fermentation experiments planning, and design were carried out by JAM, ARL and OTR. All fermentation experiments were executed by JAM, except for the fed-batch fermentation, which was performed by JAM and AR. Molecular biology and strain design was performed by AR, GG, NF and FB. Data analysis and parameter calculations were performed by JAM along with ARL and OTR. JAM designed the models and implemented the genetic algorithm. JAM and FM constructed and evaluated AR36 ΔpoxB with the support of NF. JAM wrote the manuscript with the support of AR, ARL, NF, OTR, GG and FB, who critically revised the results of this report. All authors have read and approved the final version of manuscript.

Ethics approval and consent to participate
Not Applicable.

Consent for publication
Not Applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Departamento de Ingeniería Celular y Biocatálisis, Instituto de Biotecnología, Universidad Nacional Autónoma de México (UNAM), Avenida Universidad 2001, Colonia Chamilpa, 62210 Cuernavaca, Morelos, México. [2]Departamento de Ciencias Naturales, Universidad Autonoma Metropolitana (UAM),Vasco de Quiroga 4871, Colonia Santa Fe Cuajimalpa, 05348 Delegación Cuajimalpa de Morelos, México D.F., Mexico. [3]Departamento de Medicina Molecular y Bioprocesos, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Avenida Universidad 2001, Colonia Chamilpa, 62210 Cuernavaca, Morelos, México.

References
1. Patnaik R, Spitzer RG, Liao JC. Pathway engineering for production of aromatic in *Escherichia coli*: Confirmation of stoichiometric analysis by independent modulation of arog, tkta, and pps activities. Biotech Bioeng. 1995;46:361–70. https://doi.org/doi:10.1002/bit.260460409.
2. Rodriguez A, Martínez JA, Báez-Viveros JL, Flores N, Hernández-Chávez G, Ramírez OT, Gosset G, Bolivar F. Constitutive expression of selected genes from the pentose phosphate and aromatic pathways increases the shikimic acid yield in high-glucose batch cultures of an *Escherichia coli* strain lacking pts and *pykF*. Microb Cell Factories. 2013;12:1–16. https://doi.org/doi:0.1186/1475-2859-12-86.
3. Bongaerts J, Krämer M, Müller U, Raeven L, Wubbolts M. Metabolic engineering for microbial production of aromatic amino acids and derived compounds. Metab Eng. 2001;3:289–300. https://doi.org/doi:10.1006/mben.2001.0196.
4. Yi J, Li K, Draths KM, Frost JW. Modulation of phosphoenolpyruvate synthase expression increases shikimate pathway product yields in *E. coli*. Biotechnol Prog. 2002;18:1141–8. https://doi.org/doi:10.1021/bp020101w.
5. Chandran SS, Yi J, Draths KM, von Daeniken R, Weber W, Frost JW. Phosphoenolpyruvate availability and the biosynthesis of shikimic acid. Biotechnol Prog. 2003;19:808–14. https://doi.org/doi:10.1021/bp025769p.
6. Báez-Viveros JL, Osuna J, Hernández-Chávez G, Soberon X, Bolivar F, Gosset G. Metabolic engineering and protein directed evolution increase the yield of l-phenylalanine synthesized from glucose in *Escherichia coli*. Biotech Bioeng. 2004;87:516–24. https://doi.org/doi:10.1002/bit.20159.
7. Gosset G. Production of aromatic compounds in bacteria. Curr Opin Biotechnol. 2009;20:651–8. https://doi.org/doi:0.1016/j.copbio.2009.09.012.
8. Estévez AM, Estévez RJ. A short overview on the medicinal chemistry of (-)shikimic acid. Mini-Rev Med Chem. 2012;12:1443–54. https://doi.org/doi:10.2174/138955712803832735.
9. Quiroz DD, Carmona S, Bolívar F, Escalante A. Current perspectives on applications of shikimic and aminoshikimic acids in pharmaceutical chemistry. Res Rep Med Chem. 2014;4:35–46. https://doi.org/doi:10.2147/RRMC.S46560.
10. Li S, Yuan W, Wang P, Zhang Z, Zhang W, Ownby S. Method for Rhe Extraction and Purification of Shikimic Acid; 2007, p. 20070149805A1.
11. Raghavendra TR, Vaidyanathan P, Swathi HK, Ravikanth G, Ganeshiah KN, Srikrishna A, Shaanker RU. Prospecting for alternate sources of shikimic acid, a precursor of tamiflu, a bird-flu drug. Curr Sci. 2009;96: 771–2. Scientific Correspondence.
12. Wang G-W, Hu W-T, Huang B-K, Qin L-P. *Illicium verum*: A review on its botany, traditional use, chemistry and pharmacology. J Ethnopharmacol. 2011;136:10–20. https://doi.org/doi:10.1016/j.jep.2011.04.051.
13. Ghosh S, Chisti Y, Banerjee UC. Production of shikimic acid. Biotechnol Adv. 2012;30:1425–31. https://doi.org/doi:10.1016/j.biotechadv.2012.03.001.
14. Martínez JA, Bolívar F, Escalante A. Shikimic acid production in *Escherichia coli*: from classical metabolic engineering strategies to omics applied to improve its production. Front Bioeng Biotechnol. 2015;3:1–16. https://doi.org/doi:10.3389/fbioe.2015.00145.
15. Krämer M, Bongaerts J, Bovenberg R, Kremer S, Müler U, Org S, Wubbolts M, Raeven L. Metabolic engineering for microbial production of shikimic acid. Metab Eng. 2003;5:277–83. https://doi.org/doi:10.1016/j.ymben.2003.09.001.
16. Biggs BW, Paepe BD, Santos CNS, Mey MD, Ajikumar PK. Multivariate modular metabolic engineering for pathway and strain optimization. Curr Opin Biotechnol. 2014;29:156–62. https://doi.org/doi:10.1016/j.copbio.2014.05.005.
17. Cloots L, Marchal K. Network-based funcional modeling of genomics, transcriptomics and metabolism in bacteria. Curr Opin Microbiol. 2011;14: 599–607. https://doi.org/doi:10.1016/j.mib.2011.09.003.
18. Fong SS. Computational approaches to metabolic engineering utilizing systems biology and synthetic biology. Comput Struct Biotechnol J. 2014;11:28–34. https://doi.org/doi:10.1016/j.csbj.2014.08.005.
19. Akesson M, Forster J, Nielsen J. Integration of gene expression data into genome-scale metabolic models. Metab Eng. 2004;6:284–93. https://doi.org/doi:10.1016/j.ymben.2003.12.002.

Martínez *et al. BMC Systems Biology*　　(2018) 12:102

Page 25 of 26

20. Stelling J. Mathematical models in microbial systems biology. Curr Opin Microbiol. 2004;7:513–8. https://doi.org/doi:10.1016/j.mib.2004.08.004.

21. Kim TY, Sohn SB, Kim YB, Kim WJ, Lee SY. Recent advances in reconstruction and applications of genome-scale metabolic models. Curr Opin Biotechnol. 2012;23:617–23. https://doi.org/doi:10.1016/j.copbio.2011.10.007.

22. Covert MW, Schilling CH, Famili I, Edwards JS, Goryanin II, Selkov E, Palsson BO. Metabolic modeling of microbial strains *in silico*. Trends Biochem Sci. 2001;26:179–86. https://doi.org/doi:10.1016/S0968-0004(00)01754-0.

23. Price ND, Papin JA, Schilling CH, Palsson BO. Genome-scale in microbial *in silico* models: the constraints-based approach. Trends Biotechnol. 2003;21:162–9. https://doi.org/doi:10.1016/S0167-7799(03)00030-1.

24. Jouhten P. Metabolic modelling in the development of cell factories by synthetic biology. Comput Struct Biotechnol J. 2012;3:201210009. https://doi.org/doi:10.5936/csbj.201210009.

25. Patil KR, Akesson M, Nielsen J. Use of genome-scale microbial models for metabolic engineering. Curr Opin Biotechnol. 2004;15:64–9. https://doi.org/doi:10.1016/j.copbio.2003.11.003.

26. Schuetz R, Kuepfer L, Sauer U. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. Mol Syst Biol. 2007;3:1–15. https://doi.org/doi:10.1038/msb4100162.

27. Machado D, Herrgard M. Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. PLoS Comput Biol. 2014;10:1003580. https://doi.org/doi:10.1371/journal.pcbi.1003580.

28. Saha R, Chowdhury A, Maranas CD. Recent advances in the reconstruction of metabolic models and integration of omics data. Curr Opin Biotechnol. 2014;29:39–45. https://doi.org/doi:10.1016/j.copbio.2014.02.011.

29. Long MR, Ong WK, Reed JL. Computational methods in metabolic engineering for strain design. Curr Opin Biotechnol. 2015;34:135–41. https://doi.org/doi:0.1016/j.copbio.2014.12.019.

30. OBrien EJ, Palsson BO. Computing the functional proteome: recent progress and future prospects for genome-scale models. Curr Opin Biotechnol. 2015;34:125–34. https://doi.org/doi:10.1016/j.copbio.2014.12.017.

31. Ramkrishna D, Song H-S. Dynamic models of metabolism: Review of the cybernetic approach. Bioeng, Food, Nat Prod. 2012;58:986–97. https://doi.org/doi:10.1002/aic.13734.

32. Chen PT, Chiang C-J, Wang J-Y, Lee M-Z, Chao Y-P. Genomic engineering of *Escherichia coli* for production of intermediate metabolites in the aromatic pathwaty. J Taiwan Inst Chem Eng. 2011;42:34–40. https://doi.org/doi:10.1016/j.jtice.2010.03.010.

33. Flores S, Gosset G, Flores N, de Graaf AA, Bolívar F. Analysis of carbon metabolism in *Escherichia coli* strains with an inactive phosphotransferase system by [13]c labeling and nmr spectroscopy. Metab Eng. 2002;4:124–37. https://doi.org/doi:10.1006/mben.2001.0209.

34. Oh AJ, Lee HW, Saha R, Park MS, Joon Kj, Lee D-Y. Exploring the effects of carbon sources on the metabolic capacity for shikimic acid production in *Escherichia coli* using in silico metabolic predictions. J Microbiol Biotechnol. 2008;18:1773–84. https://doi.org/doi:10.4014/jmb.0700.705.

35. Rizk ML, Liao JC. Ensemble modeling for aromatic production in *Escherichia coli*. PLoS ONE. 2009;4:6903. https://doi.org/doi:10.1371/journal.pone.0006903.

36. Flores N, Flores S, Escalante A, de Anda R, Leal L, Malpica R, Georgellis D, Gosset G, Bolívar F. Adaptation for fast growth on glucose by differential expression of central carbon metabolism and gal regulon genes in an *Escherichia coli* strain lacking the phosphoenolpyruvate: carbohydrate phosphotransferase system. Metab Eng. 2005;7:70–87. https://doi.org/doi:10.1016/j.ymben.2004.10.002.

37. Sigala JC, Flores S, Flores N, Aguilar C, de Anda R, gosset G, Bolívar F. Acetate metabolism in *Escherichia coli* strains lacking phosphoenlpyruvate: carbohydrate phosphotransferase system; evidence of carbon recycling strategies and futile cycles. J Mol Microbiol Biotechnol. 2008;16:224–35. https://doi.org/doi:10.1159/000151219.

38. Flores N, de Anda R, Flores S, Escalante A, Hernández G, Martínez A, Ramírez OT, Gosset G, Bolívar F. Role of pyruvate oxidase in *Escherichia coli* strains lacking the phosphoenolpyruvate: carbohydrate phosphotransferase system. J Mol Microbiol Biotechnol. 2004;8:209–21. https://doi.org/doi:10.1159/000086702.

39. Causey TB, Shanmugam KT, Yomano LP, Ingram LO. Engineering *Escherichia coli* for efficient conversion of glucose to pyruvate. Proc Natl Acad Sci U S A. 2004;8:2235–40. https://doi.org/doi:10.1073/pnas.0308171100.

40. Aguilar C, Escalante A, Flores N, de Anda R, Riveros-Mckay F, Gosset G, Morett E, Bolívar F. Genetic changes during a laboratory adaptive evolution process that allowed fast growth in glucose to an *Escherichia coli* strain lacking the major glucose transport system. BMC Genomics. 2012;13:1–17. https://doi.org/doi:10.1186/1471-2164-13-385.

41. Rodriguez A, Martínez JA, Flores N, Escalante A, Gosset G, Bolivar F. Engineering *Escherichia coli* to overproduce aromatic amino acids and derived compounds. Microb Cell Factories. 2014;13:1–15. https://doi.org/doi:10.1186/s12934-014-0126-z.

42. Martínez K, de Anda R, Hernández G, Escalante A, Ramírez O, Bolívar F. Coutilization of glucose and glycerol enhances the production of aromatic compounds in an *Escherichia coli* strain lacking the phosphoenolpyruvate: carbohydrate phosphotransferase system. Microb Cell Factories. 2008;22:1–12. https://doi.org/doi:10.1186/1475-2859-7-1.

43. Matsuoka Y, Shimizu K. A new insight into the main metabolic regulation of *Escherichia coli* based on systems biology approach. 12th IFAC Symp Comput Appl Biotechnol. 2013;12:16–8.

44. Rodriguez A, Martínez JA, Millard A, Gosset G, Portais J-C, Létisse F, Bolívar F. Plasmid-encoded biosynthetic genes alleviate metabolic disadvantages while increasing glucose conversion to shikimate in an engineered *Escherichia coli* strain. Biotech Bioeng. 2017;114:1319–30. https://doi.org/doi:10.1002/bit.26264.

45. Lara AR, Caspeta L, Gosset G, Bolivar F, Ramírez OT. Utility of an *Escherichia coli* strain engineered in the substrate uptake system for improved culture performance at high glucose and cell concentrations: An alternative to fed-batch cultures. Biotech Bioeng. 2007;99:893–901. https://doi.org/doi:10.1002/bit.21664.

46. Luli G, Stohl W. Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations. Appl Environ Microbiol. 1990;4:1004–11. https://doi.org/doi:0099-2240/90/041004-08\protect\T1\textdollar02.00/0.

47. Namjoshi AA, Ramkrishna D. Multiplicity and stability of steady states in continuous bioreactors: Dissection of cybernetic models. Chem Eng Sci. 2001;56:5593–607. https://doi.org/doi:10.1016/S0009-2509(01)00166-X.

48. Kompala DS, Ramkrishna D, Jansen NB, Tsao GT. Investigation of bacterial growth on mixed substrates: Experimental evaluation of cybernetic models. Biotech Bioeng. 1986;28:1044–55. https://doi.org/doi:10.1002/bit.260280715.

49. Varner J, Ramkrishna D. The non-linear analysis of cybernetic models. guidelines for model formulation. J Biotechnol. 1999;71:67–104. https://doi.org/doi:10.1016/S0168-1656(99)00016-4.

50. Siddiquee KAZ, Arauzo-bravo MJ, Shimizu K. Effect of a pyruvate kinase (*pykF*-gene) knockout mutation on the control of gene expression and metabolic fluxes in *Escherichia coli*. FEMS Microbiol Lett. 2004;235:25–33. https://doi.org/doi:10.1016/j.femsle.2004.04.004.

51. Kabir MM, Shimizu K. Gene expression patterns for metabolic pathway in *pgi* knockout *Escherichia coli* with and without *phb* genes based on rt-pcr. J Biotechnol. 2003;105:11–31. https://doi.org/doi:10.1016/S0168-1656(03)00170-6.

52. Cui Y-Y, Ling C, Zhang Y-Y, Huang J, Liu J-Z. Production of shikimic acid from *Escherichia coli* through chemically inducible chromosomal evolution and cofactor metabolic engineering. Microb Cell Factories. 2014;13:1–11. https://doi.org/doi:10.1186/1475-2859-13-21.

53. Chen X, Li M, Zhou L, Shen W, Algasan G, Fan Y, Wang Z. Metabolic engineering of *Escherichia coli* for improving shikimate synthesis from glucose. Bioresour Technol. 2014;166:64–71. https://doi.org/doi:10.1016/j.biortech.2014.05.035.

54. Nahku R, Valgepea K, Lahtve P-J, Erm S, Abner K, Adamberg K, Vilu R. Specific growth rate dependent transcriptome profiling of *Escherichia coli* k12 mg16555 in accelerostat cultures. J Biotechnol. 2010;145:60–5. https://doi.org/doi:10.1016/j.jbiotec.2009.10.007.

55. Lyubetskaya AV, Rubanov LI, Gelfand MS. Use of the flux model of amino acid metabolism of *Escherichia coli*. Biochem. 2006;71:1256–60.

56. Arense P, Bernal V, Iborra JL, Cánovas M. Metabolic adaptation of *Escherichia coli* to long-term exposure to salt stress. Process Biochem. 2010;45:1495–67. https://doi.org/doi:10.1016/j.procbio.2010.05.022.

57. Takahisa K, Takeshi K, Masako S, Kazumi H, Masayuki I. Metabolic engineering of *Corynebacterium glutamicum* for shikimate overproduction by growth-arrested cell reaction. Metab Eng. 2016;38: 204–16. https://doi.org/doi:10.1016/j.ymben.2016.08.005.

58. Song H-S, Ramkrishna D. Prediction of dynamic behavior of mutant strains from limited wild-type data. Metab Eng. 2012;14:69–80. https://doi.org/doi:10.1016/j.ymben.2012.02.003.

59. Terzer M, Stelling J. Large-scale computation of elementary flux modes with bit pattern trees. Bioinformatics. 2008;24:2229–35. https://doi.org/doi:10.1093/bioinformatics/btn401.

60. Song H-S, Ramkrishna D. Reduction of a set of elementary modes using yield analysis. Biotech Bioeng. 2009;102:554–68. https://doi.org/doi:10.1002/bit.22062.

Metabolic reconstruction of *Pseudomonas chlororaphis* ATCC 9446 to understand its metabolic potential as a phenazine-1-carboxamide-producing strain

Fabián Moreno-Avitia[1] · José Utrilla[2] · Francisco Bolívar[1] · Juan Nogales[3] · Adelfo Escalante[1]

Abstract

Pseudomonas chlororaphis is a plant-associated bacterium with reported antagonistic activity against different organisms and plant growth–promoting properties. *P. chlororaphis* possesses exciting biotechnological features shared with another *Pseudomonas* with a nonpathogenic phenotype. Part of the antagonistic role of *P. chlororaphis* is due to its production of a wide variety of phenazines. To expand the knowledge of the metabolic traits of this organism, we constructed the first experimentally validated genome-scale model of *P. chlororaphis* ATCC 9446, containing 1267 genes and 2289 reactions, and analyzed strategies to maximize its potential for the production of phenazine-1-carboxamide (PCN). The resulting model also describes the capability of *P. chlororaphis* to carry out the denitrification process and its ability to consume sucrose (Scr), trehalose, mannose, and galactose as carbon sources. Additionally, metabolic network analysis suggested fatty acids as the best carbon source for PCN production. Moreover, the optimization of PCN production was performed with glucose and glycerol. The optimal PCN production phenotype requires an increased carbon flux in TCA and glutamine synthesis. Our simulations highlight the intrinsic H_2O_2 flux associated with PCN production, which may generate cellular stress in an overproducing strain. These results suggest that an improved antioxidative strategy could lead to optimal performance of phenazine-producing strains of *P. chlororaphis*.

Key points
• *This is the first publication of a metabolic model for a strain of P. chlororaphis.*
• *Genome-scale model is worthy tool to increase the knowledge of a non model organism.*
• *Fluxes simulations indicate a possible effect of H_2O_2 on phenazines production.*
• *P. chlororaphis can be a suitable model for a wide variety of compounds.*

Keywords *Pseudomonas chlororaphis* · Metabolic reconstruction · Genome-scale model · Phenazine-1-carboxamide · Metabolic engineering

---

✉ Adelfo Escalante
adelfo@ibt.unam.mx

1 Departamento de Ingeniería Celular y Biocatálisis, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, México

2 Programa de Biología de Sistemas y Biología Sintética, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, México

3 Departamento de Biotecnología de Sistemas, Centro Nacional de Biotecnología, CSIC, Madrid, España

Introduction

In recent years, academic and industrial biotechnological research groups have turned their attention to a broad group of organisms as improved metabolite production platforms. High production requirements require more robust organisms to face different stress conditions. Another biotechnological need is to find organisms with expanded metabolic capabilities, such as the ability to degrade contaminant compounds, consume inexpensive carbon sources, and produce desirable secondary metabolites. Under this approach, many organisms have been explored both in the laboratory and industrially.

Specifically, the genus *Pseudomonas* has attracted the attention of academia and industry because many species of this genus meet the requirements mentioned above. Different members of the *Pseudomonas* genus have been studied due to their capability to degrade contaminant compounds. Moreover, such strains have shown tolerance to a high concentration of fuels or other added-value compounds considered toxic for other model bacteria such as *Escherichia coli* (Nikel et al. 2014). The two best-studied *Pseudomonas* species are *P. aeruginosa* PAO1 and *P. putida* KT2440. Despite the potential of *P. aeruginosa,* the virulence and pathogenicity of this species are major limitations for many biotechnological applications. On the other hand, *P. putida* KT2440 has been widely used for the production of polyhydroxybutyrate (PHB) and other polyhydroxyalkanoates (PHAs) and proposed as a versatile platform for production and metabolic studies (Nogales et al. 2008; Gutierrez et al. 2013; Martínez-García et al. 2014; Nikel et al. 2014; Loeschcke and Thies 2015; Nogales et al. 2017). *P. putida* KT2440 has demonstrated its relevance in metabolic engineering approaches, but there is still enormous untapped potential in other *Pseudomonas* species. Currently, the number of described *Pseudomonas* species is over 200, all of which are relevant and particularly attractive (Koehorst et al. 2016); nevertheless, among them, only a few strains have been characterized.

The extensive repertoire of secondary metabolites produced by members of the genus *Pseudomonas* includes relevant compounds such as pyrrolnitrin and phenazines, with antifungal and antibiotic activity (Gross and Loper 2009). These compounds are typically found in plant growth–associated *Pseudomonas* species, such as *P. chlororaphis*, one of the best-studied phenazine-pyrrolnitrin producers in addition to *P. aeruginosa,* with the additional advantage of being an innocuous bacterium to plants and humans. *P. chlororaphis* possesses a strong metabolic background suitable for the overexpression of biochemical pathways. This property is shared among *Pseudomonas* and characterized by the use of the Entner-Duodoroff pathway as the primary means of carbohydrate catabolism, promoting a high rate of NADH regeneration compared with other model organisms (Nikel et al. 2014).

The interest in *P. chlororaphis* as a model organism for metabolite production has increased in recent years due to its role as a biocontrol agent when growing in its natural environment (Chin-A-Woeng et al. 2007; Pierson and Pierson 2010; Calderón et al. 2015; Wullimann et al. 2016; Huang et al. 2018; Yu et al. 2018). Different strains of this *Pseudomonas* show significant antagonistic effects against many different organisms, such as rootworms, *Fusarium oxysporum*, *F. graminearum*, *Gaeumannomyces graminis*, *Phytophthora capsici*, *P. ultimum*, *Sclerotinia* spp., and more examples found in the literature (Wullimann et al. 2016; Huang et al. 2018; Yu et al. 2018). A second relevant feature

of *Pseudomonas* sp. is the increasing number of sequenced genomes of different strains, commonly used to establish and study phylogenetic relationships (Shen et al. 2012; Liu et al. 2014; Loewen et al. 2014; McCully et al. 2014; Takeuchi et al. 2014; Deng et al. 2015; Town et al. 2016; Moreno-Avitia et al. 2017). In the NCBI genome database, 80 assembled genomes of *P. chlororaphis* have been deposited, almost all of which have been published in the last 5 years. To increase the information available on *P. chlororaphis*, we recently sequenced the genome of the strain ATCC 9446. This strain has been reported as a triphenyltin (organometallic contaminant)-degrading bacterium, and this characteristic, along with other degradation clusters found in its genome, suggests that this organism has high potential in bioremediation (Inoue et al. 2000; Inoue et al. 2003).

A more relevant application of *P. chlororaphis* is its capability to produce different compounds, taking advantage of its natural features. This organism has served as a new platform for the production of trans-2,3-dihydro-3-hydroxyanthranilic acid, 2-hydroxy-phenazine-1-carboxylic acid, phenazine-1-carboxamide, 2-hydroxyphenazine, rhamnolipids, indole-3-acetic acid, muconic acid, and PHAs (Gunther et al. 2005; Dimkpa et al. 2012; Liu et al. 2016a; Wang et al. 2016; Hu et al. 2017; Sharma et al. 2017; Peng et al. 2018b; Wang et al. 2018; Yao et al. 2018). These examples clarify the relevance of *P. chlororaphis* as a promising production model for a wide variety of metabolites, some of them produced only by this bacterium.

New genomic, transcriptomic, and proteomic technologies allow researchers to acquire a large amount of phenotypic information on an organism in a short time and feed genome-scale computational models of metabolism. Consistent efforts to construct metabolic models have been performed for different *Pseudomonas* species, offering a track to build models of other species (Nogales et al. 2008; Oberhardt et al. 2011; Borgos et al. 2013; Nogales et al. 2017). These models have allowed the study of the topology of the metabolic network, visualizing the interaction of the compounds at the cellular system level and exploring their metabolic capabilities. In this work, we report the construction of a genome-scale model of *Pseudomonas chlororaphis* ATCC 9446 (PC9446) to explore its metabolic potential for phenazine production.

Methods

Metabolic draft model construction

As a starting point, we used the genome sequence of PC9446 available in GenBank under the accession number GCA_002095925.1 to explore the gene repertoire. Phylogenetically related organisms share a high number of

metabolic reactions, mainly in central metabolism (Moreno-Avitia et al. 2017). For this reason, we used the iJN1411 model of *P. putida* KT2440, one of the complete models to date, as a mold to build the first draft model of *P. chlororaphis* PC9446 (Nogales et al. 2017). Using their annotated protein sequences, we compared both organisms looking for orthologous proteins using the webserver OrthoVenn (Wang et al. 2015). The set of orthologous sequences was compared with the proteins included in the iJN1411 model, and those reactions carried by the shared enzymes were transferred to build the draft model of PC9446 (iMA1267) using the webserver MrBAC (Liao et al. 2011).

Manual gap filling and model expansion

Manual gap filling was performed to avoid the automatic addition of reactions that may disagree with reality. A genome-scale metabolic model can be used to predict cellular growth in a certain nutritional environment using the biochemical reactions encoded in the genome. To validate this prediction, it is necessary to add a biomass formation reaction to the model with information about all the biomass precursors with the correct stoichiometry. We selected the core biomass reaction from iJN1411 as the starting point.

Once the biomass reaction was fixed, we looked for those metabolites that were blocking biomass production and then looked for the reactions that produce such metabolites. This search was performed with multiple optimizations of the iMA1267 model using FBA with the Python module COBRApy. The addition of the reactions was determined by the presence of the coding gene of the responsible enzyme in the genome of *P. chlororaphis* PC9446.

Once the model was able to simulate growth, we looked for those reactions that may not be essential for growth or that occur in secondary metabolism. This step provides robustness to the model, allowing it to acquire different phenotype profiles under different constraining conditions. Moreover, the addition of secondary metabolism reactions enriches the model with compounds that may be used for different applications. For the model extension, we looked for orthologous genes from *P. aeruginosa* PAO1 and *E. coli* with the same tools mentioned above. Then, the annotation was manually reviewed for particular reactions from *P. chlororaphis* PC9446, comparing the sequence and genomic contexts with those of other *Pseudomonas* and other organisms using the Pseudomonas Genome Database (Version: 19.1) (http://www.pseudomonas.com) and KEGG (release 87.0 to release 94.1) (https://www.kegg.jp) (Ogata et al. 1998; Winsor et al. 2016). The added reactions, genes, and metabolites were reviewed from the USCD model database BiGG (version 1.6) (http://bigg.ucsd.edu), MetaCyc database (version 22.0 to version 24.0) (https://metacyc.org), and Pseudomonas Genome DB (Version: 19.1) (https://www.pseudomonas.com/) (King

et al. 2016; Winsor et al. 2016; Caspi et al. 2016; Norsigian et al. 2019).

Model manipulation and analysis

The model was analyzed using FBA (flux balance analysis) and its FVA (flux variability analysis) approximations through the diverse functional repertoire of COBRApy and CAMEO (Schellenberger et al. 2011; Ebrahim et al. 2013; Cardoso et al. 2018). The addition or deletion of genes and reactions was performed using the same Python modules. To determine the ability of the model to grow in different conditions, gene essentiality and yield calculations were performed by constraining the model with the desired values (or range of values) combined with the suitable Python script.

Phenotypic information acquisition

To acquire an extensive phenotypic profile from PC9446, we evaluated its carbon utilization capabilities using the Biolog system (Biolog Inc., Hayward, CA, USA; http://www.biolog.com). The Biolog PM1 plate consists of a 96-well plate with 95 different carbon sources, wherein oxidation of tetrazolium dye determines the utilization of the desired carbon source. The experiment was performed as recommended by the manufacturer.

The specific growth rate (μ, h^{-1}) was determined in 1-L bioreactor cultures (Applikon Ez controller) with an operating volume of 500 mL using M9 medium with 2 g/L of the respective carbon sources, and the optical density was read at 600 nm (OD_{600nm}). The culture started at 0.01 OD_{600nm}; the airflow rate was set at one vvm with agitation at 600 rpm, and the temperature was set at 30 °C. Additionally, continuous cultures (chemostats) were established to evaluate the carbon sources and oxygen uptake rates under steady-state conditions. The inlet medium flow rate was calculated to maintain three different dilution rates: 0.12, 0.25, and 0.37 h^{-1}, using a Masterflex peristaltic pump. The culture conditions were the same as those used in the batch cultures.

Results

Model reconstruction and validation

The initial iMA1267 model was obtained from the direct comparison of orthologous genes from PC9446 and genes included in the iJN1411 model. As iJN1411 is a meticulously curated model, this step offered a strong scaffold to build an accurate model. For the extension of the model, we looked for orthologous genes in other well-studied *Pseudomonas*. Figure 1 shows the shared genes among *P. aeruginosa* PAO1, *P. putida* KT2440, and *P. chlororaphis* PC9446
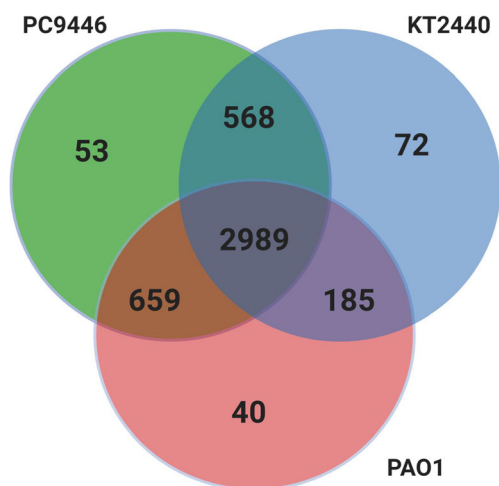
Fig. 1 Clusters shared between *P. chlororaphis* ATCC9446, *P. putida* KT2440, and *P. aeruginosa* PAO1. The figure shows a Venn diagram of the clusters in the genomes of the listed *Pseudomonas* and plots a comparison of its number of predicted clusters

classified by clusters. In this figure, we can observe that these three *Pseudomonas* species share approximately three thousand clusters, which could include the shared genes among many other *Pseudomonas* species. Additionally, PC9446 shares 568 and 659 clusters with *P. putida* KT2440 and *P. aeruginosa* PAO1, respectively.

A meticulous search was carried out to select the genes present only in *P. chlororaphis* PC9446. After the addition of the manually curated genes/reactions and modification of the biomass reaction, the new iMA1267 model contained 1267 genes, 2289 reactions, and 1815 metabolites. This model covered 20% of the 6047 predicted genes in the sequenced genome of PC9446.

As mentioned, phenazines are significant secondary metabolites produced by *P. chlororaphis* PC9446. The gene cluster responsible for producing such compounds was identified by BLAST and included in the model. Among the phenazine-producing organisms, variation exists in the specific phenazine derivative compounds they can produce; this difference is due to the genes included in such clusters (Bilal et al. 2017). In this case, *P. chlororaphis* PC9466 possesses the *phzABCDEFGH* operon (Fig. 2), which indicates that the main produced phenazine is phenazine-1-carboxamide (PCN). In the case of the pyrrolnitrin cluster, the genes were misannotated, but with the aid of the ANTISMASH server (https://antismash.secondarymetabolites.org/#!/start), a cluster was identified (Medema et al. 2011). This cluster is composed of the genes *prnC*, *prnB*, *prnA*, and *prnD* and is responsible for the conversion of L-tryptophan into pyrrolnitrin. This evidence suggests PC9446 as a model for pyrrolnitrin production.

To explore the capabilities of *P. chlororaphis* PC9446 to oxidize different carbon sources, the strain was assayed in a

Biolog plate with 95 different carbon compounds. This determination was qualitative, and we only evaluated the results as positive or negative. Table 1 shows the carbon sources potentially metabolizable by PC9446, showing the positive oxidation of 49 carbon sources, which represented 50% of the tested compounds, listed in the table below. These results match those reported in the literature (Burr et al. 2010).

As shown in Table 1, PC9446 can metabolize sucrose (Sucr), maltose, trehalose, mannose, and galactose, and carbohydrates rarely found to be consumed by the best-studied *Pseudomonas* and they are not included at their metabolic models (Oberhardt et al. 2011; Löwe et al. 2017; Löwe et al. 2018; Moehario et al. 2019; Nogales et al. 2020). The utilization of these compounds by strain PC9446 was confirmed in cultures of 3 mL of M9 medium supplemented with 2 g/L of each carbohydrate. To introduce the uptake reactions of the mentioned carbohydrates into the model, we searched for misannotated genes and their related clusters in the genome of PC9446. In the case of Scr, there are two well-studied systems in Gram-negative bacteria, the *csc* and *scr* genes. The *csc* operon is present in some *E. coli* and *Salmonella* strains (Löwe et al. 2018; Alva et al. 2020). Interestingly, the sequence comparison of these genes reveals that such operons are absent in the PC9446 genome. Another less studied Scr uptake operon (*cscRABY*) in closer organisms, such as *P. protegens* Pf-5, *P. fluorescence* AU1385, and the strain AU20219, was reported (Löwe et al. 2018) and presented similitudes to both the *csc* and *scr* clusters. Again, when comparing the sequence of the *cscRABY* system with the PC9446 genome using BLAST, no relevant similitudes were found. The only system in strain PC9446 able to transport disaccharides is the phosphotransferase system (PTS) for trehalose. This system was inferred from the close structure of the sequences ORM46423.1, ORM46424.1, and ORM46425.1 found in strain PC9446 to the trehalose degradation operon from *E. coli* (*treRBC*) and was supported by experimental evidence of trehalose uptake (Steen et al. 2014). Some reports suggest some transport promiscuity to other disaccharides, such as Scr, in this system in *E. coli* (Steen et al. 2014). We hypothesize that a similar phenomenon could occur in sugar transport in strain PC9446 (Steen et al. 2014).

The iMA1267 model was used to simulate growth under different conditions, maximizing biomass production; this was performed using FBA in COBRApy. To evaluate the utilization of carbon sources, the unique constraints used were set to zero for the uptake of any carbon source different than the tested compound. The oxygen uptake rate was set to a fixed value of 10 mmol/gDCWh. The mineral, nitrogen, and phosphate uptake rates were kept open and unconstrained. The iMA1267 model was able to predict growth in 96 carbon sources. The discrepancies found between growth-sustaining metabolites in the IMA1267 model and BIOLOG PM1 plate did not allow the comparison of some metabolites. Of the 49
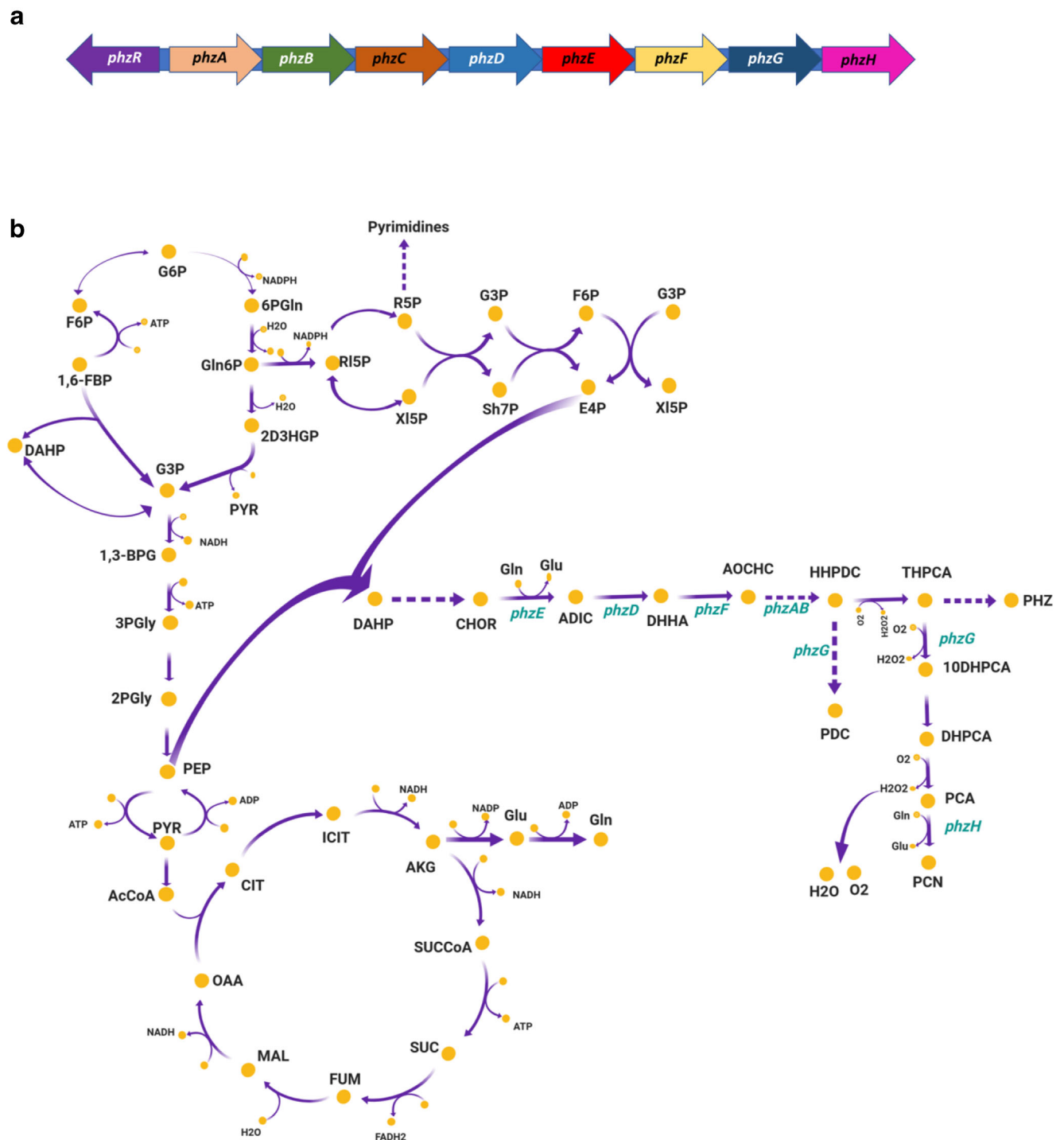
a

b

Fig. 2 **a** Phenazine pathway genetic cluster. The figure shows the organization of the genes involved in the synthesis of PCN in *P. chlororaphis* ATCC 9446. **b** The scheme shows a summarized pathway for the production of phenazine using glucose as a carbon source. Words in blue are the genes involved in phenazine biosynthesis, and those in black are the abbreviations of the compounds. G6P: glucose-6-phosphate; F6P: fructose-6-phosphate, 1,6-fructose-biphosphate; DAHP: 3-deoxy-arabino-heptulosonate-7-phosphate; CHOR: chorismate; G3P: glyceraldehyde-3-phosphate; 1,3-BPG: 1,3-bisphosphoglycerate; 3PG: 3-phosphoglycerate; 2: 2-phosphoglycerate; PEP: phosphoenolpyruvate; PYR: pyruvate; AcCoA: acetyl-coenzyme A; CIT: citrate; ICIT: isocitrate; AKG: alpha-ketoglutarate; SUCCoA: succinyl-coenzyme A; SUCC: succinate; FUM: fumarate; MAL: malate; OAA: oxaloacetate; Glu: glutamate; Gln: glutamine; 6PGln: 6-phospho-D-glucono-1,5-lactone; Gln6P: gluconate-6-phosphate; 2D3HGP: 2-dehydro-3-deoxy-gluconate-6-phosphate; RL5P: ribulose-5-phosphate; R5P: ribose-5-phosphate; XI5P: xylulose-5-phosphate; Sh7P: sedoheptulose-7-phosphate; E4P: erythrose-4-phosphate; ADIC: 2-amino-2-deoxyisochorismate; DHHA: trans-2,3-dihydro-3-hydroxyanthranilate; AOCHC: 6-amino-5-oxocyclohex-2-ene-1-carboxylic acid; HHPDC: 1,4,5,5a,6,9-hexahydrophenazine-1,6-dicarboxylate; PHZ: phenazine; PDC: phenazine-1,6-dicarboxylate; THPCA: 1,4,5,10-tetrahydrophenazine-1-carboxylate; DHPCA: 5,10-dihydrophenazine-1-carboxylate; 10HPCA: dihydrophenazine-1-carboxylate; PCA: phenazine-1-carboxylate; PCN: phenazine-1-carbozamide

Table 1 Compounds metabolizable by *P. chlororaphis* PC9446

Compound name	Result	Compound name	Result	Compound name	Result	Compound name	Result
L-Arabinose	−	D-Alanine	+	m-Tartrate	−	Adonitol	−
N-Acetyl-glucosamine	+	D-Trehalose	+	D-Glucose-1-phosphate	−	Maltotriose	−
D-Saccharic acid	+	D-Mannose	+	D-Fructose-6-phosphate	−	2-Deoxyadenosine	−
Succinate	+	Dulcitol	−	Tween 80	+	Adenosine	−
D-Galactose	+	α-Glycerol phosphate	−	α-Hydroxyglutarate-γ-lactone	+	Propionate	+
D-Serine	+	D-Xylose	−	Glycyl-L-aspartate	-	Mucate	+
D-Sorbitol	-	L-Lactate	+	Citrate	+	Glycolate	−
Glycerol	+	Formate	−	m-Inositol	+	Glyoxylate	−
L-Fucose	−	D-Mannitol	+	D-Threonine	-	D-Cellobiose	−
D-Glucuronic acid	−	L-Glutamate	+	Fumarate	+	Inosine	+
D-Gluconate	+	D-Fructose	+	Glycyl-L-glutamate	+	Acetoacetate	−
D-Glucose-6-phosphate	−	Acetate	+	Tricarballylic acid	-	N-Acetyl-β-D-mannosamine	−
D-Galatonate-6-phosphate	+	α-D-Glucose	+	L-Serine	+	Monomethyl succinic acid	−
D-Malate	+	Maltose	+	L-Threonine	+	Methylpyruvic acid	−
L-Malate	−	D-Melibiose	−	L-Alanine	+	Glucuronate	−
Tween 20	+	Thymidine	−	L-Alanyl-glycine	-	Pyruvate	+
L-Rhamnose	−	α-Ketobutyrate	−	Glycyl-L-proline	+	L-Galactonate-γ-lactone	−
L-Asparagine	+	α-Methyl-galactoside	−	p-Hydroxyphenylacetate	+	D-Galacturonate	−
D-Aspartate	+	α-D-lactose	−	m-Hydroxyphenylacetate	-	Phenylmethylamine	−
D-Glucosaminic acid	+	Lactulose	−	Tyramine	+	2-Aminoethanol	+
1,2-Propanediol	−	Sucrose	+	D-Psicose	-	L-Proline	+
Tween 40	+	Uridine	+	L-Lyxose	-	L-Glutamine	−
α-Ketoglutarate	+	α-Hydroxybutyrate	−	L-Aspartate	+	β-Methyl-D-glucoside	−

Compounds metabolizable by *P. chlororaphis* PC9446. This phenotypic analysis shows the carbon sources that PC9446 can metabolize (labeled with "+") and those that it cannot (labeled with "−")

growth-promoting carbon sources, 26% were not included in the iMA1267 model due to an initial lack of information, providing a direct guide to improve future versions of the model. The negative results in vivo for the other 46 compounds were consistent with the in silico results; no false positives were found in the model growth predictions. Originally, the simulations resulted in 16 false negatives; this result is understandable due to the diversity of the tested compounds. Among these compounds, we can find dipeptides, Tween 20, Tween 40, and Tween 80. The main reason of these false negatives is that the metabolic pathways of these molecules may not be clearly inferred from genome annotation. A 19% of the false negatives were fixed by including the uptake reactions for propionic acid, inositol, and inosine to the model. Supplementary Table S1 compares the results of in vivo and in silico growth predictions.

To evaluate quantitative model predictions, we compared in silico and in vivo μ values in experiments using different carbon sources in 500-mL bioreactor batch cultures with 2 g/L glucose (Glc), Scr, and glycerol (Gly). The strain PC9446 was able to grow at relatively high growth rates in the three carbon sources tested, with Glc being the carbon source that produced the highest μ (= 0.44 h^{-1}), closely followed by Scr (μ = 0.39

h^{-1}), and finally Gly (μ = 0.30 h^{-1}). The simulated μ values using the experimental substrate uptake rate predicted values close to the experimental data with values of 0.49 h^{-1}, 0.44 h^{-1}, and 0.34 h^{-1} for Glc, Scr, and Gly respectively (Table 2).

The model closely predicted the real values at 0.12, 0.25, and 0.37 h^{-1} dilution rates (Table 3). This result gives us a general idea of the accuracy of the current model. This experiment also allowed the calculation of the flux required for the ATP maintenance reaction (ATPM = 3.41 mmolATP/gDCWh) to constrain the model.

Table 2 Comparison of batch culture and in silico results

Conditions		O$_2$ uptake (mmol/gDCWh)	q_s (mmol/gDCWh)	μ (h^{-1})
In vivo Glc	NA		5.16	0.44
In silico Glc	10.3		5.16	0.49
In vivo Scr	NA		2.3	0.39
In silico Scr	−		2.3	0.44
In vivo Glc	NA		5.9	0.30
In silico Glc	−		5.9	0.34

Glc, Glucose; Scr, sucrose; *NA*, not available

Table 3 Comparison of chemostat and in silico results

D (h^{-1})	q_s (mmolGLC/gDCWh)	q_{O2} (mmolO$_2$/gDCWh)	μ (h^{-1})*	q_{O2}* (mmolGLC/gDCWh)
0.37	4.30	9.01	0.40	8.67
0.25	3.48	NA	0.32	7.14
0.12	2.08	NA	0.18	4.53

NA, not available

*In silico calculated data

Another relevant metabolic feature to include in the model was the denitrification capability of strain PC9446. Different strains of *P. chlororaphis* have been reported as denitrification organisms with the unique characteristic of being unable to produce nitrogen from nitrate and producing nitrous oxide instead (Greenberg and Becker 1977; Sutka et al. 2006). Based on the genome of PC9446, this strain can reduce nitrate to produce nitrous oxide, but it lacks nitrous oxide reductase, similar to the reported strains. The operon *narHGI* was found to be responsible for the reduction of nitrate to nitrite; then, *narK* converts nitrite into nitric oxide, and finally, nitrous oxide is produced by *norBC*. These reactions were added to the model to account for the denitrification capability. The capability for ATP production predicted by the model was rated against experimental data from *P. denitrificans* due to the lack of information for PC9446 and was expressed as a percentage of the ATP obtained with respect to the use of nitrate as an electron acceptor (Table 4) (Koike and Hattori 1975).

Phenazine production simulation with iMA1267

Intending to design a phenazine production derivative of strain PC9446, we used the iMA1267 model to simulate and maximize PCN production. First, we analyzed the model to find the most suitable carbon source to produce PCN. We tested all the metabolizable carbon sources that were able to sustain growth and compared their PCN yields (Fig. 3). The highest yields were obtained using medium-chain fatty acids and short peptides. Fatty acids such as hexanoate, heptanoate, octanoate, and nonanoate showed a higher cofactor yield than Glc, and these differences arise from fatty acid oxidation, which produces the FADH and NADH that participate in the

Table 4 Comparison of in silico denitrification and experimental data

Electron acceptor	molATP/molA	% respect NO$_3^-$	% respect NO$_3^-$*
NO$_3^-$	2.8	100	100
NO$_2^-$	1.4	50	65
NO	0.7	25	35

*Experimental data (Koike and Hattori 1975)

production of ATP. These compounds also facilitated a higher NADPH yield, which is needed for the synthesis of shikimic acid, a PCN precursor. Gly as a carbon source also showed a higher PCN yield than Glc, and it is an inexpensive substrate already applied in large-scale fermentation; therefore, it may be a suitable carbon source.

Among the peptides, we found leucyl-leucine and alanyl-leucine to be the best carbon sources. However, even if the best yield was obtained with the mentioned compounds, it does not mean that they can sustain high productivity due to the lack of experimental data, and more importantly, they may not be suitable for implementation in an industrial process.

Metabolic model reconstruction gives us the advantage of accelerating the exploration of metabolism under different conditions, with the overproduction of some desirable compound being one of the most exploited approaches. In this case, we decided to perform a metabolic analysis to identify the metabolic modifications required to improve PCN production using Glc and Gly as carbon sources and to understand the metabolic flux distributions that maximize PCN production. We used flux scanning based on the enforced objective flux (FSEOF) algorithm (FVA method extension) (Choi et al. 2010) (Fig. 4, Supplementary Table S2a and Table S2b). This algorithm selects the fluxes in the metabolic network that increase when product formation is enforced and uses them as an additional constraint; finally, it proposes a set of reactions in which the genes involved need to be overexpressed (Fig. 5). Among the most relevant modifications proposed by the FSEOF algorithm is the increased flux in the pentose phosphate pathway (PPP) through the transketolase (TKT1 and TKT2) and transaldolase (TALA) reactions in a factor higher than fivefold. On the other hand, FSEOF predicts that reactions that compete for pentose and ribose-5-phosphate should be diminished in their metabolic flux (phosphopentomutase, PPM; ribose-1,5-bisphosphate phosphokinase, R15BPK; and ribose-1-phosphokinase, R1PK). In addition, the reactions catalyzed by ribose-5-phosphate isomerase (RPI) and ribulose5-phosphate 3-epimerase (RPE), which produce ribose-5-phosphate and D-ribulose5-phosphate, respectively, should be increased. Due to the different mechanisms of internalization of the utilized carbon sources, simulation using Glc shows a flux increase in phosphogluconate
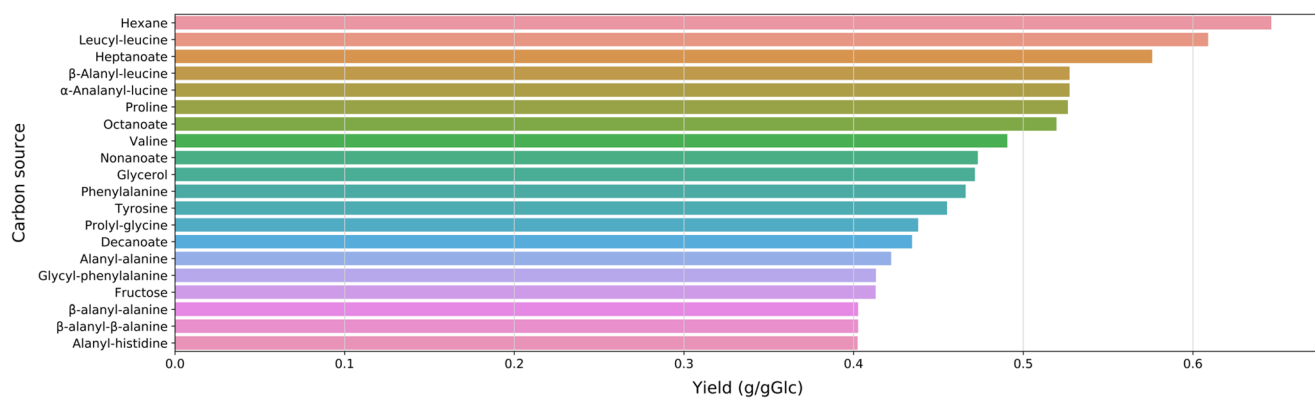
Fig. 3 Carbon source optimization. The bar plot shows the twenty best carbon sources that produce PCN with a higher yield than Glc (0.39 gPCN/gGlc), with medium-chain fatty acids and short peptides being the best substrates for this purpose (from the compounds included in the iMA1267 model)

dehydrogenase (GND). Following this idea, optimization using Glyc as a carbon source employs gluconeogenesis to produce fructose-6-phosphate. Along with these changes, a decrease in the flux of pyruvate kinase (PYK) is needed when using Glc. This result was confirmed with previous reports focused on the production of aromatic compounds in other organisms, wherein the inactivation of the *pykF* (pyruvate kinase I) gene improved shikimic acid yield (Martínez et al. 2015). The reduction of the flux through PYK appears not to be essential in the simulation with Gly. Reactions related directly or indirectly to the lower path of glycolysis show more activity, such as glycerol dehydrogenase (ALCD19) and glycerol kinase (GLYK) in both simulated conditions and in the phosphoglycerate kinase reaction using Glc.

The production of PCN requires an increase in the flux of all phenazine and shikimic acid production pathway reactions. An important metabolite in the synthesis of PCN is glutamine, which is the donor of two amino groups in two reactions of the PCN production pathway. The reactions involved in the formation of intermediaries (α-ketoglutarate and glutamate) and the direct production of glutamine need to be enforced. The increase in the flux of the α-ketoglutarate dehydrogenase (AKGDbr/AKGDBa) reaction leads to an increment in α-

ketoglutarate availability for glutamate production and subsequently conversion to glutamine by the glutamine synthetase reaction (GLNS). The availability of glutamate is augmented by a decrease in the reaction rate of valine and leucine transaminases that use this compound as an amino group donor.

This optimization suggests an increased flux across the tricarboxylic acid cycle (TCA) and the production of succinate through isocitrate lyase (ICL), accompanied by the production of glyoxylate that needs to be reincorporated to central carbon metabolism. In this case, glyoxylate is converted to 2-hydroxy-3-oxopropanoate by glyoxylate carboligase (GLXCL) and then to hydroxypyruvate by hydroxypyruvate isomerase (HPYRI). Finally, hydroxypyruvate is converted to glycerol-3-phosphate by hydroxypyruvate reductase (HPYRRy or GLYD) and glycerol kinase (GLYK) sequentially. A similar result was obtained with the simulation using Gly as a carbon source.

Active PCN production leads to high hydrogen peroxide (H_2O_2) flux in both conditions. It is known that phenazines are closely related to cellular mechanisms to overcome oxidative stress in different environments (Girard and Rigali 2011; Xie et al. 2013). This model tends to deal with high H_2O_2 production with an increase in catalase activity (CAT). To understand

Fig. 4 Flux scanning based on enforced objective flux (FSEOF) analysis for PCN production. This analysis tries to predict a new flux distribution that enforces the production of the desired metabolite target. In this case, it was implemented to increase the yield of PCN. The figure shows some reactions with essential changes in flux values
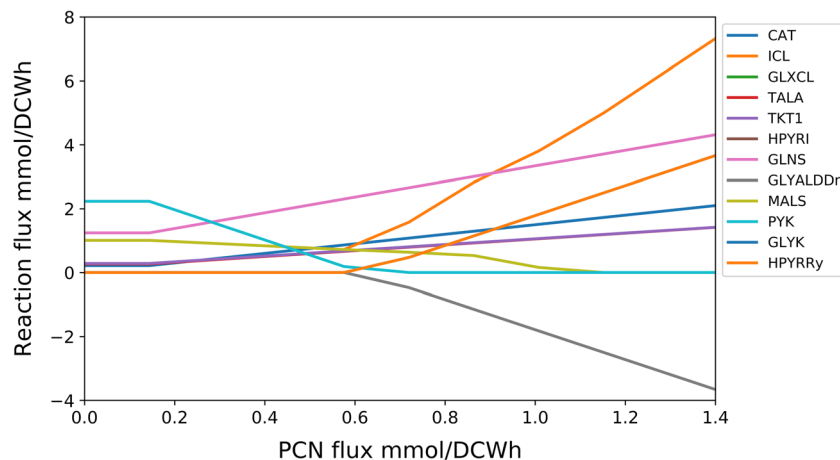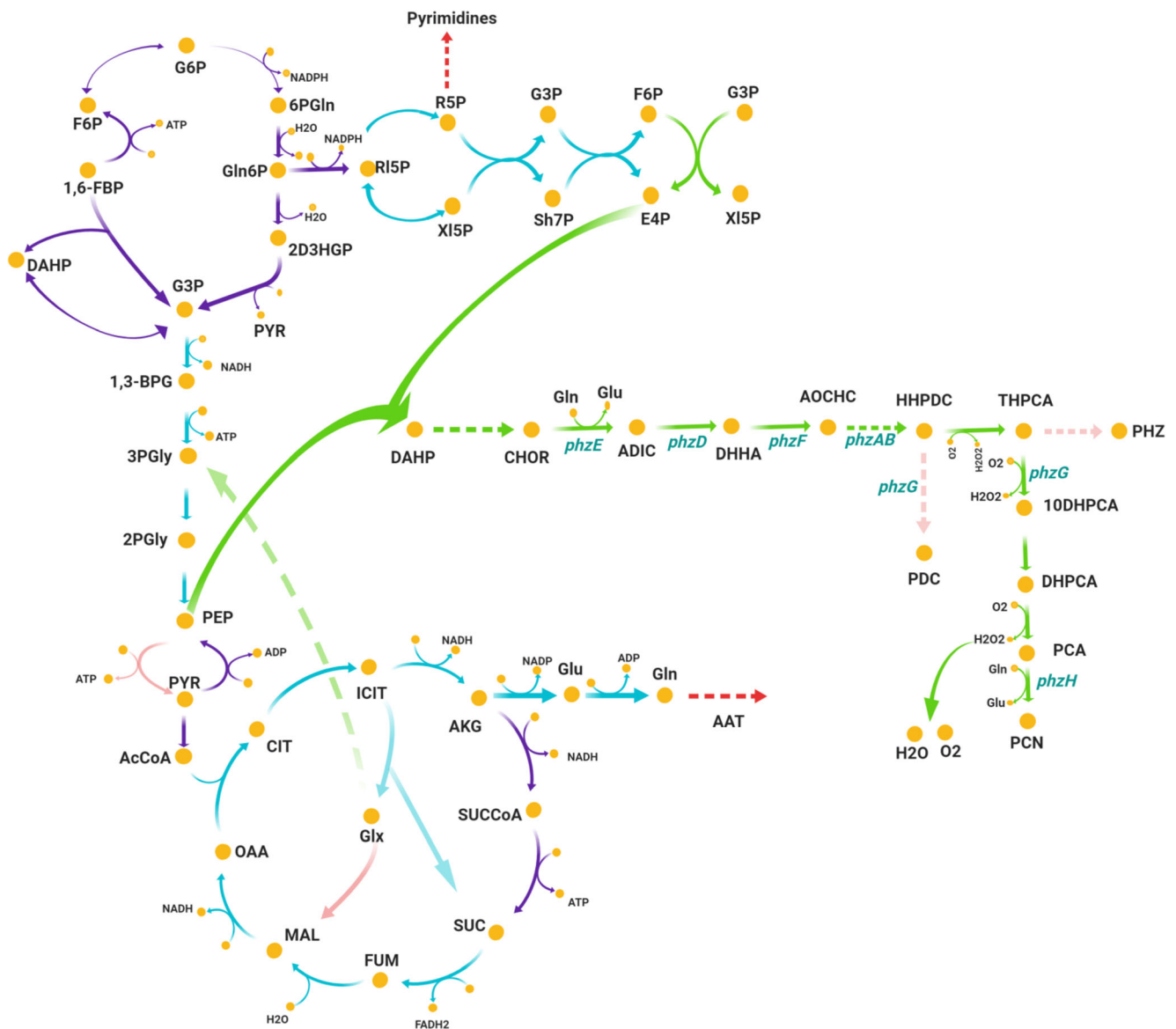
Fig. 5 Suggested flux distribution change to central metabolism for PCN production. The reactions that may be downregulated are shown in red, and those that need to be blocked are shown in light red. Reactions in blue need to be increased in a factor from 2- to 5-fold. Reactions in green are increased by a factor higher than 6-fold. Dashed lines represent multiple reactions. G6P: glucose-6-phosphate; F6P: fructose-6-phosphate, 1,6-fructose-biphosphate; DAHP: 3-deoxy-arabino-heptulosonate-7-phosphate; CHOR: chorismate; G3P: glyceraldehyde-3-phosphate; 1,3-BPG: 1,3-bisphosphoglycerate; 3PG: 3-phosphoglycerate; 2: 2-phosphoglycerate; PEP: phosphoenolpyruvate; PYR: pyruvate; AcCoA: acetyl-coenzyme A; CIT: citrate; ICIT: isocitrate; AKG: alpha-ketoglutarate; SUCCoA: succinyl-coenzyme A; SUCC: succinate; FUM: fumarate; MAL: malate; OAA: oxaloacetate; Glu: glutamate;

Gln: glutamine; 6PGln: 6-phospho-D-glucono-1,5-lactone; Gln6P: gluconate-6-phosphate; 2D3HGP: 2-dehydro-3-deoxy-gluconate-6-phosphate; RL5P: ribulose-5-phosphate; R5P: ribose-5-phosphate; Xl5P: xylulose-5-phosphate; Sh7P: sedoheptulose-7-phosphate; E4P: erythrose-4-phosphate; ADIC: 2-amino-2-deoxyisochorismate; DHHA: trans-2,3-dihydro-3-hydroxyanthranilate; AOCHC: 6-amino-5-oxocyclohex-2-ene-1-carboxylic acid; HHPDC: 1,4,5,5a,6,9-hexahydrophenazine-1,6-dicarboxylate; PHZ: phenazine; PDC: phenazine-1,6-dicarboxylate; THPCA: 1,4,5,10-tetrahydrophenazine-1-carboxylate; DHPCA: 5,10-dihydrophenazine-1-carboxylate; 10HPCA: dihydrophenazine-1-carboxylate; PCA: phenazine-1-carboxylate; PCN: phenazine-1-carbozamide. The genes whose products catalyze the indicated reaction are shown in blue.

the H_2O_2 flux behavior, we simulated the effect of increasing the PCN flux (Fig. 6). This figure shows the impact of a higher demand for oxygen and the side effects related to oxidative stress carried by a higher PCN yield achieved with CAT. This behavior may have effects on the phenotype due to oxidative stress.

Discussion

The first theme to discuss is the importance of using a closely related organism model to build a draft model. The genus *Pseudomonas* is a wide and diverse group of bacteria that share large metabolic traits, mainly in the central network.

Fig. 6 Effect of PCN production on the CAT flux, oxygen uptake, and growth rate. The plot shows the increase in the oxygen uptake rate (O₂, green) due to the higher metabolic demand of producing PCN. The same is shown for the catalase reaction (CAT, red). The growth rate (blue) is affected as a side effect of getting closer to the maximum PCN yield
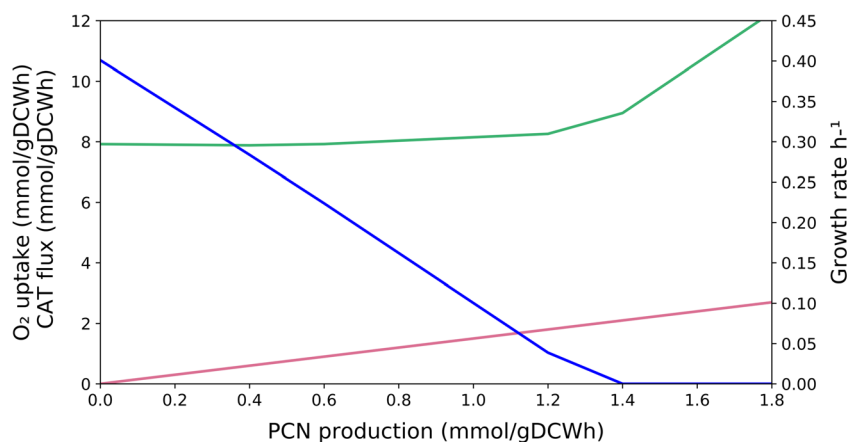
Nevertheless, when one moves away from central metabolism, the task of acquiring information becomes more challenging due to inconsistency in the annotation, forcing the model developers to curate information manually. The use of the iJN1411 model allowed us to speed up the reconstruction of the draft model of PC9446 and then extend it to obtain a well-curated specific model. As mentioned above, the iMA1267 model covers 21% of the predicted coding sequences, which means that a large portion of the metabolism is included. The large number of orthologous genes between *P. chlororaphis* PC9446, *P. aeruginosa* PAO1, and *P. putida* 2440, two well-studied *Pseudomonas*, allowed for an easier search for the addition of metabolic reactions with more accuracy.

P. chlororaphis PC9446 showed the capability of growing in 48 of the 95 tested carbon sources, some of which are attractive for different approaches. For example, strain PC9446 was able to degrade Tween 20, 40, and 80 due to the presence of esterases in its enzymatic pool. Moreover, the genome of this organism presents lipases that are not well characterized and other enzymes with biodegradation applications such as polyurethanase B (*pueB*), which has been reported in other organisms (Hung et al. 2016), highlighting the potential of strain PC9446. Some of the assayed carbon sources, such as Scr, maltose, trehalose, mannose, and galactose, are not metabolizable by *P. putida* KT2440. The capability of consuming Scr is particularly relevant from an industrial application perspective. In recent works, researchers have introduced this capability to *P. putida* KT2440 to explore the use of this inexpensive energy source (Löwe et al. 2017; Löwe et al. 2018). From this perspective, *P. chlororaphis* PC9446 can be an interesting alternative to *P. putida* due to its native capacity to efficiently use Scr, thus making it suitable for low-cost industrial processes.

In the analysis of the carbon source utilization profile, no false positives were found during the simulations. This result means that the model was not overannotated; it did not contain exchange reactions induced by a missannotation of the

genome or by manual curation. On the other hand, our simulations showed 13 false negatives (Supplementary Table S1), meaning that the model failed to predict growth when it was experimentally determined. This error is common during the model construction process and is completely reconcilable by adding the corresponding reactions, once found, by experimental evidence or updated annotations (Thiele and Palsson 2010). A strong source of information to understand the pathways involved in the metabolism of the compounds with a false negative is a transcriptomic analysis from cultures using these carbon sources as proposed by Guzmán et al. (2015). False positives appear when the metabolic network reconstructed in the model is not wide enough to include such compound uptake, or the consumption pathway is not well defined, with the latter being the most common scenario (Nogales 2014). In this case, we have the example of the uptake and metabolism of Tween 20, 40, and 80, resulting from the absence of knowledge of the exact and unexpected esterase reaction. Another case of false negatives is when the carbon source is a dipeptide or a larger peptidic chain. The specific lyase activity cannot be inferred from the genome annotation, even if similar enzymes are found in other organisms. These kinds of reactions require experimental validation. Other similar cases may occur in the model if we expand the phenotypic comparisons; some of them can be included in future versions.

This work suggests *P. chlororaphis* PC9446 as a model for different approaches. Therefore, we experimentally tested its growth in three commonly used carbon sources. Glc is one of the most used carbon sources for production processes at the laboratory level, so it is of interest to understand PC9446 performance in batch culture and the ability of the model to reproduce this performance. The in silico maximum μ in Glc was 0.49 h⁻¹; thus, the model was constrained with this experimental Glc uptake rate, a close value to the in vivo μ = 0.44 h⁻¹ result. A similar result was obtained by constraining the model using experimental uptake rates for Gly and Scr. These results point to a well-constructed metabolic network

and a correct biomass similitude assumption between *P. chlororaphis* PC9446 and *P. putida* 2440. This kind of modeling (FBA) assumes a steady-state condition, which is not represented by batch culture. For this reason, we performed three chemostat cultures with Glc at different dilution rates ($D = 0.12$, 0.25, and 0.37 h^{-1}) and compared them with the simulations using FBA. Again, the model showed good performance, giving close in silico values with respect to the experimental data. The denitrification capability of PC9446 is important for future applications, such as aquatic bioremediation; therefore, simulations under denitrifying growth are of wide interest. Simulation was performed under a natural wastewater treatment environment. The denitrification network in this model differs from other models in the compartmentalization of this process that includes the cytosol and periplasm, which implies differences in the energy requirements for compound translocation (Oberhardt et al. 2008).

Among the most relevant compounds produced by PC9446 are phenazines, whose derivative compounds may vary from one strain to another. The specific phenazine operon, formed by the *phzABCDEFGH* gene in strain PC9446, can be found in other well-studied *Pseudomonas*, such as *P. aeruginosa*, and other closely related *P. chlororaphis* in which the *phzH* gene is responsible for the capability to produce PCN. Many efforts have been made to increase the production of this compound with experimental approximations (Peng et al. 2018b; Peng et al. 2018a; Yao et al. 2018). With the same purpose, we calculated the maximum PCN theoretical yields using different metabolizable carbon sources with the iMA1267 model (Fig. 3). The yield achieved with glucose was 0.39 gPCN/gGlc; however, better results were obtained with other carbon sources, such as hexanoate, leucyl-leucine heptanoate, and other fatty acids and peptides (up to 0.64 gPCN/g with hexanoate). Degradation of those compounds produces a higher ATP yield that allows better PCN production in silico. As mentioned above, the best theoretical carbon sources may not have the highest uptake rates, which may result in lower productivity, and these compounds may not be as available as other more frequently used compounds. Gly is among the best-rated carbon sources for PCN production (0.48 gPCN/gGlyc), and it is also an inexpensive substrate. Previous studies that used Gly to obtain different phenazines support this result (Peng et al. 2018b; Yao et al. 2018).

The optimization suggests an increase in the flux through PPP in both TKT reactions, TALA, RPI, and RPE accompanied by a lower flux in the reactions PPM, R15BPK, and RPK. This combination promotes a higher availability of erythrose-4-phosphate (E4P), which participates with phosphoenolpyruvate (PEP) in the first reaction directing the carbon flux to the shikimic acid pathway. In this optimization, the flux directed to the lower glycolytic pathway is increased, and that directed toward the TCA is decreased, which directly results in more PEP available to produce aromatic compounds. A similar strategy has been used to produce 2-hydroxyphenazine, where *tktA*, *ppsA*, and the inactivation of *pykF* improved the final concentration by 150% (Liu et al. 2016b). FBA does not allow metabolite accumulation to simulate a steady state; for this reason, a set of subsequent reactions without ramifications is bound to carry the same flux. As an example of this idea, the optimization suggests that all the reactions in the shikimic acid pathway (from 3-deoxy-D-arabino-heptulosonate 7-phosphate synthetase to chorismate synthase) need to be enforced, and the same happens with the phenazine pathway. The phenazine pathway also has ramifications, but in this case, we are avoiding byproducts by setting PCN as the only desired compound.

The optimization of PCN production highlights a problem that may not be considered in most of the reported overproducing strains. The PCN-producing strains do not consider the availability of glutamine (the donor of two amino groups in the phenazine pathway). The simulation with both carbon sources points to an increase of 3.5-fold in the GLNS reaction to reach the maximum PCN yield. Additionally, the precursors of such amino acids need to be increased, which is the reason for the enforced reactions of AKGDbr and AKGDBa. This result needs to be tested due to the lack of information on the glutamine concentration under reported culture conditions, but it may become more relevant when the use of the mineral medium for PCN production is intended. Another aspect that is not considered and more interesting is the production of H$_2$O$_2$, which is not only related to PCN production but also related to cellular integrity. The phenazine pathway is highly oxygen demanding, and many steps result in H$_2$O$_2$ production and are unavoidable. As Fig. 6 shows, the production of this toxic compound is increasing when the PCN yield is augmented, and the model addresses this problem by enforcing CAT. Under in vivo conditions, H$_2$O$_2$ accumulation has many metabolic implications and forces the cell to activate its mechanism against oxidative stress, including the peroxiredoxin system, glutathione peroxidase (GTHPi), catalase, and some molecules, such as H$_2$O$_2$ scavengers. The redox balance is also affected because NADH is considered a promotor of the production of oxidative molecules, and TCA is limited to adjusting NADH production and slowing the electron transport chain. Metabolic simulation works under optimality and only uses CAT to fight stress because it is the most economical mechanism. The simulation avoids the use of GTHPi because it takes more energy (2 ATP) to build glutathione and competes for the glutamate pull. In Fig. 5, it can be observed that TCA needs to be enforced, an opposite idea of the in vivo response. The in vivo behavior uses important molecules from central metabolism against H$_2$O$_2$, diminishing the availability of α-ketoglutarate and pyruvate for other purposes, but this mechanism is not needed in silico (Mailloux et al. 2011). Experimental measurements in *P. graminis* exposed to H$_2$O$_2$

showed this exact behavior, including lower NADH, a decrease in ATP concentrations, negative impacts on TCA, and a higher $NADP^+$/NADPH ratio, because NADPH is used in many antioxidant mechanisms (Wirgot et al. 2019).

In conclusion, metabolic optimization uses cheaper routes than those of a real cell against oxidative stress; for this reason, overexpressing the CAT gene may bring closer the in silico and in vivo phenotypes. Using CAT to avoid oxidative stress, the cell could maintain a better redox balance because NAPH will not be used to reduce glutathione to maintain GTHPi activity and will not need extra glutamate to synthesize glutathione. Supporting this idea of a recent work demonstrated that the production of phenazines induces the reactive oxygen species (ROS) production, which can be alleviated by the addition of dithiothreitol (DTT) and increase the production of PCA (phenazine-1-carboxylate) in an 80% (Yue et al. 2020).

Alternatively, strategies to increase the production of NADPH may help protect against H_2O_2. By diminishing the oxidative stress, cells may be able to perform closer to the optimal phenotype predicted with the model, getting the PCN yield closer to the theoretical maximum. The strategy applied in this work may help to speed up the understanding of PC9446 and the cellular context under different conditions and may be extrapolated to increase the production of other compounds, or to study the community interactions of this organism in its natural habitat in association with plant roots. The model constructed here can be easily used as a new tool for the exploration of the phenazine pathway and PC9446 metabolism overall. The iMA1267 model is the first model for *P. chlororaphis*, and it can be improved over time as more information about this organism is released; moreover, this model can be used as a backbone for production and utilization of other closed strains.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

Ethical statement This article does not contain any studies with human participants or animals performed by any of the authors.

References

Alva A, Sabido-Ramos A, Escalante A, Bolívar F (2020) New insights into transport capability of sugars and its impact on growth from novel mutants of *Escherichia coli*. Appl Microbiol Biotechnol 104: 1463–1479. https://doi.org/10.1007/s00253-019-10335-x

Bilal M, Guo S, Iqbal HMN, Hu H, Wang W, Zhang X (2017) Engineering *Pseudomonas* for phenazine biosynthesis, regulation, and biotechnological applications: a review. World J Microbiol Biotechnol 33:1–11. https://doi.org/10.1007/s11274-017-2356-9

Borgos SEF, Bordel S, Sletta H, Ertesvåg H, Jakobsen Ø, Bruheim P, Ellingsen TE, Nielsen J, Valla S (2013) Mapping global effects of the anti-sigma factor MucA in *Pseudomonas fluorescens* SBW25 through genome-scale metabolic modeling. BMC Syst Biol 7:19. https://doi.org/10.1186/1752-0509-7-19

Burr SE, Gobeli S, Kuhnert P, Goldschmidt-Clermont E, Frey J (2010) *Pseudomonas chlororaphis* subsp. *piscium* subsp. nov., isolated from freshwater fish. Int J Syst Evol Microbiol 60:2753–2757. https://doi.org/10.1099/ijs.0.011692-0

Calderón CE, Ramos C, de Vicente A, Cazorla FM (2015) Comparative genomic analysis of *Pseudomonas chlororaphis* PCL1606 reveals new insight into antifungal compounds involved in biocontrol. Mol Plant-Microbe Interact 28:249–260. https://doi.org/10.1094/mpmi-10-14-0326-fi

Cardoso JGR, Jensen K, Lieven C, Hansen ASL, Galkina S, Beber M, Özdemir E, Herrgård MJ, Redestig H, Sonnenschein N (2018) Cameo: a Python library for computer aided metabolic engineering and optimization of cell factories. ACS Synth Biol 7:1163–1166. https://doi.org/10.1021/acssynbio.7b00423

Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Ong Q, Paley S, Subhraveti P, Weaver DS, Karp PD (2016) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res 44:D471–D480. https://doi.org/10.1093/nar/gkv1164

Chin-A-Woeng TFC, Bloemberg GV, Mulders IHM, Dekkers LC, Lugtenberg BJJ (2007) Root colonization by phenazine-1-carboxamide-producing bacterium *Pseudomonas chlororaphis* PCL1391 is essential for biocontrol of tomato foot and root rot. Mol Plant-Microbe Interact 13:1340–1345. https://doi.org/10.1094/mpmi.2000.13.12.1340

Choi HS, Lee SY, Kim TY, Woo HM (2010) In silico identification of gene amplification targets for improvement of lycopene production. Appl Environ Microbiol 76:3097–3105. https://doi.org/10.1128/AEM.00115-10

Deng P, Wang X, Baird SM, Lu S-E (2015) Complete genome of *Pseudomonas chlororaphis* strain UFB2, a soil bacterium with antibacterial activity against bacterial canker pathogen of tomato. Stand Genomic Sci 10:117. https://doi.org/10.1186/s40793-015-0106-x

Dimkpa CO, Zeng J, McLean JE, Britt DW, Zhan J, Anderson AJ (2012) Production of indole-3-acetic acid via the indole-3-acetamide pathway in the plant-beneficial bacterium *Pseudomonas chlororaphis* O6 is inhibited by ZnO nanoparticles but enhanced by CuO nanoparticles. Appl Environ Microbiol 78:1404–1410. https://doi.org/10.1128/AEM.07424-11

Ebrahim A, Lerman JA, Palsson BO, Hyduke DR (2013) COBRApy: constraints-based reconstruction and analysis for Python. BMC Syst Biol 7. https://doi.org/10.1186/1752-0509-7-74

Girard G, Rigali S (2011) Role of the phenazine-inducing protein Pip in stress resistance of *Pseudomonas chlororaphis*. Microbiology 157: 398–407. https://doi.org/10.1099/mic.0.043075-0

Greenberg EP, Becker GE (1977) Nitrous oxide as end product of denitrification by strains of fluorescent pseudomonads. Can J Microbiol 23:903–907. https://doi.org/10.1139/m77-133

Gross H, Loper JE (2009) Genomics of secondary metabolite production by *Pseudomonas* spp. Nat Prod Rep 26:1408–1446. https://doi.org/10.1039/b817075b

Gunther NW, Nuñez A, Fett W, Solaiman DKY (2005) Production of rhamnolipids by *Pseudomonas chlororaphis*, a nonpathogenic bacterium. Appl Environ Microbiol 71:2288–2293. https://doi.org/10.1128/AEM.71.5.2288-2293.2005

Gutierrez M, Choi MH, Tian B, Xu J, Rho JK, Kim MO, Cho Y-H, Yoon SC (2013) Simultaneous inhibition of rhamnolipid and polyhydroxyalkanoic acid synthesis and biofilm formation in *Pseudomonas aeruginosa* by 2-bromoalkanoic acids: effect of inhibitor alkyl-chain-length. PLoS One 8:e73986. https://doi.org/10.1371/journal.pone.0073986

Guzmán GI, Utrilla J, Nurk S, Brunk E, Monk JM, Ebrahim A, Palsson BO, Feist AM (2015) Model-driven discovery of underground metabolic functions in *Escherichia coli*. Proc Natl Acad Sci 112:929–934. https://doi.org/10.1073/pnas.1414218112

Hu H, Li Y, Liu K, Zhao J, Wang W, Zhang X (2017) Production of trans-2,3-dihydro-3-hydroxyanthranilic acid by engineered *Pseudomonas chlororaphis* GP72. Appl Microbiol Biotechnol 101:6607–6613. https://doi.org/10.1007/s00253-017-8408-0

Huang R, Feng Z, Chi X, Sun X, Lu Y, Zhang B, Lu R, Luo W, Wang Y, Miao J, Ge Y (2018) Pyrrolnitrin is more essential than phenazines for *Pseudomonas chlororaphis* G05 in its suppression of Fusarium graminearum. Microbiol Res 215:55–64. https://doi.org/10.1016/j.micres.2018.06.008

Hung CS, Zingarelli S, Nadeau LJ, Biffinger JC, Drake CA, Crouch AL, Barlow DE, Russell JN, Crookes-Goodson WJ (2016) Carbon catabolite repression and Impranil polyurethane degradation in *Pseudomonas protegens* strain Pf-5. Appl Environ Microbiol 82:6080–6090. https://doi.org/10.1128/AEM.01448-16

Inoue H, Takimura O, Fuse H, Murakami K, Kamimura K, Yamaoka Y (2000) Degradation of triphenyltin by a fluorescent pseudomonad. Appl Environ Microbiol 66:3492–3498. https://doi.org/10.1128/AEM.66.8.3492-3498.2000

Inoue H, Takimura O, Kawaguchi K, Nitoda T, Fuse H, Murakami K, Yamaoka Y (2003) Tin-carbon cleavage of organotin compounds by pyoverdine from *Pseudomonas chlororaphis*. Appl Environ Microbiol 69:878–883. https://doi.org/10.1128/AEM.69.2.878-883.2003

King ZA, Lu J, Dräger A, Miller P, Federowicz S, Lerman JA, Ebrahim A, Palsson BO, Lewis NE (2016) BiGG models: a platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Res 44:D515–D522. https://doi.org/10.1093/nar/gkv1049

Koehorst JJ, Van Dam JCJ, Van Heck RGA, Saccenti E, Dos Santos VAPM, Suarez-Diez M, Schaap PJ (2016) Comparison of 432 *Pseudomonas* strains through integration of genomic, functional, metabolic and expression data. Sci Rep 6:1–13. https://doi.org/10.1038/srep38699

Koike I, Hattori A (1975) Energy yield of denitrification: an estimate from growth yield in continuous cultures of *Pseudomonas denitrificans* under nitrate-, nitrite- and nitrous oxide-limited conditions. J Gen Microbiol 88:11–19. https://doi.org/10.1099/00221287-88-1-11

Liao YC, Chen JCY, Tsai MH, Tang YH, Chen FC, Hsiung CA (2011) MrBac: a web server for draft metabolic network reconstructions for bacteria. Bioeng Bugs 2:284–287. https://doi.org/10.4161/bbug.2.5.16113

Liu Y, Lu S-E, Baird SM, Qiao J, Du Y (2014) Draft genome sequence of *Pseudomonas chlororaphis* YL-1, a biocontrol strain suppressing plant microbial pathogens. Genome Announc 2. https://doi.org/10.1128/genomeA.01225-13

Liu K, Hu H, Wang W, Zhang X (2016a) Genetic engineering of *Pseudomonas chlororaphis* GP72 for the enhanced production of 2-hydroxyphenazine. Microb Cell Factories 15:1–12. https://doi.org/10.1186/s12934-016-0529-0

Liu K, Hu H, Wang W, Zhang X (2016b) Genetic engineering of *Pseudomonas chlororaphis* GP72 for the enhanced production of 2-hydroxyphenazine. Microb Cell Factories 15:131. https://doi.org/10.1186/s12934-016-0529-0

Loeschcke A, Thies S (2015) *Pseudomonas putida*—a versatile host for the production of natural products. Appl Microbiol Biotechnol 99:6197–6214. https://doi.org/10.1007/s00253-015-6745-4

Loewen PC, Villenueva J, Fernando WGD, de Kievit T (2014) Genome sequence of *Pseudomonas chlororaphis* strain PA23. Genome Announc 2:1002784. https://doi.org/10.1128/genomea.00689-14

Löwe H, Schmauder L, Hobmeier K, Kremling A, Pflüger-Grau K (2017) Metabolic engineering to expand the substrate spectrum of *Pseudomonas putida* toward sucrose. Microbiologyopen 6:1–9. https://doi.org/10.1002/mbo3.473

Löwe H, Sinner P, Kremling A, Pflüger-Grau K (2018) Engineering sucrose metabolism in *Pseudomonas putida* highlights the importance of porins. Microb Biotechnol 13:1–10. https://doi.org/10.1111/1751-7915.13283

Mailloux RJ, Lemire J, Appanna VD (2011) Metabolic networks to combat oxidative stress in *Pseudomonas fluorescens*. Antonie Van Leeuwenhoek 99:433–442. https://doi.org/10.1007/s10482-010-9538-x

Martínez JA, Bolívar F, Escalante A (2015) Shikimic acid production in *Escherichia coli*: from classical metabolic engineering strategies to omics applied to improve its production. Front Bioeng Biotechnol 3:1–16. https://doi.org/10.3389/fbioe.2015.00145

Martínez-García E, Nikel PI, Aparicio T, de Lorenzo V (2014) *Pseudomonas* 2.0: genetic upgrading of *P. putida* KT2440 as an enhanced host for heterologous gene expression. Microb Cell Factories 13:159. https://doi.org/10.1186/s12934-014-0159-3

McCully LM, Bitzer AS, Spence CA, Bais HP, Silby MW (2014) Draft genome sequence of rice isolate *Pseudomonas chlororaphis* EA105. Genome Announc 2. https://doi.org/10.1128/genomeA.01342-14

Medema MH, Blin K, Cimermancic P, De Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R (2011) AntiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res 39:339–346. https://doi.org/10.1093/nar/gkr466

Moehario LH, Boestami HP, Edbert D, Tjoa E (2019) Automation for the identification of *Pseudomonas aeruginosa*: comparison of TDR-300B, VITEK®2, and VITEK®-MS. bioRxiv:510107. https://doi.org/10.1101/510107

Moreno-Avitia F, Lozano L, Utrilla J, Bolívar F, Escalante A (2017) Draft genome sequence of *Pseudomonas chlororaphis* ATCC 9446, a nonpathogenic bacterium with bioremediation and industrial potential. Genome Announc 5:1–2. https://doi.org/10.1128/genomeA.00474-17

Nikel PI, Martínez-García E, de Lorenzo V (2014) Biotechnological domestication of pseudomonads using synthetic biology. Nat Rev Microbiol 12:368–379. https://doi.org/10.1038/nrmicro3253

Nogales J (2014) A practical protocol for genome-scale metabolic reconstructions. pp 197–221

Nogales J, Palsson BØ, Thiele I (2008) A genome-scale metabolic reconstruction of *Pseudomonas putida* KT2440: iJN746 as a cell factory. BMC Syst Biol 2:79. https://doi.org/10.1186/1752-0509-2-79

Nogales J, Gudmundsson S, Duque E, Ramos L, Bernhard O (2017) Expanding the computable reactome in *Pseudomonas putida* reveals metabolic cycles providing robustness. https://doi.org/10.1101/139121

Nogales J, Mueller J, Gudmundsson S, Canalejo FJ, Duque E, Monk J, Feist AM, Ramos JL, Niu W, Palsson BO (2020) High-quality

genome-scale metabolic modelling of *Pseudomonas putida* highlights its broad metabolic capabilities. Environ Microbiol 22:255–269. https://doi.org/10.1111/1462-2920.14843

Norsigian CJ, Pusarla N, McConn JL, Yurkovich JT, Dräger A, Palsson BO, King Z (2019) BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. Nucleic Acids Res:1–5. https://doi.org/10.1093/nar/gkz1054

Oberhardt MA, Puchałka J, Fryer KE, Martins Dos Santos VAP, Papin JA (2008) Genome-scale metabolic network analysis of the opportunistic pathogen *Pseudomonas aeruginosa* PAO1. J Bacteriol 190:2790–2803. https://doi.org/10.1128/JB.01583-07

Oberhardt MA, Puchałka J, dos Santos VAPM, Papin JA (2011) Reconciliation of genome-scale metabolic reconstructions for comparative systems analysis. PLoS Comput Biol 7:e1001116. https://doi.org/10.1371/journal.pcbi.1001116

Ogata H, Goto S, Fujibuchi W, Kanehisa M (1998) Computation with the KEGG pathway database. BioSystems 47:119–128. https://doi.org/10.1016/S0303-2647(98)00017-3

Peng H, Tan J, Bilal M, Wang W, Hu H, Zhang X (2018a) Enhanced biosynthesis of phenazine-1-carboxamide by *Pseudomonas chlororaphis* strains using statistical experimental designs. World J Microbiol Biotechnol 34:0. https://doi.org/10.1007/s11274-018-2501-0

Peng H, Zhang P, Bilal M, Wang W, Hu H, Zhang X (2018b) Enhanced biosynthesis of phenazine-1-carboxamide by engineered *Pseudomonas chlororaphis* HT66. Microb Cell Factories 17:1–12. https://doi.org/10.1186/s12934-018-0962-3

Pierson LS, Pierson EA (2010) Metabolism and function of phenazines in bacteria: impacts on the behavior of bacteria in the environment and biotechnological processes. Appl Microbiol Biotechnol 86:1659–1670. https://doi.org/10.1007/s00253-010-2509-3

Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S, Kang J, Hyduke DR, Palsson B (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc 6:1290–1307. https://doi.org/10.1038/nprot.2011.308

Sharma PK, Munir RI, de Kievit T, Levin DB (2017) Synthesis of polyhydroxyalkanoates (PHAs) from vegetable oils and free fatty acids by wild-type and mutant strains of *Pseudomonas chlororaphis*. Can J Microbiol 63:1009–1024. https://doi.org/10.1139/cjm-2017-0412

Shen X, Chen M, Hu H, Wang W, Peng H, Xu P, Zhang X (2012) Genome sequence of *Pseudomonas chlororaphis* GP72, a root-colonizing biocontrol strain. J Bacteriol 194:1269–1270. https://doi.org/10.1128/JB.06713-11

Steen JA, Bohlke N, Vickers CE, Nielsen LK (2014) The trehalose phosphotransferase system (PTS) in *E. coli* W can transport low levels of sucrose that are sufficient to facilitate induction of the csc sucrose catabolism operon. PLoS One 9. https://doi.org/10.1371/journal.pone.0088688

Sutka RL, Ostrom NE, Ostrom PH, Breznak JA, Gandhi H, Pitt AJ, Li F (2006) Distinguishing nitrous oxide production from nitrification and denitrification on the basis of isotopomer abundances. Appl Environ Microbiol 72:638–644. https://doi.org/10.1128/AEM.72.1.638-644.2006

Takeuchi K, Noda N, Someya N (2014) Complete genome sequence of the biocontrol strain *Pseudomonas protegens* Cab57 discovered in Japan reveals strain-specific diversity of this species. PLoS One 9. https://doi.org/10.1371/journal.pone.0093683

Thiele I, Palsson BØ (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc 5:93–121. https://doi.org/10.1038/nprot.2009.203

Town J, Audy P, Boyetchko SM, Dumonceaux TJ (2016) Genome sequence of *Pseudomonas chlororaphis* strain 189. Genome Announc 4. https://doi.org/10.1128/genomeA.00581-16

Wang Y, Coleman-Derr D, Chen G, Gu YQ (2015) OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res 43:W78–W84. https://doi.org/10.1093/nar/gkv487

Wang D, Yu JM, Dorosky RJ, Pierson LS, Pierson EA (2016) The phenazine 2-hydroxy-phenazine-1- carboxylic acid promotes extracellular DNA release and has broad transcriptomic consequences in *Pseudomonas chlororaphis* 30-84. PLoS One 11:1–22. https://doi.org/10.1371/journal.pone.0148003

Wang S, Bilal M, Zong Y, Hu H, Wang W, Zhang X (2018) Development of a plasmid-free biosynthetic pathway for enhanced muconic acid production in *Pseudomonas chlororaphis* HT66. ACS Synth Biol 7:1131–1142. https://doi.org/10.1021/acssynbio.8b00047

Winsor GL, Griffiths EJ, Lo R, Dhillon BK, Shay JA, Brinkman FSL (2016) Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. Nucleic Acids Res 44:D646–D653. https://doi.org/10.1093/nar/gkv1227

Wirgot N, Lagrée M, Traïkia M, Besaury L, Amato P, Canet I, Sancelme M, Jousse C, Diémé B, Lyan B, Delort AM (2019) Metabolic modulations of *Pseudomonas graminis* in response to H2O2 in cloud water. Sci Rep 9:1–14. https://doi.org/10.1038/s41598-019-49319-2

Wullimann MF, Becker CG, Bernhardt RR, Schachner M, Dutta A, Ghosh S, Lee YS, Lin VW, Silver J, Doller CM, Malouf AT, Silver J, Tretter Y, Alzheimer C, Werner S, Schluesener HJ, Adibzahdeh M, Schwab JM, Duncan MR, Schwartz M, Quaglia X, Zurn AD, Barbarese E, Mcintosh TK (2016) A selective insecticidal protein from *Pseudomonas* for controlling corn rootworms. Science (80-) 6056:1–8. https://doi.org/10.1126/science.aaf6056

Xie K, Peng H, Hu H, Wang W, Zhang X (2013) OxyR, an important oxidative stress regulator to phenazines production and hydrogen peroxide resistance in *Pseudomonas chlororaphis* GP72. Microbiol Res 168:646–653. https://doi.org/10.1016/j.micres.2013.05.001

Yao R, Pan K, Peng H, Feng L, Hu H, Zhang X (2018) Engineering and systems-level analysis of *Pseudomonas chlororaphis* for production of phenazine-1-carboxamide using glycerol as the cost-effective carbon source. Biotechnol Biofuels 11:1–15. https://doi.org/10.1186/s13068-018-1123-y

Yu JM, Wang D, Pierson LS, Pierson EA (2018) Effect of producing different phenazines on bacterial fitness and biological control in *Pseudomonas chlororaphis* 30-84. Plant Pathol J 34:44–58. https://doi.org/10.5423/PPJ.FT.12.2017.0277

Yue SJ, Huang P, Li S, Jan M, Hu HB, Wang W, Zhang XH (2020) Enhanced production of 2-hydroxyphenazine from glycerol by a two-stage fermentation strategy in *Pseudomonas chlororaphis* GP72AN. J Agric Food Chem 68:561–566. https://doi.org/10.1021/acs.jafc.9b05033