



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

Modelación matemática y análisis de superficies de respuesta  
para una cepa de *Escherichia coli* modificada por ingeniería  
de vías metabólicas para la producción de ácido shikímico

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Doctor en Ciencias

PRESENTA:

M.C. Juan Andrés Martínez Álvarez

TUTOR PRINCIPAL:

Dr. Francisco Gonzalo Bolívar Zapata

[Instituto de Biotecnología-UNAM](#)

MIEMBROS DEL COMITÉ TUTOR:

Dr. Octavio T. Ramírez Reivich

[Instituto de Biotecnología-UNAM](#)

Dr. Alvaro R. Lara Rodríguez

[Departamento de procesos y tecnología-UAM](#)

Dra. Noemí Flores Mejía

[Instituto de Biotecnología-UNAM](#)

Ciudad de México, marzo 2019



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

AMDG...

A mi Familia, Polioptro, Gloria, Daniel y Gaby...

A Pamela ...

y a mis amigos y compañeros...

*La paciencia  
Todo lo alcanza  
Quien a Dios tiene  
Nada le falta  
Sólo Dios basta.*

Este trabajo fue realizado en las instalaciones del Instituto de Biotecnología - UNAM en Cuernavaca, Morelos, México. Este trabajo fue realizado con el apoyo del Consejo Nacional de Ciencia y Tecnología (CONACYT, México) beca de Ciencia Básica 240519 y el apoyo de la beca DGAPA-PAPIIT UNAM, México IN209618. El presente trabajo ha sido publicado y presentado en diferentes instancias gracias al “Programa de Apoyo a los Estudios del Posgrado” (PAEP).

## Integrantes del comité tutor

**Dr. Francisco Gonzalo Bolívar Zapata** . (Tutor principal)

Departamento de ingeniería celular y biocatálisis, Instituto de Biotecnología, UNAM. Miembro de El Colegio Nacional (ECN), México.

**Dr. Octavio Tonatiuh Ramírez Reivich** .

Departamento de medicina molecular y bioprocesos, Instituto de Biotecnología, UNAM.

**Dr. Alvaro Raúl Lara Rodríguez** .

Departamento de procesos y tecnología, Universidad Autónoma Metropolitana, Unidad Cuajimalpa, UAM .

**Dra. Noemí Flores Mejía** .

Departamento de ingeniería celular y biocatálisis, Instituto de Biotecnología, UNAM.

## Integrantes del jurado de examen de grado

**Dr. Lorenzo Patrick Segovia Forcella** .

Departamento de microbiología molecular, Instituto de Biotecnología, UNAM.

**Dr. Adán Oswaldo Guerrero Cárdenas** .

Laboratorio nacional de microscopía avanzada, Instituto de Biotecnología, UNAM.

**Dr. José Utrilla Carreri** .

Laboratorio de biología de sistemas y biología sintética, Centro de Ciencias Genómicas, UNAM.

**Dr. Maximino Aldana González** .

Departamento de fenómenos no lineales y complejidad, Instituto de Ciencias Físicas, UNAM.

**Dr. Juan Carlos Sigala Alanís** .

Departamento de procesos y tecnología, Universidad Autónoma Metropolitana, Unidad Cuajimalpa, UAM .

## Agradecimientos

Se agradece a los investigadores Dr. Francisco Gonzalo Bolívar, Dr. Guillermo Gosset y Dr. Alfredo Martínez por el espacio y los recursos utilizados en el Laboratorio de ingeniería de vías metabólicas del consorcio Bolívar-Gosset-Martínez del departamento de Ingeniería Celular y Biocatálisis del Instituto de Biotecnología UNAM, para la realización de esta investigación.

Se agradece también al personal técnico, M.C. Georgina Hernández, M.C. Ramón de Anda, M.C. Luz Ma. Martínez y en especial a la Dra. Noemí Flores por su apoyo y conocimientos aportados durante la realización del proyecto. Se agradece así mismo, a los estudiantes de este laboratorio, en especial al Dr. Alberto Rodríguez y al M.C. Fabián Moreno por su discusión y apoyo a lo largo del desarrollo de la investigación. Se agradece además al personal de apoyo y administrativo, que permitieron el correcto desarrollo del proyecto, Mercedes, Manuel, Delia, Sonia, Rubí y Aurelia.

Se agradece de forma especial al Dr. Francisco Gonzalo Bolívar, al Dr. Octavio T. Ramírez y al Dr. Alvaro R. Lara por su trabajo, apoyo y valiosa dirección del proyecto a través de las diferentes etapas del mismo. Especialmente, por su participación crítica y aportaciones intelectuales, que permitieron la concepción, desarrollo y realización de este trabajo y de todos los relacionado con el desarrollo profesional y personal del M.C. Juan Andrés Martínez durante esta etapa. Se agradece particularmente al Dr. Guillermo Gosset y al Dr. Adelfo Escalante por su participación y apoyo en la revisión crítica de los artículos publicados durante la estancia doctoral, así como por su dirección personal y profesional en el laboratorio. Se agradece también de forma particular al Profesor Doraiswami Ramkrishna y al Dr. Huyn-Seob Song por el apoyo y conocimientos aportados en el desarrollo de modelos matemáticos.

Finalmente, el presente trabajo no hubiera sido posible sin los apoyos del Consejo Nacional de Ciencia y Tecnología (CONACYT, México) Ciencia Básica 240519 y el apoyo de la DGAPA-PAPIIT UNAM IN209618, México.

---

# Índice general

Agradecimientos . . . . .	IV
Resumen . . . . .	XI
Summary . . . . .	XIII
<b>1. INTRODUCCIÓN y ANTECEDENTES</b>	<b>1</b>
1.1. INTRODUCCIÓN . . . . .	2
1.1.1. Importancia del ácido shikímico . . . . .	3
1.1.2. Producción de ácido shikímico en la industria . . . . .	5
1.1.3. Ingeniería de vías metabólicas para elevar y optimizar la producción de ácido shikímico . . . . .	6
1.1.4. La modelación matemática como herramienta para la ingeniería de vías metabólicas . . . . .	13
1.2. ANTECEDENTES . . . . .	16
1.2.1. Cepas de <i>Escherichia coli</i> carentes del sistema PTS, incluyendo la AR36 . . . . .	16
1.2.2. Modelos metabólicos para la optimización de la producción de ácido shikímico . . . . .	20
1.2.3. Modelado cibernético . . . . .	21
<b>2. PLANTEAMIENTO DEL PROBLEMA DE INVESTIGACIÓN.</b>	<b>26</b>
2.1. JUSTIFICACIÓN e HIPÓTESIS . . . . .	27
2.1.1. Justificación . . . . .	27
2.1.2. Hipótesis . . . . .	27
2.2. OBJETIVOS . . . . .	28
2.2.1. Objetivo general . . . . .	28

---

2.2.2. Objetivos particulares . . . . .	28
<b>3. MATERIALES, MÉTODOS y MODELOS</b>	<b>29</b>
3.1. MATERIALES Y MÉTODOS . . . . .	30
3.1.1. Diseño experimental . . . . .	30
3.1.2. Cepas, cultivos y procedimientos analíticos . . . . .	31
3.2. MODELOS MATEMÁTICOS . . . . .	33
3.2.1. Diseño y construcción de modelos para describir el comporta- miento fisiológico en variaciones de las concentraciones de los dos sustratos utilizados. . . . .	33
3.2.2. Construcción y validación de superficies de respuesta . . . . .	37
3.2.3. Diseño y construcción de modelos para describir el comporta- miento de las distribuciones de flujos metabólicos en variacio- nes de la concentración de los dos sustratos utilizados. . . . .	38
3.3. PROCESO DE PRODUCCIÓN DE ÁCIDO SHIKÍMICO . . . . .	45
3.3.1. Diseño de un bioproceso mejorado para la producción de ácido shikímico . . . . .	45
<b>4. RESULTADOS y DISCUSIÓN</b>	<b>46</b>
4.1. RESULTADOS . . . . .	47
4.1.1. Caracterización del comportamiento fisiológico de la cepa AR36 mediante la construcción y análisis de superficies de respuesta de los parámetros modelados en diferentes etapas de la fermen- tación. . . . .	47
4.1.2. Validación de las superficies modeladas mediante la compara- ción de predicciones del comportamiento contra experimentos no contenidos en el diseño experimental. . . . .	52
4.1.3. Caracterización del comportamiento de las distribuciones de flujo mediante la construcción y análisis de superficies de res- puesta de los flujos modelados en diferentes etapas de la fer- mentación. . . . .	54
4.2. DISCUSIÓN . . . . .	70



---

4.2.1. Discusión, análisis y comparación de los resultados obtenidos de los modelos fisiológicos y metabólicos para la determinación de blancos genéticos y metabólicos para el mejoramiento de la producción de ácido shikímico. . . . .	70
4.3. APLICACIÓN: PRODUCCIÓN DE ÁCIDO SHIKIMICO . . . . .	81
4.3.1. Mejoramiento de la producción de ácido shikímico mediante el uso de la información obtenida por modelación matemática. .	81
<b>5. CONCLUSIONES, PERSPECTIVAS y PRODUCTOS</b>	<b>86</b>
5.1. CONCLUSIONES . . . . .	87
5.2. PERSPECTIVAS . . . . .	89
5.3. PRODUCTOS DEL TRABAJO . . . . .	90
5.3.1. Publicaciones científicas . . . . .	90
5.3.2. Trabajo de apoyo a docencia y desarrollo profesional de estudiantes. . . . .	92
5.3.3. Colaboraciones con revistas indexadas, industria y apoyo a actividades de investigación. . . . .	92
<b>Bibliografía</b>	<b>xv</b>
<b>ANEXOS</b>	<b>xxvii</b>

---

# Índice de figuras

1.1. Representación esquemática del sistema principal de transporte de glucosa, metabolismo central de carbono (CCM) y su interconexión con la ruta metabólica del SA y aminoácidos aromáticos en <i>Escherichia coli</i> . . . . .	8
1.2. Porcentajes molares de los compuestos aromáticos producidos en la cepa AR36 en cultivos lote. . . . .	19
3.1. Mapa simplificado de las reacciones del metabolismo central de carbono tomadas en cuenta para el modelo cibernético y el análisis de flujos. . . . .	40
4.1. Resultados para las nueve condiciones de fermentación realizadas con la cepa AR36, aproximación de los modelos fisiológicos a los datos experimentales. . . . .	50
4.2. Contornos de las superficies de respuesta para los parámetros estimados por los modelos. . . . .	51
4.3. Predicción fisiológica de los modelos a tres experimentos de validación y puntos experimentales. . . . .	53
4.4. Comparación de los datos experimentales con los perfiles de metabolitos externos resultantes de la modelación de las distribuciones de flujos para cada punto del diseño experimental. . . . .	55
4.5. Contornos de superficies de respuesta para la fase de crecimiento pt.1 (% Flujo relativo al consumo de GLC). . . . .	59
4.6. Contornos de superficies de respuesta para la fase de crecimiento pt.2 (% Flujo relativo al consumo de GLC). . . . .	62

---

4.7. Perfiles de fermentación de la cepa AR36 $\Delta$ <i>poxB</i> cultivada en condiciones iniciales 145:45 GLC:YE g/L . . . . .	63
4.8. Contornos de superficies de respuesta para la fase estacionaria pt.1 (% Flujo relativo al consumo de GLC). . . . .	66
4.9. Contornos de superficies de respuesta para la fase estacionaria pt.2 (% Flujo relativo al consumo de GLC). . . . .	68
4.10. Perfil de fermentación del lote alimentado en condiciones iniciales de 80 g/L GLC y 40 g/L YE. . . . .	83

---

# Índice de cuadros

1.1. Cepas productoras de ácido shikímico, producción y rendimiento. Modificada de Martínez et al. (2015). . . . .	12
3.1. Puntos experimentales para la matriz experimental diseñada para la caracterización fisiológica y metabólica de la AR36 en variaciones de sustrato. . . . .	30
3.2. Composición del medio de cultivo base empleado en los experimentos	32
3.3. Definición de las reacciones de la red metabólica utilizada para la modelación . . . . .	39
4.1. Parámetros promedio y desviaciones estándar para el punto central experimental, 100 g/L GLC y 30 g/L YE. . . . .	47
4.2. Valores estadísticos para la comparación de los perfiles experimentales con los resultados de los modelos . . . . .	48
4.3. Valores experimentales, modelados y error promedio para la validación de las superficies. . . . .	54

## Resumen

Los intermediarios de la vía de los aminoácidos aromáticos tienen papeles importantes en las industrias farmacéuticas y alimentarias. La producción biotecnológica de estos compuestos ha incrementado en interés en los últimos años gracias a la capacidad de modificación genética que ha permitido la construcción de cepas bacterianas modificadas por ingeniería genética capaces de producirlos con rendimientos relativamente elevados. Sin embargo, las estrategias clásicas de ingeniería de vías metabólicas a menudo inducen desequilibrios en el metabolismo central de carbono causando resultados no deseados para la conversión de sustratos a los productos requeridos. Por ello existe interés de estudiar de forma sistémica y global los procesos metabólicos. En este sentido, muchos modelos matemáticos han sido construidos para describir el estado metabólico de los microorganismos, la identificación de blancos genéticos y la optimización de parámetros de fermentación para mejorar la producción de compuestos de interés. Sin embargo, pocos han sido desarrollados para representar procesos en condiciones dinámicas y/o en medios complejos, condiciones comunes en diversos procesos de producción industrial. En este trabajo, se empleó una aproximación de modelación dinámica para describir el comportamiento fisiológico y metabólico en medio complejo de una cepa de *Escherichia coli* carente del sistema PTS (principal sistema de transporte y fosforilación de glucosa), la cual se modificó genéticamente para la sobreproducción de ácido shikímico. Se realizó el análisis de flujos de carbono y de parámetros mediante superficies de respuesta construidas sobre un diseño experimental de variaciones en la concentración inicial de sustratos. Los modelos fisiológicos fueron construidos con ecuaciones logísticas permitiendo describir el consumo y producción de metabolitos con errores menores al 5%. De los modelos se obtuvieron parámetros para la caracterización del comportamiento de la cepa en las diferentes condiciones iniciales de sustratos. Se construyó una red del metabolismo central de carbono de *Escherichia coli* la cual fue reducida a 9 modos elementales mediante el análisis de rendimientos obtenidos de las fermentaciones. Se construyeron ecuaciones diferenciales basadas en los modelos fisiológicos para determinar el comportamiento dinámico de los modos elementales seleccionados, los cuales se regularon mediante variables cibernéticas referentes a la asignación de recursos para la síntesis de enzimas y sus actividades. Esta aproximación cibernética permitió el

cálculo de distribuciones de flujo de carbono dinámicas las cuales revelaron diferentes estados metabólicos dependientes de las concentraciones iniciales de los substratos en el medio. De esta forma se encontró una alta proporción de flujo de carbono dirigido a través de las vías de las pentosas fosfato y de Entner-Doudoroff señalando a la eritrosa-4-fosfato como el compuesto limitante para la producción de ácido shikímico. Se observó también un incremento en la actividad del ciclo de producción-consumo de acetato entre la piruvato oxidasa y la acetil-coenzima A sintetasa para contender con la saturación del metabolismo central de carbono en las condiciones de cultivo. Además, el análisis de las superficies de respuesta construidas tanto para los flujos de carbono como para los parámetros fisiológicos permitió localizar zonas con diferentes comportamientos metabólicos y fisiológicos. Con estos análisis se diseñó un bioproceso para aumentar la productividad del shikimato. Este proceso fue diseñado para mantener constantes las concentraciones iniciales de los substratos al extender la fase de crecimiento preservando las condiciones metabólicas especificadas por las superficies de respuesta calculadas para la fase estacionaria. Dicho proceso resultó en 60 g/L de título final de ácido shikímico (40 % de aumento) con una productividad volumétrica de 2.45 g/L\*h (70 % de aumento), manteniendo los altos rendimientos obtenidos con esta cepa en investigaciones anteriores.

## Summary

Aromatic amino acid pathway intermediates play important roles on the pharmaceutical and alimentary industries. Their biotechnological production has increased on interest on recent years thanks to the construction of bacterial strains genetically engineered capable of producing them with relatively high yields. However, classic metabolic engineering strategies often induce significant flux imbalances to microbial metabolism, causing undesirable outcomes for valuable metabolite production. Therefore, there has been an increasing interest on metabolism global and systemic processes characterization. On that regard, many mathematical models have been constructed to describe the metabolic state of microorganisms, genetic modification target identification and process parameter optimization to enhance the production of valuable metabolites. Despite this, few models have been developed to represent or characterize processes under dynamic conditions and/or complex media, which are common on many industrial production processes. On this report, a dynamic modeling approach was used to describe the physiological and metabolic behavior of an *Escherichia coli* strain that lacks the PTS (main glucose transport system), modified for the overproduction of shikimic acid, under complex media conditions. Physiological and metabolic flux parameter behavioral characterization was performed by response surface analysis on a experimental design with variant substrate conditions. Physiological models were constructed with logistic math functions to describe growth, consumption and production of metabolites, resulting on less than 5% error compared to experimental data. This helped parameterize and characterize the output of the strain under different initial substrate conditions. For metabolic flux modeling, an *Escherichia coli* central carbon metabolism network was constructed and reduced to 9 elementary modes by yield space analysis. Dynamic behavior of these elementary modes were described by differential equations constructed based on the physiological models and regulated by cybernetic variables accounting for resource allocation for enzymatic synthesis and activity. This approach allowed the calculation of dynamic flux distributions which revealed different metabolic states dependent on the initial concentration of substrates on media composition. 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 observed physiological and flux with response surfaces, enabled zone allocation for different physiological outputs within variant substrate conditions. Information was then used for an improved fed-batch process designed to preserve the metabolic conditions for shikimic acid productivity enhancement. 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.



---

# Capítulo 1

## INTRODUCCIÓN y ANTECEDENTES

## 1.1. INTRODUCCIÓN

A principios de la última década la producción anual de aminoácidos se estimó en el orden de 4.5 millones de toneladas, dentro de los cuales los aminoácidos aromáticos triptófano (TRP) y fenilalanina (PHE) presentaron tamaños de mercado superiores a las 14,000 y 30,000 toneladas anuales respectivamente (Rodríguez et al., 2014). Esta demanda se debe a que estas moléculas son utilizadas en la producción de compuestos de alto valor comercial como L-DOPA, fenilalanina, aspartame, p-aminobenzoato, pirrolnitrina, ciclomarina, entre otros (Flores et al., 1996; Patnaik y Liao, 1994; Polen et al., 2005). Los aminoácidos aromáticos son producidos de forma natural en diversos organismos a través de la ruta metabólica de los aminoácidos aromáticos (AAAP), la cual se encuentra presente en bacterias, plantas, hongos y algunos parásitos del género apicomplexa (Díaz-Quiroz et al., 2014; Herrmann y Weaver, 1999). Además de los compuestos finales de esta ruta metabólica, los compuestos intermedios y derivados tienen también un papel importante en la industria farmacéutica y alimentaria, ya sea como materias primas, aditivos o productos terminados (Báez et al., 2001; Báez-Viveros et al., 2004; Bongaerts et al., 2001; Chávez et al., 2005; Gosset, 2009; Patnaik y Liao, 1994; Yi et al., 2003).

La AAAP comienza con la condensación aldólica entre el fosfoenolpiruvato (PEP) y la eritrosa-4-fosfato (E4P) provenientes del metabolismo central de carbono, específicamente de las rutas de Embden-Meyerhoff-Parnas (EMP) y de la ruta de las pentosas fosfato (PPP) respectivamente (Figura 1.1). Dicha reacción produce D-arabino-heptuloso-7-fosfato (DAHP), el primer intermediario de la vía. El DAHP se transforma posteriormente en corismato (CHO) por medio de reacciones secuenciales en lo que se conoce como ruta común de los aminoácidos aromáticos o vía del shikimato, para ramificarse en varias rutas individuales para la producción de los aminoácidos aromáticos y otros compuestos aromáticos (Díaz-Quiroz et al., 2014; Flores et al., 1996; Patnaik y Liao, 1994). La AAAP es esencial para el crecimiento de los organismos, para la producción de los aminoácidos aromáticos esenciales: PHE, TRP y tirosina (TYR) y para la producción de una gran diversidad de compuestos relevantes para diferentes procesos biológicos como cofactores, vitaminas, sideróforos entre otros (Díaz-Quiroz et al., 2014; Herrmann y Weaver, 1999).

Por estas razones, esta ruta metabólica se encuentra altamente regulada en los organismos para mantener un nivel homeostático y de producción adecuada para el crecimiento y mantenimiento celular (Gosset, 2005; Patnaik y Liao, 1994; Rodríguez et al., 2013). Debido a la importancia de la AAAP para el crecimiento de los organismos, así como la gran relevancia industrial de varios de sus compuestos intermediarios, diferentes grupos de investigación han desarrollado alternativas de modificación, control y optimización de esta ruta con fines de obtener fuentes de producción biotecnológica por diferentes organismos de forma confiable y con altos rendimientos.

### 1.1.1. Importancia del ácido shikímico

Uno de los intermediarios de la AAAP que más ha generado interés de investigación e industrial es el ácido shikímico (SA). Este es un compuesto conformado por un anillo de seis carbonos altamente funcionalizado, el cual contiene tres centros asimétricos en su estructura, por lo que es usado como materia prima precursora para la producción de compuestos aromáticos. Además, sus características estructurales han facilitado su uso en la industria como precursor enantiomérico para la producción de moléculas biológicas de alto valor agregado (Díaz-Quiroz et al., 2014; Estévez y Estévez, 2012). Hasta la fecha el SA ha sido utilizado como materia prima para la producción de compuestos con actividades antipiréticas, antioxidantes, anticoagulantes, antitrombóticas, antiinflamatorias, analgésicas, antibacterianas, hormonales y antivirales (Díaz-Quiroz et al., 2014; Estévez y Estévez, 2012). Dentro de los derivados más importantes provenientes del SA se encuentra la ciclohexano (-)zeylenona, la cual presenta actividad inhibitoria y citotóxica para células de carcinoma de Ehrlich y otras líneas tumorales humanas (Díaz-Quiroz et al., 2014; Estévez y Estévez, 2012; Liu et al., 2012). Otro compuesto derivado del SA interesante es el triacetil shikimato, el cual ha sido utilizado como agente antitrombótico y protector cerebral durante episodios de isquemia focal (Díaz-Quiroz et al., 2014; Huang et al., 2002). Además de estos, se han estudiado otros derivados para el tratamiento de colitis ulcerosa, entre otras enfermedades (Díaz-Quiroz et al., 2014; Estévez y Estévez, 2012; Ma et al., 2004; Tang et al., 2009).

Sin embargo, el uso más importante del SA en la industria en los años recientes es como precursor del Oseltamivir fosfato (Tamiflu®), un antiviral ampliamente usado en el tratamiento de la influenza A y B, influenza aviar por virus H5N1 y la influenza humana H1N1 (Chandran et al., 2003; Díaz-Quiroz et al., 2014; Escalante et al., 2010; Krämer et al., 2003). En el año 2009 Roche<sup>TM</sup> reportó ventas de Tamiflu® por 3.5 billones de dólares y la capacidad de producir hasta 33 millones de tratamientos al mes, es decir 400 millones de tratamientos al año (Roche, 2009). Sin embargo, esta capacidad de producción podría resultar insuficiente en caso de un brote de influenza pandémica, ya que el estimado para la producción necesaria es de 30 billones de dosis, requiriendo 3.9 millones de kilogramos de SA (Bradley, 2005; Rawat et al., 2013). Tan sólo en Europa, Japón y Estados Unidos alrededor de 100 millones de personas son afectadas anualmente por cepas comunes de influenza, mientras que en pandemias anteriores al año 2010, entre el 20 % y el 40 % de la población mundial fue infectada causando más de 20 millones de muertes (Roche, 2009). La limitación de la existencia de tratamientos y su obtención en casos de pandemia se exagera para países en vías de desarrollo como México. Según el reporte de la World Health Organization (WHO) sobre la preparación para brotes de influenza, en el 2011 se tenía cobertura sólo para 66 millones de personas con ingresos medios y bajos en países en desarrollo, lo que representa sólo el 2.25 % de su población (WHO, 2011). En este sentido, el inventario de este medicamento en almacenes de las instituciones de salud gubernamentales y privadas debe ser aumentado y renovado constantemente, y aún con estos a máxima capacidad, la disponibilidad podría ser insuficiente (Neumann et al., 2009; WHO, 2011). Las razones anteriores, aunadas a la capacidad de mutación, recombinación y rearreglo del virus que aumentan la probabilidad de generar nuevas cepas de influenza con capacidades pandémicas, supone serios retos a los organismos de salud pública y a las comunidades científicas para la búsqueda de alternativas de logística, obtención y producción de medicamentos para el combate de la influenza (Kao et al., 2010; Neumann et al., 2009; Raghavendra et al., 2009).

Por lo tanto, siendo el SA el precursor principal de este medicamento y de otros derivados importantes para la industria farmacéutica y de la salud, este se convierte en un blanco de estudio de gran interés científico e industrial. En el caso de la biotecnología, se busca el diseño de cepas con capacidades de producción mejoradas y procesos de fermentación optimizados capaces de competir con las formas de producción tradicionales (Johansson y Lidén, 2006; Nie et al., 2009; Raghavendra et al., 2009).

### 1.1.2. Producción de ácido shikímico en la industria

La mayoría de los compuestos pertenecientes a la AAAP incluyendo el SA son producidos a escala industrial por medio de síntesis química, desde materias primas no renovables como el petróleo, y/o son extraídos directamente de plantas (Báez et al., 2001; Gosset, 2009). Sin embargo, estas estrategias por lo general resultan costosas e ineficientes (Krämer et al., 2003). Para el caso específico del ácido shikímico, la mayor parte de la producción se obtiene de la semilla de anís chino *Illicium verum*. Dicha semilla contiene entre el 2 % y el 7 % de ácido shikímico (Raghavendra et al., 2009). Las técnicas de extracción, aunque son de bajo costo y fáciles de implementar, sólo permiten obtener rendimientos bajos, cercanos a los 30 mg/Kg (Li et al., 2007; Raghavendra et al., 2009; Rawat et al., 2013; Wang et al., 2011). Además, la obtención de las semillas ocurre hasta después de 6 años de crecimiento de la planta en su primera floración y de ahí en adelante estacionalmente en los meses de septiembre y octubre (Li et al., 2007; Raghavendra et al., 2009; Wang et al., 2011). A pesar de estas limitaciones, reportes del 2007 indican que aproximadamente el 90 % de la cosecha de semillas de anís en China fue utilizada por Roche para la manufactura de Tamiflú<sup>®</sup> (Chandran et al., 2003; Escalante et al., 2010; Krämer et al., 2003; Li et al., 2007), siendo esta la fuente primordial de obtención de SA por parte de esta compañía. Sin embargo, a través de la última década la proporción de uso de otras fuentes de obtención como la biotecnológica, ha ido incrementando gracias a la aparición de mejores cepas y procesos fermentativos (Chandran et al., 2003; Díaz-Quiroz et al., 2014; Martínez et al., 2015; Rodríguez et al., 2014).

Una de las alternativas más viables a los procesos extractivos clásicos, es la producción mediante microorganismos recombinantes con altos rendimientos y productividades. Por este motivo, muchos grupos de investigación se han enfocado en la producción de SA en diferentes microorganismos. Los avances en este campo van desde la búsqueda de cepas naturalmente productoras, la modificación genética y metabólica de las capacidades de acumulación de este compuesto, la optimización de procesos biotecnológicos en reactores de gran capacidad hasta el mejoramiento de los procesos de recuperación y purificación desde los caldos de cultivo (Díaz-Quiroz et al., 2014; Ghosh et al., 2012; Martínez et al., 2015). La conceptualización y construcción de cepas sobreproductoras es resultado de la ingeniería de vías metabólicas (IVM), la cual diseña e implementa modificaciones genéticas para alterar las propiedades celulares y aumentar así la capacidad de formación de productos de interés (Stephanopoulos y Sinskey, 1993). En este sentido, la bacteria *E. coli* ha sido la más extensivamente utilizada por varios grupos de investigación; cepas genéticamente modificadas de esta bacteria ya han incluso alcanzado niveles de uso industrial para la producción de SA (Chen et al., 2013; Frost et al., 2002; Li et al., 2007).

### 1.1.3. Ingeniería de vías metabólicas para elevar y optimizar la producción de ácido shikímico

El proceso de desarrollo de cepas sobreproductoras por IVM, comienza por la selección de organismos adecuados para la producción. Estos organismos deben poseer las características adecuadas para realizar diversas modificaciones a sus rutas metabólicas, eliminar los subproductos no deseados y desregular las reacciones deseadas para llegar al metabolito de interés (Patnaik y Liao, 1994). Una de las bacterias más utilizadas por la IVM en la producción biotecnológica de metabolitos, tanto autólogos como heterólogos, es *Escherichia coli*. Particularmente, esta bacteria ha sido estudiada de forma extensiva para la producción de SA debido al amplio conocimiento de sus características, genéticas, fisiológicas y metabólicas, su capacidad de crecimiento con bajos requerimientos nutricionales, a las variadas y poderosas metodologías que existen para su modificación genética y a su amplio y estable uso en la industria (Chen et al., 2013; Flores et al., 2004; Frost et al., 2002; Li et al., 2007; Yu et al., 2011).

Una vez seleccionado el organismo a modificar, se estudian sus particularidades genéticas y metabólicas con respecto a la producción del compuesto de interés. En este sentido, la producción del ácido shikímico en *Escherichia coli* comienza con la condensación de fosfoenolpiruvato (PEP) y eritrosa-4-fosfato E4P por medio de tres iso-enzimas DAHP sintetas (DAHPs) codificada por los genes *aroG*, *aroF* y *aroH* (Figura 1.1). Estas enzimas se encuentran altamente reguladas por la concentración de aminoácidos aromáticos en el citoplasma, por lo que son reprimidas en su actividad cuando estos últimos llegan a concentraciones suficientes para mantener el crecimiento (Gosset, 2005; Keseler et al., 2013; Martínez et al., 2015). Posteriormente el D-arabino-heptulosonato-7-fosfato (DAHP) es convertido en ácido dehidroquinico (DHQ) por la enzima 3-dehidroquinato sintasa codificada en el gen *aroB*, seguida por una deshidratación catalizada por la 3-dehidroquinato deshidratasa codificada por el gen *aroD* y generando ácido dehidroshikímico (DHS). Este ácido es posteriormente convertido en ácido shikímico por medio de la enzima ácido shikímico deshidrogenasa codificada por el gen *aroE*. Finalmente, el ácido shikímico es dirigido hacia la síntesis de CHO y aminoácidos aromáticos mediante las isoenzimas ácido shikímico cinasas codificadas en los genes *aroK* y *aroL*. El esquema general de la ruta metabólica del shikimato en *Escherichia coli* se puede observar en la Figura 1.1.

De acuerdo con el conocimiento que se tenga de la vía, se diseñan modificaciones al metabolismo a través de las estrategias clásicas de IVM con el objetivo de mejorar el comportamiento celular y los parámetros como consumo de substrato y de producción del metabolito, así como de tolerancia a compuestos tóxicos, altas temperaturas, fluctuaciones de pH entre otras, que otorguen un mejor desempeño para la producción. Entre las estrategias específicas de IVM que se han utilizado para la producción de SA en *Escherichia coli* se encuentran: i) el aumento de la producción de los compuestos precursores de la vía (Chandran et al., 2003; Chen et al., 2012; Cui et al., 2014; Escalante et al., 2010; Flores et al., 2005, 1996; Krämer et al., 2003; Lin et al., 2014; Rodriguez et al., 2013), ii) la eliminación de las rutas metabólicas de competencia (Chen et al., 2012; Escalante et al., 2010; Rodriguez et al., 2013), iii) la desregulación de la ruta metabólica deseada a nivel de actividad enzimática y

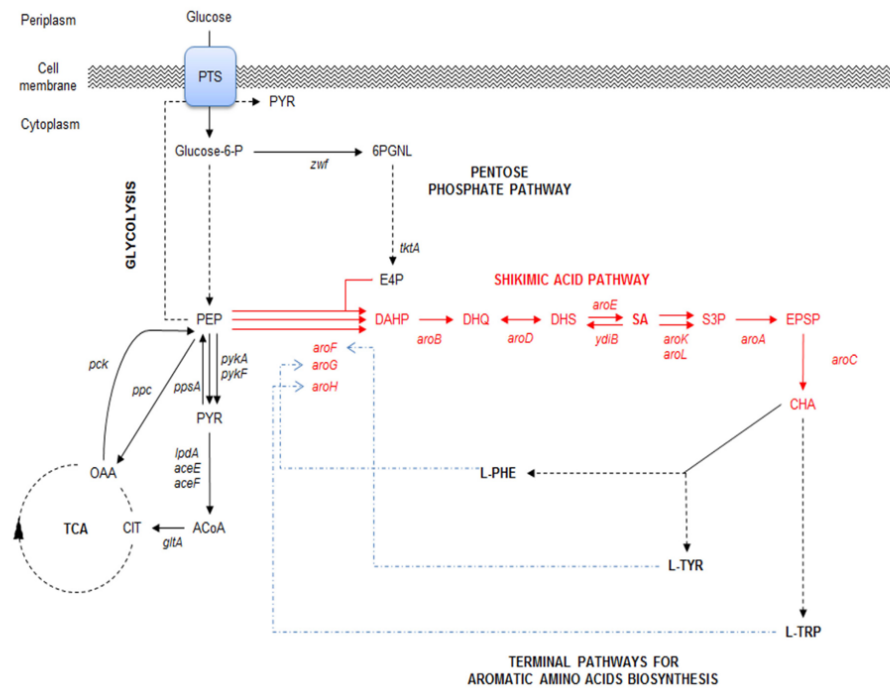


Figura 1.1: Representación esquemática del sistema principal de transporte de glucosa, metabolismo central de carbono (CCM) y su interconexión con la ruta metabólica del SA y aminoácidos aromáticos en *Escherichia coli*. PTS: sistema de fosfotransferasa:PEP:glucosa; TCA: ciclo de los ácidos tricarboxílicos; E4P: eritrosa-4-fosfato; PGNL: 6-fosfo-d-glucono-1,5-lactona; PEP: fosfoenolpiruvato; PYR: piruvato; ACoA: acetil-CoA; CIT: citrato; OAA: oxaloacetato; zwf: glucosa-6-fosfato-1-deshidrogenasa; tktA: transcetolasa I; pykA,pykF: piruvato cinasas II y I; lpda, aceE y aceF: genes codificantes para las subunidades de la piruvato deshidrogenasa; gltA: citrato sintasa; pck: PEP carboxilasa; ppc: PEP carboxilasa; ppsA: PEP sintasa; DAHP: 3-deoxy-d-arabino-heptulosonato-7-fosfato; DHQ: 3-hidroquinato; DHS: 3-dehidroshikimato; SA: ácido shikímico; S3: SK3-3fosfato; EPSP: 5-enolpiruvil-shikimato-3-fosfato; CHA: Corismato; aroF, aroG y aroH: DAHP sintasas; aroB: DHQ sintasa; aroD: DHQ deshidratasa; aroE y ydiB: SHK deshidroquenasa y SHK deshidrogenasa/quinato deshidrogenasa; aroA: 3-fosfosjikimato-1-carboxiviniltransferasa; aroC: CHA sintasa. TRP: triptófano, PHE: fenilalanina; TYR: tirosina. Flechas continuas indican reacciones enzimáticas individuales; flechas discontinuas indican varias reacciones enzimáticas conjuntas; discontinuas-punteadas (azules) muestran la represión alostérica y circuitos de regulación. Tomada de (Martínez et al., 2015)



de niveles de transcripción (Chandran et al., 2003; Chen et al., 2012; Cui et al., 2014; Escalante et al., 2010; Lin et al., 2014; Rodriguez et al., 2013) y iv) la sobreexpresión de enzimas para los pasos limitantes de reacción (Chandran et al., 2003; Cui et al., 2014; Rodriguez et al., 2013).

Siguiendo la primera estrategia de aumentar la biodisponibilidad de los precursores de la ruta del SA (PEP y E4P), una de las principales modificaciones es la eliminación del sistema principal de transporte y fosforilación de carbohidratos (glucosa entre otros) por el sistema de fosfotransferasas (PTS), el cual consume PEP para tal efecto (Chandran et al., 2003; Escalante et al., 2010; Flores et al., 2004, 1996, 2002; Rodriguez et al., 2014). Esta eliminación permite aumentar la disponibilidad de PEP además de incrementar el rendimiento máximo teórico molar de producción de SA en *Escherichia coli* de 43% a 86% (Chandran et al., 2003; Escalante et al., 2010). Sin embargo, esta modificación resulta en una disminución significativa de la capacidad de consumo de glucosa de las cepas (Flores et al., 2004, 2005). Por tal motivo, diversos estudios se han enfocado en el desarrollo de estrategias para recobrar la capacidad de consumo y crecimiento como: la sobreexpresión de enzimas que transporten y fosforilen glucosa (*glf* y *glk* de *Zymomonas mobilis*) (Chandran et al., 2003; Frost et al., 2002) o la adaptación adaptativa en reactores continuos (Aguilar et al., 2012; Escalante et al., 2010; Flores et al., 1996, 2002; Rodriguez et al., 2013). Además de la eliminación del PTS, se han realizado estudios de sobreexpresión de la enzima fosfoenolpiruvato sintasa (PpsA) para la redirección del piruvato (PYR) hacia PEP (Chandran et al., 2003; Chen et al., 2012; Cui et al., 2014), y además de la sobreexpresión de diversas enzimas de la ruta de las pentosas fosfato (PPP) para aumentar el flujo de carbono dirigido a la producción de E4P (Bongaerts et al., 2001; Chandran et al., 2003; Chen et al., 2012; Cui et al., 2014; Draths et al., 1999; Escalante et al., 2010; Patnaik et al., 1995; Rodriguez et al., 2013).

Con respecto a la estrategia de eliminación de las rutas metabólicas de competencia, diferentes grupos de investigación han estudiado la eliminación de las enzimas piruvato cinasas, para evitar el consumo de PEP hacia PYR (Escalante et al., 2010) y la eliminación de las shikimato cinasas para evitar el consumo de SA hacia CHO

(Chandran et al., 2003; Chen et al., 2012; Cui et al., 2014; Escalante et al., 2010; Rodriguez et al., 2013). Esta última modificación provoca que las cepas se tornen auxótrofas a los aminoácidos aromáticos, por lo que deben entonces ser crecidas en medios enriquecidos. Para este efecto, usualmente se utiliza extracto de levadura como substrato, debido a su viabilidad industrial comparada contra medios definidos formulados con los aminoácidos purificados. Otros estudios también han tenido por objetivo la generación de mutantes carentes de enzimas de la ruta que desvían intermediarios hacia otros compuestos no deseados, como por ejemplo YdiB, que convierte el DHQ en ácido quínico (Knop et al., 2001).

Continuando con la estrategia de desregulación de la ruta metabólica, se han utilizado o desarrollado variantes mutantes de las iso-enzimas DAHPs que presentan insensibilidad a la inhibición alostérica por producto. De esta forma se evita la limitación del flujo hacia la AAAP en presencia de los aminoácidos aromáticos esenciales para el crecimiento (Chandran et al., 2003; Escalante et al., 2010; Keseler et al., 2013; Lin et al., 2014; Patnaik et al., 1995; Rodriguez et al., 2014). Además, se han realizado diversos experimentos de evolución adaptativa en fermentadores utilizando diferentes fuentes de estrés y de incremento de mutación, para generar cepas variantes con flujos mayores hacia la producción de SA, como limitación de carbono y evolución cromosomal inducida por agentes químicos (CIChE) como triclosán (Cui et al., 2014; Flores et al., 2005, 2007, 1996). Finalmente, con respecto a la estrategia de sobreexpresión de las enzimas de la vía para evitar pasos limitantes de flujo, se han sobreexpresado los genes que codifican para las enzimas AroB, AroD y AroE en diversos sistemas de expresión, a nivel de plásmidos o de integraciones cromosomales, y bajo diferentes promotores para regular la expresión de las enzimas en diferentes etapas del cultivo, y la proporción entre estas, así como para aumentar su fuerza de expresión (Cui et al., 2014; Rodriguez et al., 2013).

El uso de estas estrategias de IVM llevaron a la construcción de diversas cepas sobreproductoras de SA, de entre las cuales sobresalen las mostradas en el Cuadro 1.1. Dentro de estas, la cepa con mayor capacidad productora hasta el momento es la cepa *E. coli* SP1.1*pts*-/pSC6.090B (Chandran et al., 2003), la cual alcanza un título

final de 87 g/L de ácido shikímico y un rendimiento mol/mol del 36 % cuando es cultivada en glucosa en modo lote alimentado de 10 L. Esta cepa fue construida a partir de la SP1.1 construida por Draths y colaboradores en 1999 y contiene la inserción del gen *aroB* en el locus de *serA*, además de la interrupción de los genes *aroK* y *aroL* (Draths et al., 1999). Chandran y colaboradores le interrumpieron el sistema PTS de transporte de glucosa dependiente de PEP para acumular el intermediario y redirigirlo hacia la vía de los aminoácidos aromáticos (Chandran et al., 2003). Esta cepa SP1.1*pts-* fue posteriormente transformada con el plásmido pSC6.090B, el cual contiene los genes *glf* y *glk* de *Zymomonas mobilis*, que codifican para un transportador de glucosa y la enzima glucosa cinasa, con lo que lograron recuperar la capacidad de la cepa para internalizar y fosforilar glucosa sin el sistema PTS. Además, también contiene los genes *tktA*, *aroE* y *serA* que codifican para la transcetolasa A, la shikimato deshidrogenasa y la D-3-fosfoglicerato deshidrogenasa respectivamente. De esta forma Chandran y colaboradores lograron aumentar la biodisponibilidad de los metabolitos precursores de la vía (PEP y E4P) por medio de la interrupción del sistema PTS y mediante la expresión del gen *tktA*, aumentar el flujo de carbono hacia shikimato por medio de los genes *aroE* y *aroB*, permitiendo la acumulación de shikimato por medio de la interrupción de los genes *aroK* y *aroL* (Chandran et al., 2003).

Los estudios de este grupo de investigación derivaron en una patente para un proceso fermentativo para la producción de ácido shikímico desde glucosa, con cepas modificadas de *Escherichia coli* que, alcanzaron producciones de hasta 71 g/L de SA en el cultivo con rendimientos del 27 % mol/mol (Frost et al., 2002). Esta patente a su vez derivó en la creación de una empresa que logró en el 2005 tener una capacidad instalada de producción de 200 toneladas de ácido shikímico por año (Bradley, 2005). Este éxito aumentó el interés en el desarrollo de cepas genéticamente modificadas para la sobreproducción de SA a escala laboratorio e industrial, permitiendo llegar a rendimientos relativamente altos (entre 40-50 % mol/mol) (Díaz-Quiroz et al., 2014; Ghosh et al., 2012; Krämer et al., 2003; Martínez et al., 2015; Rodríguez et al., 2013). En este sentido la cepa AR36 es la que presenta el rendimiento molar más alto al momento (0.42 mol/mol en glucosa, al rededor del 50 % del rendimiento máximo

teórico) (Rodríguez et al., 2013), aunque el título final es menor al reportado por Chandran et al. (2003) y aún se encuentra lejana al máximo teórico (86 % mol/mol).

Cuadro 1.1: Cepas productoras de ácido shikímico, producción y rendimiento. Modificada de Martínez et al. (2015).

Cepa	Características Fenotípicas	Observaciones	Título (g/L)	Rendimiento (mol/mol)
SA112 (Cui et al., 2014)	BW25113 $\Delta$ aroKL, Ppps::PlacQ1,PcsrB::PlacQ1Pt5 -pps,PT5-csrB, 5Ptac-tktA	CICHÉ: evolución para optimizar SA	1.70	0.25
DHPYAAS-T7 (Chen et al., 2012)	DH5 $\alpha$ $\Delta$ ptsH $\Delta$ err, $\Delta$ aroKL, $\Delta$ ydiB pAOC-TGEFB:aroE, aroB, glk, tkta, aroFfbr	Expresión plasmídica de genes relacionados al SA	1.066	0.23
PB12.SA22 (Escalante et al., 2010)	JM101 $\Delta$ ptsH,ptsI,cr::Kmr $\Delta$ arokL::cmr pJLBaroGfbrtktA, pTOPO-aroBaroE	Evolucionada en laboratorio derivada PTS-	7.05	0.22
SA5 (Chen et al., 2014)	B0013 $\Delta$ arokL::dif $\Delta$ ptsG::dif $\Delta$ ydiB:dif $\Delta$ ackA-pta::dif pTH-aroGfbr-ppsA-tktA	Expresión plasmídica de genes relacionados al SA	14.6	0.3
SA114 (Cui et al., 2014)	BW25113 $\Delta$ aroKL, Ppps::PlacQ1, PcsrB::PlacQ1 PT5-pps,PT5-csrB, 5Ptac-tktA, 5Ptac-pntAB	CICHÉ: evolución para optimizar SA	2.99	0.31
SP1.1pts-/ pSC6.090B (Chandran et al., 2003)	DH5 $\alpha$ $\Delta$ ptsH-ptsI-err $\Delta$ serA::aroB $\Delta$ aroL::Tn10 $\Delta$ aroK::Cmr Ptacglf glk, aroFfbr tktA PtacaroE, serA	glk y glf heterólogos de <i>Z. mobilis</i>	84	0.33
SA116 (Cui et al., 2014)	BW25113 $\Delta$ aroKL, Ppps::PlacQ1,PcsrB::PlacQ1 PT5-pps, PT5-csrB, 5Ptac-tktA, 5Ptac-nadK	CICHÉ: evolución para optimizar SA	3.12	0.33
AR36 (Rodríguez et al., 2013)	JM101 $\Delta$ ptsH, ptsI, cr::Kmr $\Delta$ arokL::cmr $\Delta$ pykF $\Delta$ lacI pTrcAro6-aroB, tktA, aroGfbr, aroE, aroD zwf	Expresión fuerte, constitutiva por medio de operón sintético en plásmido	41.8	0.42

Para que estas cepas puedan ser competitivas en la industria es necesario crear procesos de producción de bajo costo manteniendo la eficiencia de la síntesis a gran escala de forma que los costos de producción sean menores o similares a los de las síntesis tradicionales (Báez et al., 2001; Báez-Viveros et al., 2004; Chávez et al., 2005; Leib et al., 2001; Patnaik et al., 1995). En específico para el proceso de obtención de SA, es necesario aumentar tanto el rendimiento como las productividades del proceso. Esto es debido a que las características, del producto y del mercado, requieren que sea producido en reactores de gran escala.

La ingeniería de vías metabólicas (IVM) ha permitido redirigir flujos de carbono en una red bioquímica hacia la síntesis del ácido shikímico por medio de la manipulación genética. Sin embargo, a menudo estas modificaciones inducen efectos secundarios y desequilibrios en el CCM que provocan resultados no deseados o usos incompletos de las capacidades de conversión de la cepa. Esto se debe a que pueden

interferir con la disponibilidad de algunos precursores importantes y los balances de energía en el organismo, causando la acumulación de intermediarios y subproductos no deseados. De esta forma reducen la aptitud de la cepa para crecer y producir compuestos de interés (Biggs et al., 2014). Dicha problemática surge de la dificultad de predecir los resultados finales de un sistema complejo con información parcial, como lo es el metabolismo celular compuesto por varias capas de información interrelacionadas (genoma, transcriptoma, proteoma y metaboloma, entre otros), y donde generalmente se utilizan estrategias de modificación local para las rutas metabólicas en estudio (Cloots y Marchal, 2011; Fong, 2014; Matsuoka y Shimizu, 2013). Debido a la importancia no sólo de encontrar modificaciones a la vía metabólica particular sino también de información metabólica sistémica y su relación con la fisiología celular, para entender con mayor detalle y atender las limitaciones en la sobreproducción de metabolitos de interés, es necesaria una aproximación de biología de sistemas. Esta aproximación ha crecido en interés gracias a la posibilidad de obtener información global de los diferentes componentes de los organismos (genes, transcritos, proteínas y metabolitos, entre otros) como de la capacidad de estudiarlos y modelarlos como sistemas complejos.

#### **1.1.4. La modelación matemática como herramienta para la ingeniería de vías metabólicas**

En años recientes se ha generado un creciente interés en el desarrollo de técnicas y modelos que ayuden a describir y comprender en mayor detalle los procesos metabólicos y su regulación a nivel sistémico. Impulsado por las más recientes técnicas de obtención de información de alto rendimiento (*ómicas*) se ha logrado un importante avance en modelos matemáticos para resolver y clarificar comportamientos de las redes metabólicas complejas y sus interacciones (Akesson et al., 2004; Covert et al., 2001a; Kim et al., 2012; Stelling, 2004). En efecto, la construcción de modelos matemáticos ha probado ser capaz de generar hipótesis, información y predicciones que pueden ser acotados y validados experimentalmente. A la fecha, se han desarrollado varios marcos de modelación matemática y sus aproximaciones han resultado útiles para la identificación de blancos genéticos de modificación (genes, rutas y sistemas), y para la selección y mejoramiento de características fenotípicas relevantes para el

mejoramiento de la producción de compuestos de interés (Jouhten, 2012; Long et al., 2015). Dichos marcos de modelación han ayudado a revelar propiedades de los mecanismos de regulación y han clarificado algunas de las complejas interacciones de la red metabólica, así como de características sistémicas metabólicas (Stelling, 2004).

De forma general, los modelos inician con la generación de una representación matricial de redes metabólicas construidas con suficiente detalle y curadas con los datos obtenidos de genomas, transcriptomas, metabolomas u otras técnicas. Dicha matriz es después cruzada con operadores matemáticos derivados de información estequiométrica, termodinámica y/o cinética disponible para las condiciones deseadas. Debido a la complejidad de los microorganismos, que comprenden de cientos a miles de reacciones, el conjunto de parámetros requeridos para describir el comportamiento y las características con estas redes metabólicas en estos operadores matemáticos es generalmente mayor a la cantidad de datos posibles a obtener. Esto significa que los sistemas de ecuaciones resultantes son indeterminados y por lo tanto es necesario desarrollar aproximaciones de modelación y computo intensivas y eficientes (Price et al., 2003). Por estos motivos, la mayoría de los modelos hoy en día se construyen para condiciones con el menor número de substratos posibles y así reducir los balances de materia necesarios asumiendo estados pseudo-estacionarios (Akesson et al., 2004; Fong, 2014; Jouhten, 2012; Kim et al., 2012). Aun así, este tipo de sistemas de ecuaciones indeterminados presenta la característica de no tener una solución única, por lo que se genera un espacio convexo que contiene todas las soluciones individuales del sistema. Dicho espacio puede ser acotado después con datos experimentales adicionales o algunas suposiciones derivadas del conocimiento del comportamiento celular que contribuyan a encontrar una solución significativa y útil (Akesson et al., 2004; Fong, 2014; Jouhten, 2012; Kim et al., 2012). Dentro de las aproximaciones desarrolladas para encontrar dichas soluciones se encuentran las basadas en mecanismos, en interacciones y variables de regulación, y las basadas en restricciones. Estas últimas son las más usadas, debido a que pueden reducir el espacio a una sola distribución de flujo, incluso con poca información sobre el sistema (Price et al., 2003; Stelling, 2004).

A pesar del rápido desarrollo del área, todavía existe un campo desafiante para la construcción de modelos en condiciones dinámicas y con medios de crecimiento poco o indefinidos, así como una mayor integración de información derivada de las técnicas ómicas (Kim et al., 2012; Long et al., 2015; Machado y Herrgard, 2014; O'Brien y Palsson, 2015; Patil et al., 2004; Ramkrishna y Song, 2012; Saha et al., 2014; Schuetz et al., 2007; Stelling, 2004). De esta forma el presente trabajo tiene la finalidad de contribuir al desarrollo de la producción de SA en *Escherichia coli* mediante el uso de modelación matemática para la determinación de los flujos metabólicos en una cepa sobreproductora de SA previamente construida con estrategias clásicas de ingeniería de vías metabólicas y de esta forma determinar posibles blancos de modificación genética y parámetros de fermentación para el mejoramiento de sus capacidades de producción.

## 1.2. ANTECEDENTES

### 1.2.1. Cepas de *Escherichia coli* carentes del sistema PTS, incluyendo la AR36

El grupo de investigación del Dr. Bolivar, ha generado varias cepas sobreproductoras de ácido shikímico. En específico, se ha trabajado con cepas a las que se les eliminó el sistema principal de transporte y fosforilación de glucosa (PTS) (Escalante et al., 2010; Flores et al., 1996). Flores et al. (1996) reportaron la construcción de una cepa, derivada de la cepa silvestre de *Escherichia coli* JM101, carente de este sistema PTS (PB11). Esta cepa presenta una capacidad reducida de crecimiento en medio mínimo con glucosa como única fuente de carbono (fenotipo PTS- Glucosa-), aproximadamente  $0.1 \text{ h}^{-1}$  (reducción del 85 % con respecto a la cepa silvestre). Para recuperar la capacidad de crecimiento en glucosa, la cepa PB11 fue sometida a un experimento de evolución adaptativa en cultivo continuo en medio mínimo con glucosa como única fuente de carbono, aumentando la tasa de dilución del sistema lavando así células con menor tasa de crecimiento que la tasa de dilución. Con este sistema se seleccionaron mutantes espontáneas dentro de las cuales destaca la PB12 que presentó un aumento en la velocidad de crecimiento hasta valores de  $0.4 \text{ h}^{-1}$  debido al aumento de la capacidad de consumo de glucosa (fenotipo PTS- Glucosa+) (Flores et al., 1996). Las mutaciones sufridas por la cepa PB12 durante la evolución adaptativa fueron estudiadas por Aguilar y colaboradores al secuenciar, analizar y comparar su genoma respecto al de la PB11. De esta forma se encontró que la cepa PB12 carece de un fragmento cromosomal de 10,328 pares de bases que contiene los genes *rpph*, *mutH* y *galR*. Se encontraron también varias mutaciones puntuales en genes reguladores como *arcB*, *rpoS*, *barA*, *rna*, *yjjU*, *rssA* y *ypdA* (Aguilar et al., 2012, 2018). Además, debido a la escisión del operón *ptsHIcrr* del sistema PTS, ambas cepas carecen de represión catabólica por glucosa, haciéndolas capaces de co-utilizar varias fuentes de carbono (Hernandez-Montalvo et al., 2001; Martínez et al., 2008). Los efectos de las diversas mutaciones fueron estudiados posteriormente en diversos trabajos publicados por el grupo de investigación y se discuten a continuación.



Las mutaciones en los genes *arcB*, *rpoS* y *rppH* en la cepa PB12 generaron mayores niveles de transcripción de los genes glicolíticos y otros genes del metabolismo central del carbono, incluyendo los de TCA y los involucrados en el metabolismo de ppGpp (Flores et al., 2008, 2005, 2007, 1996, 2002). La mutación en *arcB* fue relacionada con la modificación de la capacidad de represión del sistema de dos componentes ArcA-ArcB, al posiblemente alterar la conformación de ArcB por un cambio de aminoácido tirosina a cisteína en la posición 71, y provocando la sobreexpresión de genes de producción y consumo de acetato como *poxB* y *acs* (Flores et al., 2004, 2005). Se estudiaron además los efectos de inactivar el gen *arcA* en PB12, lo que generó un aumento de la transcripción de los genes de TCA y un incremento en la velocidad de crecimiento del 246 % (Flores et al., 2007). Por su parte la inactivación total de *rpoS* en PB12 provocó una disminución de la velocidad de crecimiento (Flores et al., 2008). La inactivación del gen *galR*, el cual codifica para el represor del operón *gal*, fue responsable del aumento en la expresión del gen *galP* (transportador GalP). GalP es el responsable del transporte de glucosa en medio mínimo en ausencia del sistema PTS (Flores et al., 2005, 2002). La inactivación del gen *rppH*, que codifica para una RNA pirofosfodihidrolasa que inicia la degradación del mRNA, fue establecida como la responsable de las altas concentraciones de transcrito observadas en PB12 comparadas con las cepas JM101 y PB11 (Aguilar et al., 2012; Flores et al., 2008, 2005, 2002). Esta última característica permitió la sobreexpresión de los genes de glicólisis y, por lo tanto, un aumento conjunto del transporte por GalP y la fosforilación por Glk (Flores et al., 2002). Por su parte, la ausencia del gen *mutH* que codifica para la endonucleasa del complejo MthHLS, fue implicada en la acumulación de mutaciones en la cepa PB12 durante la evolución adaptativa (Aguilar et al., 2012; Flores et al., 2005).

La cepa PB12 se caracterizó para producción de aminoácidos aromáticos y algunos intermediarios de esta vía, resultando capaz de redirigir el fosfoenolpiruvato que no utiliza para el transporte y fosforilación de glucosa hacia la vía del shikimato (Flores et al., 2005, 1996, 2002). Para lograr que esta cepa obtuviera mejores rendimientos y títulos, en el laboratorio del Dr. Francisco Bolivar se llevaron a cabo estrategias de modificación genética basadas en estrategias clásicas de ingeniería de

vías metabólicas como las mencionadas en la sección de introducción. Dentro de las cepas construidas resalta la cepa PB12.SA22/JLBaroG<sup>fbr</sup>tktA/pTOPOaroBaroE, la cual es capaz de una producción de shikimato de 7.5 g/L con un rendimiento de 0.29 mol/mol con respecto a la glucosa en volúmenes de cultivo de 500 mL (Escalante et al., 2010). Esta cepa tiene inactivos los genes *aroK* y *aroL*, que limitan la vía de los aromáticos hasta la producción de shikimato. Además, contiene dos plásmidos: pJLBaroG<sup>fbr</sup>tktA y pTOPOaroBaroE, con los cuales se sobrexpresan los genes *aroG<sup>fbr</sup>*, *tktA*, *aroB* y *aroE* haciendo uso de promotores *lacUV5* inducibles por IPTG (Escalante et al., 2010). Con construcciones derivadas de esta cepa se estableció que la inactivación adicional del gen *pykF* permite obtener rendimientos más altos de compuestos aromáticos totales (0.5 molTAC/molGlc), indicando que la cepa PB12 puede redirigir aún más carbono hacia la vía de los aromáticos (Escalante et al., 2010).

Estos antecedentes llevaron finalmente a la construcción de la cepa AR36 (Rodríguez et al., 2013), cepa derivada de la PB12 a la cual se le eliminaron los genes *aroK* y *aroL* para permitir la acumulación del SA, y el gen *pykF* para disminuir el consumo de PEP por vías alternas a la ruta del SA. Finalmente, a la cepa se le removió el gen *lacI* para hacer constitutiva la expresión de los genes que dependen de los promotores inducibles por IPTG. Esta cepa además contiene el plásmido de alto número de copias pTrcAro6, en el que se encuentra un operón sintético compuesto por los genes: *aroB*, *tktA*, *aroG<sup>fbr</sup>*, *aroE*, *aroD* y *zwf* (Rodríguez et al., 2013). La expresión sincrónica y constitutiva de todos los genes involucrados en la ruta de los aminoácidos aromáticos hasta el SA, y de los genes *tktA* y *zwf* para aumentar la biodisponibilidad de E4P, permitió que la cepa acumulara hasta 43.3 g/L de shikimato, con un rendimiento de 0.42 mol/mol y una productividad volumétrica de 1.44 gSA/Lh (Rodríguez et al., 2013). Además, este sistema de expresión permitió reducir la acumulación de otros intermediarios de la vía en cantidades menores al 6.7% mol/mol GLC. Es importante notar que estos resultados fueron obtenidos en fermentadores con 100 g/L de glucosa suplementados con 30 g/L de extracto de levadura, condiciones de alta presión osmótica y saturación metabólica, lo que generalmente provoca grandes producciones de acetato y la concomitante pérdida de aptitud celu-

lar. Sin embargo, las fermentaciones presentaron concentraciones relativamente bajas de acetato (cerca de 8.65 g/L) y no disminuyeron las velocidades de crecimiento y consumo de glucosa (Rodríguez et al., 2013). Esto fue atribuido específicamente al sistema de expresión y al operón sintético utilizado en AR36, dado que se comprobó que su inclusión en la cepa aumenta el metabolismo glicolítico, la velocidad de crecimiento y disminuye la proporción de acetato producido al aumentar la producción de SA por medio de experimentos de metabolómica y fluxómica en medio mínimo (Rodríguez et al., 2017).

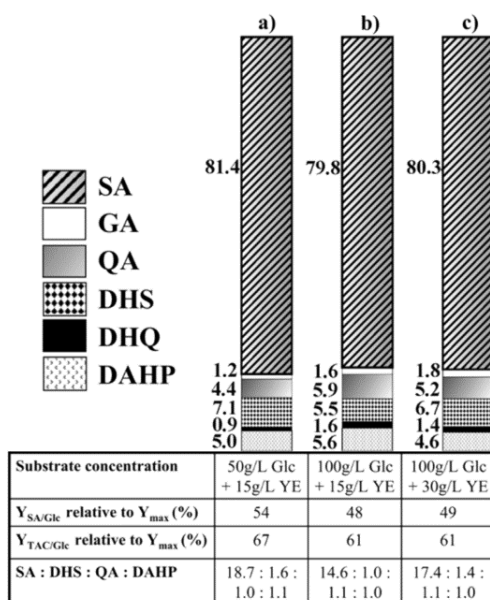


Figura 1.2: Porcentajes molares de los compuestos aromáticos producidos en la cepa AR36 en cultivos lote con concentraciones altas de sustrato. Tomada de Rodríguez et al. (2013).

Debido a estas características, la cepa de *Escherichia coli* AR36, presenta gran potencial de uso industrial ya que está más cerca del rendimiento máximo teórico, no necesita inductor durante las fermentaciones, presenta poca acumulación de intermediarios no deseados de la vía de SA (lo que facilita la purificación de SA) y produce relativamente bajas cantidades de acetato en condiciones de alto contenido de sustratos. Sin embargo, está claro que aún existen retos importantes para el mejoramiento de la producción de SA en esta cepa, tanto en aumentar el rendimiento a

valores cercanos al máximo teórico como en mejores estrategias de fermentación que aumenten la productividad del proceso utilizando la visión de la biología de sistemas.

### 1.2.2. Modelos metabólicos para la optimización de la producción de ácido shikímico

Con respecto a la producción de SA en *E. coli*, pocos estudios de modelación existen en la literatura. Chen y colaboradores (2011) usaron un análisis basado en el balance de flujos (FBA), asumiendo cero crecimiento y utilizando la producción de SA como el objetivo de optimización. Los modelos identificaron a los genes *aroF*, *tktA*, *ppsA* y *glf* como candidatos para sobreexpresión, así como a los genes *ldhA* y *ackA* para inactivación (Chen et al., 2011). Dichos genes y nodos están en concordancia con reportes previos para la producción de intermediarios AAAP (Chandran et al., 2003; Gosset, 2009; Martínez et al., 2015). Sin embargo, los modelos también identificaron al gen *zwf* como responsable de la redirección de flujo de carbono hacia la AAAP. Su sobreexpresión ha resultado en conversiones molares de GLC hacia aminoácidos aromáticos de hasta en un 47% (Chen et al., 2011; Flores et al., 2002; Rodriguez et al., 2013). Por su parte, Ahn et al. (2008) construyeron un modelo para maximizar la producción de SA desde GLC, el cual resaltó la importancia de genes del CCM como *tktA* y *zwf* (Ahn et al., 2008). Sin embargo, tampoco en este modelo fueron considerados el crecimiento y los requerimientos de mantenimiento. Rizk y Liao (2009) (Rizk y Liao, 2009) usaron un modelado por conjuntos, basados en mecanismos de reacción, para identificar a *tktA* como responsable de la reacción de control, lo que significó que la sobreexpresión del gen *ppsA* aumenta la producción de intermediarios aromáticos sólo cuando *tktA* se sobreexpresa simultáneamente (Rizk y Liao, 2009).

Aun con estas aportaciones, todavía existen muchos desafíos que deben afrontarse y atender con respecto a la construcción e implementación de modelos matemáticos para este sistema. Los modelos a menudo se encuentran limitados por supuestos específicos como el del estado estacionario. En específico, el asumir el estado estacionario provee sólo información limitada sobre las propiedades dinámicas de regulación del sistema metabólico. Esta limitación puede resultar en algunas contradicciones de

los resultados de los modelos con el comportamiento real de las cepas en condiciones dinámicas de crecimiento, causadas por la existencia de los complejos mecanismos de regulación. Nuevos modelos y herramientas que ayuden a tomar en cuenta los sistemas biológicos de forma sistémica y en condiciones dinámicas podrían resultar en un mejor entendimiento del comportamiento de los microorganismos sobreproductores de SA.

De hecho, para cepas de *E. coli* construidas para la sobreproducción de SA, la mayoría del trabajo ha sido enfocado en el diseño e implementación de plataformas de expresión, estudio de diferentes fondos genéticos y uso de estrategias de cultivo basados en información obtenida por métodos clásicos. Pocos estudios se han enfocado en la generación de información basada en la modelación de metabolismo para entender y diseñar cepas productoras de SA, y aún menos se ha hecho en la construcción de modelos para cepas usando medios complejos de crecimiento y/o en condiciones dinámicas, las cuales son comunes en los procesos de producción de SA u otros compuestos. En este reporte, se describe una aproximación dinámica de modelado del comportamiento fisiológico y de las distribuciones de flujo metabólico para una cepa de *E. coli* modificada por ingeniería de vías metabólicas para la sobreproducción de SA, en medio complejo.

### 1.2.3. Modelado cibernético

Como se estableció en las secciones anteriores, el desarrollo de estrategias matemáticas que nos permitan estudiar de forma sistémica las relaciones entre las diferentes enzimas, metabolitos, flujos y sus características de regulación, es de vital importancia para el desarrollo de nuevas estrategias de IVM y de mejores procesos de producción. El metabolismo celular está compuesto por diversas reacciones que van desde el consumo de nutrientes hasta su transformación en numerosos componentes intracelulares y extracelulares (Ramkrishna y Song, 2012). Estas reacciones forman una compleja red de caminos metabólicos catalizados por enzimas específicas. Dentro de cada uno de estos caminos metabólicos puede existir una o más enzimas críticas o limitantes para la velocidad de flujo de carbono a través de estos (Ramkrishna y

Song, 2012). Los marcos matemáticos de modelación desarrolladas hasta la fecha se han concentrado en general en estrategias basadas en la estequiometría y diseñando sistemas lineales de ecuaciones diferenciales en estado estacionario. La inclusión de la temporalidad y la regulación en este tipo de sistemas de ecuaciones ha sido poco estudiada. Esto se debe a la complejidad y cantidad de parámetros desconocidos o difíciles de calcular experimentalmente que hacen, casi imposible, obtener una descripción completa. Esto representa un serio reto para lograr el objetivo último de la modelación matemática: la generación de estructuras y conocimientos sistémicos de la complejidad del metabolismo que sean capaces de ser predictivos y contribuyan en la toma de decisiones para el desarrollo de nuevas cepas y procesos productivos (Rizk y Liao, 2009).

Lo anterior significa que es necesario desarrollar nuevas estrategias que tomen en cuenta funciones de estado dependientes de las concentraciones de los metabolitos en el sistema a partir de una distribución estequiométrica dada restringida por uno o varios objetivos metabólicos; Sino también incluir funciones sobre la dinámica de la regulación derivada de perturbaciones o cambios constantes en el medio en el que se crecen los microorganismos. Con respecto a este objetivo dinámico de modelación, la mayoría del esfuerzo realizado ha sido enfocado a la deducción de parámetros cinéticos mecanísticos y a la construcción de sistemas de ecuaciones diferenciales deterministas mediante el uso y aproximación de ecuaciones clásicas de cinética enzimática (Rizk y Liao, 2009). Con respecto a la regulación la mayoría de los modelos incluyen estos efectos como valores de restricción y en formulaciones *ad hoc*. Por ejemplo, inactivando o activando flujos de rutas metabólicas por medio de operadores booleanos derivados de datos transcriptómicos (Covert et al., 2001a,b; Ramkrishna y Song, 2012; Rizk y Liao, 2009).

En las últimas décadas diferentes esfuerzos no convencionales de racionalización del problema dinámico y de regulación derivaron en la construcción de estrategias de modelado capaces de atacar tanto la dinámica como la asignación de recursos limitados a través del sistema celular, aún con poca información mecanística detallada sobre los procesos. El principal marco de modelación matemática con estas carac-

terísticas particulares es el desarrollado por Ramkrishna y colaboradores (2012), denominado marco cibernético. Esta aproximación de modelación permite incluir respuestas de regulación dinámica a través de la red metabólica y por tanto tomar en cuenta la plasticidad metabólica y su multiplicidad atendiendo a cambios constantes de los parámetros extracelulares e intracelulares (Ramkrishna y Song, 2012). El marco cibernético de modelación es de hecho uno de los pocos, si no el único, que permite una representación sistemática de la influencia de la regulación metabólica en cuanto al control de la actividad enzimática, así como de su expresión de forma dinámica (Ramkrishna y Song, 2012; Varner y Ramkrishna, 1999).

La base conceptual del marco cibernético recae en la hipótesis de que los sistemas metabólicos han evolucionado hacia la optimización de “programas” orientados a resultados. Es decir, que su proceso de evolución les ha permitido construir sistemas de reacción concretos, específicos y eficientes para contender con perturbaciones extracelulares, intracelulares e inclusive genéticas, permitiendo su supervivencia y mejorando su aptitud en diversos ambientes (Ramkrishna y Song, 2012; Varner y Ramkrishna, 1999). El uso de este tipo de objetivos de optimización es habitual en marcos de modelación como el análisis basado en flujos (FBA), en el cual se utiliza por lo general la optimización del rendimiento de biomasa como función objetivo. Esto se realiza bajo la premisa fundamental de que los organismos utilizan su maquinaria para maximizar su biomasa, aunque también se han utilizado otros objetivos metabólicos con buenos resultados como el consumo de oxígeno, consumo de carbono, producción de ATP, entre otros. Sin embargo, en los marcos basados en restricciones más comunes, se utiliza generalmente una solución única del espacio convexo y es representada por ecuaciones de estado estacionario. Mientras que, por su parte, la modelación cibernética usa por lo general varias soluciones del espacio convexo de forma simultánea y coordina su participación en el metabolismo mediante la inclusión de funciones dinámicas de regulación. Dichas funciones evalúan a través del tiempo cada una de las distribuciones de flujo asociadas a diversos objetivos celulares para representar el resultado final del comportamiento metabólico. Tal es el caso de la combinación y competencia de diferentes rutas metabólicas por los recursos necesarios para asegurar su aptitud de acuerdo al medio externo y las alternativas

de crecimiento, producción y consumo de metabolitos (Ramkrishna y Song, 2012; Varner y Ramkrishna, 1999).

De forma específica, el marco de modelación cibernética introduce la regulación mediante el uso de dos vectores  $u \equiv [u_1, u_2, u_3, \dots, u_m]$  y  $v \equiv [v_1, v_2, v_3, \dots, v_m]$ . Estos vectores están asociados con la distribución fraccional de los recursos necesarios para la síntesis de enzimas y su actividad (Ramkrishna y Song, 2012). Estas variables son calculadas a lo largo del proceso biológico y modifican la participación de cada conjunto de flujos expresados como modos elementales (EMs) obtenidos del análisis estequiométrico de la matriz metabólica. Estos EMs son conjuntos de reacciones en rutas metabólicas mínimas e indecomposables que relacionan metabolitos de entrada y de salida. El conjunto total de EMs describe, por tanto, todas las posibilidades metabólicas del organismo. La descripción matemática utilizando todos los EMs resultaría en un problema de optimización matemática imposible, debido al gran número de parámetros necesarios para el cálculo de un resultado integrado. Sin embargo, no todos los EMs son utilizados de forma simultánea por el microorganismo, por lo que es posible seleccionar un subconjunto de EMs que puedan describir el comportamiento metabólico en una escala paramétrica asequible. Para estos efectos, se utilizan estrategias de reducción por medio análisis de rendimientos o flujos, para obtener el conjunto mínimo útil para describir el comportamiento en las condiciones experimentales (Song y Ramkrishna, 2009). Una vez seleccionados los EMs participantes se puede construir el marco de ecuaciones diferenciales que van a dictar las velocidades de flujo para cada uno de ellos. El marco de modelación cibernética incluye también las velocidades de síntesis de enzimas críticas y/o limitantes para cada EMs, lo que lo ha hecho distintivo de otros marcos de modelación (Ramkrishna y Song, 2012). Este marco describe la velocidad de flujo y la velocidad de producción de enzimas clave por medio de ecuaciones tipo Michaelis-Menten, y son estas velocidades las que se regulan por los valores de  $v$  y  $u$  respectivamente para cada EMs, multiplicándolas con valores entre 0 y 1 derivados de una o varias funciones objetivo para cada tiempo  $t$  dado. Estas variables son calculadas de forma general por medio de ecuaciones basadas en leyes de igualación. Generalmente para  $u$  se relaciona el retorno de la función objetivo de cada EM con la sumatoria de



los retornos de todos los EMs, mientras que para  $v$  se relaciona el retorno de cada EMs con el retorno más alto encontrado para el conjunto de EMs. De esta forma, las variables cibernéticas regulan la participación final de los flujos producidos, por cada EMs en el metabolismo para cada diferencial de tiempo ( $t + \Delta t$ ) aun teniendo poca información mecanística de las particularidades de la regulación celular (Ramkrishna y Song, 2012; Song y Ramkrishna, 2009; Varner y Ramkrishna, 1999).

En el presente trabajo se hizo uso de este marco de modelación matemática para poder revelar las características de las distribuciones de flujo de la cepa AR36 de forma dinámica para fermentaciones en condiciones variadas de concentración de los substratos utilizados para su crecimiento y producción de shikimato. Los detalles particulares de la construcción del modelo se tratarán más a fondo en la sección de materiales, métodos y modelos.

---

## Capítulo 2

# PLANTEAMIENTO DEL PROBLEMA DE INVESTIGACIÓN.

## 2.1. JUSTIFICACIÓN e HIPÓTESIS

### 2.1.1. Justificación

La mayoría del trabajo reportado para mejorar la producción biotecnológica de SA en *Escherichia coli*, ha sido enfocado en el diseño e implementación de plataformas de expresión, estudio de diferentes fondos genéticos y de estrategias de cultivo basados en información obtenida por métodos clásicos de ingeniería metabólica. Sin embargo, estas estrategias a menudo inducen desequilibrios en el CCM que provocan resultados no deseados. Esto se debe a que pueden modificar la disponibilidad de algunos precursores biosintéticos y energéticos, causando la acumulación de subproductos no deseados o el agotamiento de metabolitos o cofactores. Pocos estudios se han enfocado a la generación de información basada en la modelación del metabolismo celular. Menos aún se han desarrollado modelos para cepas modificadas por ingeniería de vías metabólicas creciendo en medios complejos conteniendo más de un substrato (glucosa y extracto de levadura) y/o en condiciones dinámicas, las cuales son comunes en los procesos de producción industrial de SA. Por tal motivo, la generación de nuevos modelos y herramientas que ayuden a tomar en cuenta el comportamiento de forma sistémica en condiciones dinámicas, pudieran favorecer un mejor entendimiento del comportamiento de los microorganismos y con ello el poder mejorar las capacidades de producción de compuestos de interés como el SA.

### 2.1.2. Hipótesis

El comportamiento fisiológico y metabólico de la cepa AR36 en fermentaciones realizadas en medio complejo y a diferentes concentraciones de substrato podrán ser representados por medio de modelos dinámicos basados en un marco de modelación cibernética. Dicha representación permitirá generar información suficiente y relevante para el mejoramiento de la producción de ácido shikímico con AR36.

## 2.2. OBJETIVOS

### 2.2.1. Objetivo general

Diseñar, construir y validar modelos dinámicos fisiológicos y metabólicos que permitan caracterizar el comportamiento de la cepa *Escherichia coli* AR36 sobreproductora de ácido shikímico en medio complejo con variaciones en la concentración de glucosa y extracto de levadura, y que sirvan para identificar parámetros que permitan mejorar la producción de SA.

### 2.2.2. Objetivos particulares

1. Diseñar y construir modelos para describir el comportamiento fisiológico en variaciones de la concentración de los dos sustratos utilizados.
2. Caracterizar el comportamiento fisiológico mediante la construcción y análisis de superficies de respuesta de los parámetros modelados, en diferentes etapas de la fermentación.
3. Validar las superficies modeladas mediante la comparación de predicciones del comportamiento contra experimentos no contenidos en el diseño experimental.
4. Diseñar y construir modelos para describir el comportamiento de las distribuciones de flujos metabólicos en variaciones de las concentraciones de los dos sustratos utilizados.
5. Caracterizar el comportamiento de las distribuciones de flujo mediante la construcción y análisis de superficies de respuesta de los flujos modelados en diferentes etapas de la fermentación.
6. Analizar y comparar los resultados obtenidos de los modelos fisiológicos y metabólicos para la determinación de blancos genéticos y metabólicos para el mejoramiento de la producción de ácido shikímico
7. Mejorar la producción de ácido shikímico mediante el uso de la información obtenida por la modelación matemática.

---

## Capítulo 3

# MATERIALES, MÉTODOS y MODELOS

## 3.1. MATERIALES Y MÉTODOS

### 3.1.1. Diseño experimental

Para realizar el estudio de caracterización fisiológica y metabólica, se utilizó una matriz experimental simétrica para las variables de los substratos glucosa y extracto de levadura a 3 niveles. El rango experimental utilizado fue de 75 a 125 g/L de glucosa y de 15 a 45 g/L del extracto de levadura. Esto genera una matriz de 9 puntos con el punto central siendo la condición inicial de 100 g/L de glucosa y 30 g/L de extracto de levadura de acuerdo con el Cuadro 3.1. Esta zona experimental fue diseñada tomando como centro el punto experimental donde fue encontrada la mejor condición de producción de ácido shikímico en la investigación de Rodríguez y colaboradores. (Rodriguez et al., 2013). De esta forma la matriz nos permitirá explorar condiciones donde la cepa AR36 tenga mejor producción, así como estudiar la respuesta metabólica de la cepa ante los cambios de concentración de las fuentes de carbono.

Cuadro 3.1: Puntos experimentales para la matriz experimental diseñada para la caracterización fisiológica y metabólica de la AR36 en variaciones de sustrato.

Experimento	Glucosa (GLC)	E. Levadura (YE)
1	75	15
2	75	30
3	75	45
4	100	15
5*	100	30
6	100	45
7	125	15
8	125	30
9	125	45

\*Punto central realizado por triplicado.

Los experimentos de la matriz experimental fueron realizados de acuerdo con la subsiguiente sección, midiendo las respuestas de biomasa, consumo de glucosa y producción de ácido shikímico y acético. Los datos obtenidos permitieron el cálculo de los siguientes parámetros de fermentación: velocidad de crecimiento, biomasa máxima, concentración final de shikimato y acetato, glucosa consumida, productividades volumétricas globales, productividades específicas de fase exponencial y estacionaria

y velocidades de consumo específicas de fase exponencial y estacionaria. Con estos parámetros fueron construidas superficies de respuesta para la caracterización del comportamiento de fermentación, así como para develar el metabolismo celular de la cepa, en las diferentes composiciones de glucosa y extracto de levadura como se describirá en los siguientes apartados.

### 3.1.2. Cepas, cultivos y procedimientos analíticos

La cepa *E. coli* AR36 construida por Rodríguez y colaboradores (2013) (Rodríguez et al., 2013) fue utilizada para todos los experimentos y cálculos. Esta cepa es auxótrofa a aminoácidos aromáticos debido a la inactivación de los genes *aroK* y *aroL*, por lo que extracto de levadura (YE) (BD Bacto) y GLC (Fermentas) fueron utilizados como fuente de carbono y nitrógeno (Rodríguez et al., 2013). Para los experimentos de fermentación se utilizaron reactores Applikon con controladores ADI 1010 y consolas ADI 1025 y jarras con volumen máximo de 1 L, usando 0.5 L de volumen de trabajo. El medio de cultivo base tiene la composición mostrada en el Cuadro 3.2. Las concentraciones de glucosa y el extracto de levadura fueron ajustados en cada experimento de acuerdo con las condiciones deseadas de cultivo. El medio de cultivo de los fermentadores se prepara disolviendo  $K_2HPO_4$ ,  $KH_2PO_4$ , Ac. Cítrico y Citrato de Amonio Fe(III) hasta un volumen de 310 ml con agua bidestilada desionizada, se le agregan 600  $\mu L$  de  $H_2SO_4$  concentrado y se neutraliza con  $NH_4OH$  10 N después de disolver. Este medio se coloca en el fermentador armado con los electrodos de pH y de tensión de oxígeno y se esteriliza por 20 minutos a 121 centígrados. Las demás sales inorgánicas son adicionadas en soluciones 100 X esterilizadas previamente por filtrado, al igual que la tiamina, tetraciclina y la betaína. La glucosa es preparada por separado a 400 g/L y esterilizada por filtración para los experimentos. Todos los componentes del medio son adicionados al fermentador en condiciones de esterilidad justo antes de realizar la inoculación, llevando el volumen a 500 mL con la adición de agua bidestilada desionizada estéril.

Los inóculos fueron hechos en matraz de 500 mL con 50 mL de volumen de trabajo con composición de medio base igual a la utilizada en los reactores, pero con 15 g/L de levadura y 25 g/L de glucosa. Los matraces fueron puestos en incubadora a 300

rpm y 37° C durante 12 horas para que crecieran alrededor de 7 D.O. (600 nm). Se inocularon 25 mL del inoculo a los reactores para obtener una concentración en el reactor de alrededor 0.3 D.O.

Cuadro 3.2: Composición del medio de cultivo base empleado en los experimentos

Compuesto	Concentración final
K <sub>2</sub> HPO <sub>4</sub>	7.5 g/L
KH <sub>2</sub> PO <sub>4</sub>	7.5 g/L
Ac. Cítrico	2.1 g/L
Citrato de Amonio Fe(III)	0.3 g/L
MgSO <sub>4</sub>	0.64 g/L
CaCl <sub>2</sub>	0.06 g/L
(NH <sub>4</sub> ) <sub>6</sub> (Mo <sub>7</sub> O <sub>24</sub> )	0.0037 g/L
ZnSO <sub>4</sub>	0.0029 g/L
H <sub>3</sub> BO <sub>3</sub>	0.0247 g/L
CuSO <sub>4</sub>	0.0025 g/L
MnCl <sub>2</sub>	0.0158 g/L
CoCl <sub>2</sub>	0.00129 g/L
Tiamina	1 mg/L
Tetraciclina	30 ug/ml
Betaína	2mM

Los reactores fueron operados a 37 °C, 1 vvm de aireación, manteniendo la D.O.T. mayor al 20 % por medio de cascadas de agitación de forma manual que variaron entre los 500 y 1200 rpm. El pH en los experimentos realizados fue mantenido en 7 por medio de la adición de base(NH<sub>4</sub>OH al 10 % ) y ácido (H<sub>3</sub>PO<sub>4</sub> al 3.5 %). Los puntos muestrales fueron tomados cada 2 horas durante las primeras 12 horas, cada cuatro horas desde la hora 12 hasta la hora 24 y finalmente cada 6 horas de la hora 24 hasta el fin de la fermentación. La concentración de biomasa fue medida por medio de la densidad óptica a 600 nm con un espectrofotómetro Beckman DU700. La GLC, el SA y el AC fueron determinado por HPLC con un equipo Waters (600E bomba cuaternaria, inyector automático 717, detector de índice de refracción 2410 y un detector de arreglo de fotodiodos 966) con una columna aminex HPX-87H (300 x 7.8 mm; 9 μm), usando H<sub>2</sub>SO<sub>4</sub> 5 mM como fase móvil a 50°C. Los parámetros medidos fueron corregidos volumétricamente con adición de ácido o base. Las muestras tomadas fueron de 3 mL en viales Eppendorf plásticos con tapa de 1.5 mL donde inmediatamente después de la toma de muestra uno de ellos fue centrifugado durante 2 minutos, decantado y congelado para análisis de sobrenadante por HPLC. Las muestras de sobrenadante fueron diluidas en la fase móvil del HPLC 0.5mM de



H<sub>2</sub>SO<sub>4</sub> para su análisis. La otra muestra se utilizó para la corrección de la medición de la densidad óptica; se centrifuga durante 2 minutos, decanta y utiliza para la medición de D.O. a 600 nm y realizar así la corrección del medio de cultivo.

## 3.2. MODELOS MATEMÁTICOS

### 3.2.1. Diseño y construcción de modelos para describir el comportamiento fisiológico en variaciones de las concentraciones de los dos substratos utilizados.

Para poder caracterizar y describir el comportamiento fisiológico de la cepa AR36 dentro del área de diseño experimental, se construyó un sistema de modelos extracelulares para caracterizar el comportamiento del crecimiento celular, el consumo de glucosa y la producción de shikimato. De esta forma para la descripción del crecimiento celular se utilizó un modelo logístico clásico, del cual se establecieron los parámetros de velocidad máxima de crecimiento y la máxima biomasa producida, de acuerdo a las ecuaciones siguientes:

$$\frac{dX}{dt} = \mu_{max} X \left( 1 - \frac{X}{X_{max}} \right) \quad (3.1)$$

$$\Rightarrow X(t) = \frac{X_0 e^{\mu_{max} t}}{1 - \left( \frac{X_0}{X_{max}} (1 - e^{\mu_{max} t}) \right)} \quad (3.2)$$

donde  $X(t)$  es la biomasa calculada al tiempo  $t$  por el modelo de crecimiento logístico,  $X_0$  es la biomasa inicial,  $X_{max}$  es el parámetro de biomasa máxima y  $\mu_{max}$  es el parámetro de velocidad máxima de crecimiento. Este modelo permite seguir la concentración de la biomasa a través de la fase exponencial de crecimiento, la fase de desaceleración y la fase estacionaria. Este modelo es comúnmente usado en la descripción de procesos de fermentación tanto en investigación como en la industria biotecnológica, donde ha demostrado ser robusto y eficiente para diferentes microorganismos, cepas y procesos. Estas características permiten tener una buena base para la construcción de los modelos de consumo y producción de metabolitos. Dicha construcción inició mediante la representación simplificada del proceso de fermentación

por medio de una reacción de primer orden:



donde  $S$  se refiere a substrato, en este caso GLC, y  $P$  se refiere a producto, SA en este diseño experimental. Por lo que las velocidades de aparición de producto y desaparición de substrato pueden ser descritas como:

$$\frac{dS}{dt} = -q_s \frac{dX}{dt} \quad (3.4)$$

$$\frac{dP}{dt} = q_p \frac{dX}{dt} \quad (3.5)$$

donde  $q$  se refiere a la velocidad específica para cada uno de estos, como marca el subíndice. Dado que en el fermentador se encontrarán poblaciones de células en crecimiento y poblaciones en estado estacionario, reescribimos las ecuaciones de la forma siguiente:

$$\frac{dS}{dt} = -q_s^{exp} \frac{dX^{exp}}{dt} - q_s^{sta} \frac{dX^{sta}}{dt} \quad (3.6)$$

$$\frac{dP}{dt} = q_p^{exp} \frac{dX^{exp}}{dt} + q_p^{sta} \frac{dX^{sta}}{dt} \quad (3.7)$$

donde  $\frac{dX^{exp}}{dt}$  está relacionada con la aparición de nuevas células en estado de crecimiento y  $\frac{dX^{sta}}{dt}$  con la aparición de células que han dejado de crecer. Estas diferenciales se pueden describir con respecto a  $X$  total de forma tal que se construyeron las siguientes ecuaciones:

$$\frac{dX_{exp}}{dt} = \frac{dX}{dt} \Psi^{exp} \quad (3.8)$$

$$\frac{dX_{sta}}{dt} = \frac{dX}{dt} \Psi^{sta} \quad (3.9)$$

donde  $\Psi^{exp}$  y  $\Psi^{sta}$  son funciones que describen la proporción de células en cada estado de acuerdo con las siguientes ecuaciones:

$$\Psi^{exp} = \left(1 - \frac{X}{X_{max}}\right) \quad (3.10)$$

$$\Psi^{sta} = \left(\frac{X}{X_{max}}\right) \quad (3.11)$$

De esta forma integrando e intercambiando los subíndices por los substratos y productos utilizados en esta investigación, llegamos a las siguientes ecuaciones construidas para describir el comportamiento de substratos y productos en nuestro diseño experimental:

$$S_{(t)} = S_{(t-1)} - \left[ (q_{glc}^{exp} X_{(t)}(\Delta t)) \left(1 - \frac{X_{(t)}}{X_{max}}\right) \right] - \left[ (q_{glc}^{sta} X_{(t)}(\Delta t)) \left(\frac{X_{(t)}}{X_{max}}\right) \right] \quad (3.12)$$

$$P_{(t)} = P_{(t-1)} + \left[ (q_{sa}^{exp} X_{(t)}(\Delta t)) \left(1 - \frac{X_{(t)}}{X_{max}}\right) \right] + \left[ (q_{sa}^{sta} X_{(t)}(\Delta t)) \left(\frac{X_{(t)}}{X_{max}}\right) \right] \quad (3.13)$$

donde  $q_{glc}^{exp}$  y  $q_{sa}^{exp}$  son las velocidades específicas en la fase exponencial para el consumo de GLC y la producción de SA respectivamente.  $q_{glc}^{sta}$  y  $q_{sa}^{sta}$  son las velocidades específicas en la fase estacionaria para el consumo de GLC y la producción de SA respectivamente. Se observa que en estas ecuaciones la participación de cada una de estas velocidades específicas está determinada por el término que contiene el radio de relación entre la biomasa existente al tiempo  $t$  y la biomasa máxima  $X_{max}$ , obtenidas de  $\Psi^{exp}$  y  $\Psi^{sta}$ . Finalmente, para determinar los límites del modelo se utilizaron restricciones basadas en desigualdades las cuales se presentan a continuación:

$$\frac{dS}{dt} = 0 \quad ; \quad S_{(t)} \leq 0 \quad (3.14)$$

$$\frac{dP}{dt} = 0 \quad ; \quad P_{(t)} \geq P_{max} \quad (3.15)$$

Los parámetros para los modelos construidos en esta investigación fueron aproximados por medio de la minimización de la suma cuadrática del error (SEE) por medio de programación en MATLAB con respecto a los datos experimentales. Para todos los modelos el escalón de tiempo  $\Delta t$  para la integración numérica fue fijado a 1 minuto. Una vez optimizados los parámetros se evaluaron por medio de la estimación

del error calculado por la relación entre el SSE y la suma del cuadrado de los puntos experimentales (SSEP) de acuerdo con la ecuación  $\%ERR = \left(\frac{SSE}{SSEP}\right) 100$ . Además, la aproximación matemática fue descrita como una regresión lineal entre los valores experimentales y los obtenidos del modelo. Esto permitió la construcción de un indicador de porcentaje de desviación con respecto a la pendiente esperada (1 para la equivalencia entre el modelo y los experimentos), un coeficiente de regresión de Pearson (describiendo la dispersión) y un parámetro de significancia determinado por el p-valor, los cuales permitieron calificar la aceptación de los modelos como descriptores del comportamiento experimental observado.

Es importante notar que el AC presento un comportamiento distinto al SA, describiendo un patrón de consumo y producción simultánea a lo largo de la fermentación. Por este motivo no pudo ser descrito por las ecuaciones previamente presentadas. Sin embargo, velocidades específicas iniciales en la fase exponencial ( $q_b^{exp}$ ) y velocidades específicas iniciales en la fase estacionaria ( $q_b^{sta}$ ), fueron aproximadas por las siguientes ecuaciones:

$$q_b^{exp} = Y_{ac/x}^{exp} \mu_{max} \quad (3.16)$$

$$q_b^{sta} = \frac{Q_b^{sta}}{X_{max}} \quad (3.17)$$

donde el rendimiento fue calculado con una regresión lineal del AC contra la biomasa y la velocidad volumétrica fue calculada con la regresión lineal del acetato contra el tiempo en la fase estacionaria.

### 3.2.2. Construcción y validación de superficies de respuesta

Los parámetros obtenidos de la modelación fueron utilizados para construir superficies de respuesta tridimensionales con una ecuación de segundo orden polinomial bivariada:

$$Z = a + b * (GLC) + c * (YE) + d * (GLC)^2 + e * (YE)^2 + f * (GLC) * (YE) \quad (3.18)$$

Otras ecuaciones polinomiales de mayores ordenes fueron probadas, sin embargo, esta fue la del tamaño mínimo encontrada capaz de describir el comportamiento de los parámetros a través del área del diseño experimental. A pesar de que ecuaciones polinomiales de mayor orden pueden dar resultados con valores de SSE más pequeños, su uso podría llevar a una sobreestimación de la morfología de superficie y por ende limitar sus capacidades predictivas. Por este motivo, y resultando suficiente, la de menor grado fue utilizada, su evaluación y caracterización se describirá en las siguientes secciones.

La aproximación no lineal de las superficies fue realizada en MATLAB por medio de la reducción del SSE. Las ecuaciones de las superficies de respuesta resultantes fueron analizadas cualitativamente por medio de la obtención de gráficas tridimensionales y mapas de contorno. Así mismo, evaluadas matemáticamente por medio del criterio de LaGrange. El criterio de LaGrange permite la definición de la naturaleza geométrica de los puntos críticos observados. Este criterio consiste en el cálculo de la determinante de Hessian para la función modelo de la superficie  $R$  dada por:

$$H = \begin{bmatrix} \frac{\delta^2 Z}{\delta GLC^2} & \frac{\delta^2 Z}{\delta GLC \delta YE} \\ \frac{\delta^2 Z}{\delta YE \delta GLC} & \frac{\delta^2 Z}{\delta YE^2} \end{bmatrix}$$

$$H = \left( \frac{\delta^2 Z}{\delta GLC^2} \right) \left( \frac{\delta^2 Z}{\delta YE^2} \right) - \left( \frac{\delta^2 Z}{\delta GLC \delta YE} \right) \left( \frac{\delta^2 Z}{\delta YE \delta GLC} \right)$$

donde esta los puntos críticos se obtienen de la solución del sistema de ecuaciones resultante de las derivadas de  $Z$  con respecto a las diferentes variables igualadas a 0 y se pueden evalúan de la siguiente forma: máximo si  $H(GLC, YE) > 0$  y  $\delta^2 Z / \delta GLC^2 < 0$ , mínimo si  $H(GLC, YE) > 0$  y  $\delta^2 Z / \delta GLC^2 > 0$  y punto de silla si  $H(GLC, YE) < 0$  (Ferreira et al., 2004).

Finalmente, las superficies fueron evaluadas por coeficientes de regresión, valores  $p$ , SSE y porcentajes de error. Su capacidad de descripción y predicción del comportamiento fue validada con la predicción del comportamiento de puntos experimentales en condiciones iniciales de 75:20, 80:40 y 115:45 GLC:YE. Los parámetros obtenidos de las superficies de respuesta para estas condiciones fueron utilizados para el cálculo de los modelos logísticos de crecimiento, consumo y producción y comparados con los puntos experimentales. Además, se calcularon los parámetros derivados de los experimentos y comparados con los estimados por las superficies mediante porcentaje de error.

### **3.2.3. Diseño y construcción de modelos para describir el comportamiento de las distribuciones de flujos metabólicos en variaciones de la concentración de los dos substratos utilizados.**

Para caracterizar y describir el comportamiento metabólico de la cepa AR36 dentro del área de diseño experimental, se construyó una red metabólica basada en el metabolismo central (CCM) de *Escherichia coli*. La red metabólica del CCM fue conformada por 60 reacciones, 44 metabolitos internos y 6 metabolitos externos, correspondientes a las vías Embden-Meyerhoff-Parnas (EMP), la vía de las pentosas fosfato (PPP), ciclo de los ácidos tricarboxílicos (TCA), el metabolismo del piruvato, las reacciones anapleróticas, reacciones de respiración y metabolismo energético, reacciones de consumo de YE, reacciones biosintéticas de SA y reacciones de generación de biomasa. Los metabolitos externos definidos fueron AC, GLC, SA, YE, biomasa y el mantenimiento. El consumo de YE fue introducido a la red metabólica como un solo metabolito y su consumo deriva en un precursor de biomasa (BIOMp), aminoácidos aromáticos (tomados como un metabolito unificado), alanina (ALA) y ácido glutámico (GLU). Los valores estequiométricos para la conversión de YE a estos metabolitos fueron estimados del promedio de composición descrita por el fabricante. La representación simplificada de la red metabólica utilizada en este trabajo se puede observar en la Figura 3.1, los nombres de las reacciones serán referidos de ahora en adelante como se indican en esta figura. La representación completa de

todas las reacciones incluidas en el modelo se muestra en el Cuadro 3.3.

Cuadro 3.3: Definición de las reacciones de la red metabólica utilizada para la modelación

EMP	
GalP	GLC + HEXT => GLC <sub>i</sub>
Glk	GLC <sub>i</sub> + ATP => G6P
Pgi	G6P <=> F6P
Pfk	F6P + ATP => FDP
Fba	FDP <=> T3P1 + T3P2
TpiA	T3P2 <=> T3P1
GapA	T3P1 <=> NADH + 13PDG
Pgk	13PDG <=> 3PG + ATP
Pgm	3PG <=> 2PG
Eno	2PG <=> PEP
PykA	PEP => PYR + ATP
PpsA	PYR + 2 ATP => PEP
LpdA	PYR => ACCOA + NADH
PPP	
G6Pdh	G6P <=> D6PGL + NADPH
Pgl	D6PGL => D6PGC
Gnd	D6PGC => NADPH + RL5P
Rpi	RL5P <=> R5P
Rpe	RL5P <=> X5P
TktA1	R5P + X5P <=> T3P1 + S7P
TktA2	F6P + T3P1 <=> X5P + E4P
TalA	T3P1 + S7P <=> E4P + F6P
PGdh	D6PGC => 2KD6PG
KDPGa	2KD6PG => T3P1 + PYR
TCA	
Csyn	ACCOA + OA => CIT
Acn	CIT => ICIT
IcdA	ICIT <=> NADPH + AKG
KGdh	AKG => NADH + SUCCOA
SucCD	SUCCOA <=> ATP + SUCC
SdhABCD	SUCC => FADH + FUM
Fum	FUM <=> MAL
Mdh	MAL <=> NADH + OA
Metabolismo de piruvato	
Pta	ACCOA <=> ACTP
AckA	ACTP <=> ATP + ACin
Acs	2 ATP + ACin <=> ACCOA
ActPin	AC => ACin
ActPout	ACin => AC
Reacciones anapleróticas	
PckA	OA + ATP => PEP
Ppc	PEP => OA
MaeB	MAL => NADPH + PYR
MaeA	MAL => NADH + PYR
Icl	ICIT => GLX + SUCC
Msn	ACCOA + GLX => MAL
Energía y respiración	
Ndh	NADH => QH2
Nuo	NADH => QH2 + 3.5 HEXT
PoxB	PYR => ACin + QH2
Cyo	QH2 => 2 HEXT
Sdh	FADH <=> QH2
PntA	NADPH => NADH
PntB	NADH + 2 HEXT => NADPH
AtpABCD	4 HEXT <=> ATP
Catabolismo de YE	
YEa	YE + HEXT => 0.51 BIOM <sub>p</sub> + 0.25 AAA + 0.63 GLU + 0.62 ALA
Gdh	AKG + NADPH <=> GLU
DadA	ALA => PYR + FADH
Biosíntesis de SA	
DAHPS	E4P + PEP => DAHP
DHQd	DAHP => DHQ
DHSS	DHQ => DHS
SAdh	DHS + NADPH => SA

Biomasa y mantenimiento :

BIOMSp 1.496 3PG + 3.7478 ACCOA + 59.8100 ATP + 0.0709 F6P + 0.1290 T3P1 + 0.2050 G6P + 5.1464 GLU + 13.0279 NADPH + 1.7867 OA + 0.1581 PEP + 2.8328 PYR + 0.8977 R5P => BIOMp + 4.1 AKG + 3.5 NADH BIOMS BIOMp + 0.3610 AAA => BIOM

ATPdrain ATP => MAINT

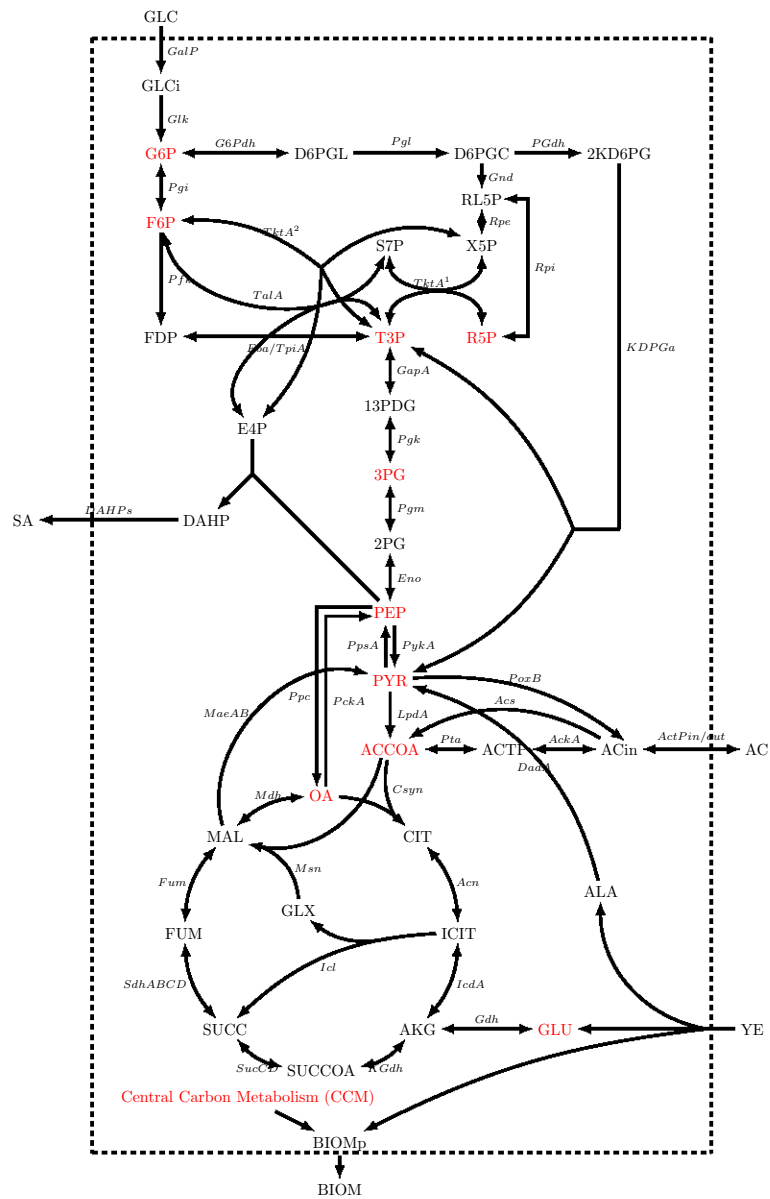


Figura 3.1: Mapa simplificado de las reacciones del metabolismo central de carbono tomadas en cuenta para el modelo cibernético y el análisis de flujos.

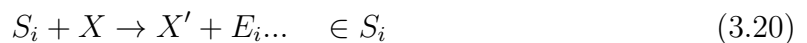


Con esta red metabólica se realizó el cálculo y extracción de los EMs, por medio del protocolo de efmtool (Terzer y Stelling, 2008) embebido en un programa de MATLAB. El conjunto de EMs fue después reducido mediante el protocolo de análisis de rendimientos usando los parámetros obtenidos del punto central como describe Song et al. (Song y Ramkrishna, 2009). Para estos efectos se construyeron dos familias de análisis, la primera contendría los EMs preferidos en fase exponencial al restringir la selección a aquellos EMs que consumieran simultáneamente GLC y YE a su vez que produjeran SA. La segunda familia de EMs contendría las restricciones de aquellos que sólo consumieran GLC y produjeran SA sin consumir YE, y por lo tanto deberían ser preferidos en la fase estacionaria de las fermentaciones. La selección de los EMs fue realizada para cada familia fue realizada con los rendimientos de biomasa, shikimato y acetato con respecto a la glucosa calculados por medio de la relación de las velocidades específicas de cada uno de estos para el punto central experimental. De esta forma se seleccionaron 9 modos elementales, 6 para la familia de fase exponencial y 3 para la familia de la fase estacionaria.

Una vez seleccionados los EMs, se construyó un conjunto de modelos metabólicos dinámicos basados en el marco cibernético publicado por Ramkrishna y colaboradores como se describe a continuación (Kompala et al., 1986; Ramkrishna y Song, 2012, 2016; Varner y Ramkrishna, 1999). Este marco inicia con el supuesto de que cada modo elemental puede ser descrito de forma simple como la asimilación de un substrato o un conjunto de substratos ( $S_i$ ) catalizados por una enzima crítica ( $E_i$ ) para la producción de biomasa y/o otros productos.  $E_i$  representa el conjunto total de enzimas que catalizan reacciones en dicho modo elemental. De forma que se puede escribir:



De la misma forma que la biomasa la enzima  $E_i$  es catalizada para su síntesis por la presencia del substrato específico  $S_i$ , lo que se puede expresar como:



Donde  $X'$  representa la biomasa excluyendo la enzima crítica  $E_i$ . Estas dos reacciones pueden describirse cinéticamente por medio de ecuaciones tipo Michaelis-Menten, utilizadas generalmente para reacciones de catálisis enzimática. Por lo tanto, se escriben de la siguiente forma de manera general:

$$r_i = \frac{k_i e_i s_i X}{K_i + s_i} \quad (3.21)$$

$$r_{E_i} = \frac{\alpha s_i X}{K'_i + s_i} \quad (3.22)$$

donde  $r_i$  es la velocidad de la reacción enzimática y  $r_{E_i}$  es la velocidad de la producción de enzima. El parámetro  $e_i$  es la concentración específica de la enzima  $E_i$  de forma que  $e_i X$  es la concentración de esta.  $\alpha$  es la velocidad de síntesis enzimática máxima y  $k_i e_i$  substituye a la velocidad de flujo máxima del modelo clásico de Michaelis-Menten. Estas velocidades máximas de reacción por su parte, son modificadas en el metabolismo por medio de procesos regulatorios de inhibición/activación y represión/inducción, por lo que se introducen en este marco las variables cibernéticas  $u$  y  $v$  de forma que la velocidad real de síntesis de enzima  $E_i$  se escribe como :

$$r_{E_i} u_i \quad (0 < u_i < 1 ; \sum_i u_i = 1) \quad (3.23)$$

Además, introduciendo el efecto del crecimiento, disolución y decaimiento de la proteína la velocidad para  $e_i$  se puede entonces describir como:

$$\frac{de_i}{dt} = \frac{\alpha_i s_i}{K'_i + s_i} u_i + \frac{d}{dt}(\ln X) e_i - \beta_i e_i \quad (3.24)$$

donde  $\alpha_i$  y  $\beta_i$  son constantes de producción y degradación enzimática conocidas para *Escherichia coli* (Kompala et al., 1986; Ramkrishna y Song, 2012). Dado que es prácticamente imposible medir las concentraciones de las enzimas  $e_i$ , además de ser una enzima crítica representativa, se realiza la siguiente inferencia, dado que podemos

definir una velocidad máxima en relación a una cantidad máxima de enzima como:

$$k_i^{max} = k_i e_i^{max} \quad (3.25)$$

$$e_i^{max} = \frac{\alpha_i}{k_i^{max} + \beta_i} \quad (3.26)$$

Se puede deducir que para la modelación dinámica el valor real de la enzima no es relevante sino su valor relativo, dado que:

$$k_i e_i = k_i^{max} = \left[ \frac{e_i}{e_i^{max}} \right] \quad (3.27)$$

Finalmente, de la misma forma que para la síntesis de enzima, tanto los substratos como los productos derivados del flujo de cada EMs ( $m$ ) son sujetos a control de actividad enzimática por medio de la variable cibernética  $v$  de forma que:

$$\frac{dm_i}{dt} = \sum r_i v_i \quad (0 \leq v_i \leq 1) \quad (3.28)$$

Las variables cibernéticas  $u$  y  $v$  son calculadas a través del tiempo por medio de leyes de igualación para la comparación de “retornos” comparados con una función objetivo, ya sea el crecimiento, el consumo de carbono, el consumo de oxígeno entre otros. Donde  $R_i$  representa el retorno de cada EMs alternativo con respecto a la función objetivo. Las ecuaciones que gobiernan dichas variables se escriben de la siguiente forma:

$$u_i = \frac{R_i}{\sum_j R_j} \quad (3.29)$$

$$v_i = \frac{R_i}{\max_j(R_j)} \quad (3.30)$$

De esta forma se puede introducir este valor para los cálculos de las velocidades de flujo y de síntesis enzimática. Además de la estructura del modelo construido por Ramkrishna y colaboradores (Kompala et al., 1986; Ramkrishna y Song, 2012, 2016; Varner y Ramkrishna, 1999) y previamente detallada, en esta investigación la ecuación de velocidad de flujo de los diferentes EMs fueron descritos por ecuaciones diferenciales modificadas del tipo Michaelis-Mentem. Estas ecuaciones de velocidad

fueron diseñadas para acoplar cada familia a su fase de la fermentación por medio de la inclusión de términos similares al de las ecuaciones integradas de los modelos fisiológicos  $\Psi^{exp}$  y  $\Psi^{sta}$ . Además, se les agregó un término para tomar en cuenta la inhibición por AC. Las ecuaciones utilizadas se presentan a continuación:

$$r_i^M = \left( \frac{k_{max,i}^M [GLC]_{(t)}}{K_{m,i}^M + [GLC]_{(t)}} \right) \left( 1 + \frac{[AC]_{(t)}}{K_I^{ac}} \right)^{-1} \left( 1 - \frac{X_{(t)}}{X_{max}} \right) \quad (3.31)$$

$$r_i^G = \left( \frac{k_{max,i}^G [GLC]_{(t)}}{K_{m,i}^G + [GLC]_{(t)}} \right) \left( 1 + \frac{[AC]_{(t)}}{K_I^{ac}} \right)^{-1} \left( \frac{X_{(t)}}{X_{max}} \right) \quad (3.32)$$

donde los índices  $M$  y  $G$  se refieren a los EMs de la familia exponencial (consumo mixto) y a la familia estacionaria (consumo sólo de GLC) respectivamente. El índice  $i$  se refiere a cada uno de los  $i^s$  EM de cada una de las familias;  $K_{max}$  y  $K_m$  son los parámetros de Michaelis-Menten para cada ecuación,  $K_I^{ac}$  es la constante de inhibición y  $X_{max}$  se refiere la máxima biomasa calculada del modelo logístico de la caracterización fisiológica para cada uno de los puntos del diseño experimental.  $[GLC]_{(t)}$ ,  $[AC]_{(t)}$ ,  $X_{(t)}$  se refieren a las concentraciones de GLC, AC y biomasa en cada tiempo  $t$  en  $mmol/L$  para los primeros dos y en  $g/L$  para la biomasa. La concentración de enzima relativa inicial fue asignada como 0.95 para el primer EM, y 0.5 para los subsiguientes EMs de la primera familia y 0.1 para la los EMs de la segunda familia. Todos los otros parámetros del modelo, como velocidades de producción y degradación enzimática entre otras, fueron utilizadas como se describe por Ramkrishna et al. (Kompala et al., 1986; Ramkrishna y Song, 2012; Song y Ramkrishna, 2012). Los valores de los parámetros  $K_{max}$  y  $K_m$  fueron aproximados mediante el uso de un algoritmo genético.

Las distribuciones de flujo fueron usadas para construir superficies de respuesta tridimensionales usando ecuaciones polinomiales de segundo orden bivariadas para diferentes puntos de la fermentación. Las superficies fueron calificadas de acuerdo con las técnicas descritas anteriormente y utilizadas para analizar el comportamiento del metabolismo de la cepa durante la producción de SA. Durante el análisis de rendimientos para la selección de EMs, se estudió por separado la distribución del porcentaje de flujo sobre el nodo de la Fosfoglucosa isomerasa (Pgi)/Glucosa 6-

fosfato-1-deshidrogenasa (G6Pdh). Los conjuntos fueron evaluados por regresiones lineales entre los valores experimentales y los valores modelados para cada uno de los metabolitos externos con respecto al porcentaje de desviación de la pendiente, el porcentaje de error y los valores de los coeficientes regresión de Pearson y sus p-valores. El conjunto con la mejor regresión fue el utilizado con 0.5 de rendimiento. Los conjuntos de proporciones superiores resultaron en la selección de EMs que elevaban el SSE y en su inspección mostró una baja adecuación a los datos experimentales. Por estos motivos se utilizaron los EMs del set de 0.5 para los análisis posteriores.

### **3.3. PROCESO DE PRODUCCIÓN DE ÁCIDO SHIKÍMICO**

#### **3.3.1. Diseño de un bioproceso mejorado para la producción de ácido shikímico**

Para evaluar la utilidad de los modelos y las superficies de respuestas construidas, un proceso fermentativo fue diseñado con la información adquirida de los flujos y la caracterización fisiológica para incrementar la productividad de SA. El proceso fue diseñado para mantener constantes las concentraciones iniciales de GLC y YE. Esto en teoría mantendrían constantes las variables cibernéticas (las cuales son una representación de parámetros de regulación) y por lo tanto se mantendrían constantes las distribuciones de flujos y su evolución dinámica. Las condiciones iniciales escogidas fueron 80:45 g/L GLC:YE. El proceso fue operado en modo lote donde la alimentación fue regulada mediante la manipulación manual cada 15 minutos de una bomba peristáltica de acuerdo con las necesidades de alimentación calculadas con los valores de velocidad de crecimiento y consumo. La alimentación consistió en dos soluciones añadidas simultáneamente con 400 g/L GLC y amortiguador de fosfatos adicionado con 400 g/L YE. Después de 12 horas de alimentación fue detenida y la fermentación entró a fase estacionaria para consumir la GLC restante. Oxígeno fue añadido según lo necesario para mantener la DOT sobre 20% en conjunto con las cascadas de agitación. Los otros componentes del medio, como sales, y antibióticas fueron adicionados a las soluciones de alimentación para evitar su disolución.

---

## Capítulo 4

# RESULTADOS y DISCUSIÓN

## 4.1. RESULTADOS

### 4.1.1. Caracterización del comportamiento fisiológico de la cepa AR36 mediante la construcción y análisis de superficies de respuesta de los parámetros modelados en diferentes etapas de la fermentación.

La Figura 4.1 muestra los resultados para las nueve fermentaciones de la cepa AR36 realizadas en el diseño experimental, con el punto central realizado por triplicado. Los valores promedio y las desviaciones estándar para la condición central (100:30 GLC:YE g/L) se encuentran resumidas en el Cuadro 4.1. Las desviaciones estándar muestran valores relativamente bajos, teniendo en cuenta que se utilizaron experimentos con extracto de levadura (YE) de tres diferentes lotes. La desviación estándar más alta corresponde al SA final producido ( $[SA]_f$ ), la glucosa consumida final ( $\Delta[GLC]$ ) y la velocidad de consumo específica en fase exponencial ( $q_s^{exp}$ ). A pesar de esto, como se observa en la Figura 4.1, los modelos siguen en buena forma los datos experimentales tanto para el punto central como para todas las otras condiciones experimentales. El comportamiento observado y los datos estadísticos obtenidos para el seguimiento de estos modelos con respecto a los datos experimentales sugieren que los modelos son adecuados para describir y parametrizar el crecimiento, el consumo de glucosa y la producción de SA en la cepa AR36 como se muestra en el Cuadro 4.2.

Cuadro 4.1: Parámetros promedio y desviaciones estándar para el punto central experimental, 100 g/L GLC y 30 g/L YE.

	$X_{max}$	$\Delta[GLC]$	$[SA]_f$	$[AC]_f$	$\mu$	$q_{glc}^{exp}$	$q_{sa}^{exp}$	$q_{glc}^{sta}$	$q_{sa}^{sta}$	$Y_{ps}$	$Y_{px}$	$q_{ac}^{exp}$	$q_{ac}^{sta}$
	[g/L]	[g/L]	[g/L]	[g/L]	[h <sup>-1</sup> ]	[g/Lh]	[g/Lh]	[g/Lh]	[g/Lh]	[g/g]	[g/g]	[g/Lh]	[g/Lh]
$\bar{p}$	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
$\sigma^2$	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

De acuerdo con lo que se describe en el capítulo de materiales, modelos y métodos, las superficies de respuesta construidas para los parámetros se pueden observar en la Figura 4.2. La superficie de respuesta para la biomasa máxima ( $X_{max}$ ) presentada en la Figura 4.2 A, muestra un aumento de la cantidad de biomasa principalmente con respecto del aumento de la concentración de YE, y sólo pequeños aumentos con

incrementos de la concentración de GLC. Por su parte la superficie de respuesta para  $\Delta[\text{GLC}]$  (Figura 4.2 B) muestra que el consumo de GLC aumenta proporcionalmente con respecto a concentraciones iniciales más altas de GLC y YE. Esta observación es más evidente en concentraciones altas de GLC donde se puede establecer que son necesarios al menos  $\approx 40$  g/L de YE para consumir completamente concentraciones de GLC mayores a  $\approx 110$  g/L. Con respecto a la concentración final de SA ( $[\text{SA}]_f$ ), la morfología de su superficie es similar a la de  $\Delta[\text{GLC}]$ . Sin embargo, el análisis de LaGrange revela la aparición de un punto crítico máximo en la condición  $\approx 110:40$  g/L GLC:YE (Figura 4.2 C).

Cuadro 4.2: Valores estadísticos para la comparación de los perfiles experimentales con los resultados de los modelos

GLC	75	75	75	100	100	100	125	125	125
YE	15	30	45	15	30	45	15	30	45
Biomass %ERR	0.10	0.10	0.14	0.07	1.88	0.31	0.34	0.28	1.27
GLC %ERR	0.02	0.23	0.02	0.07	0.72	0.03	0.14	0.66	0.12
SA %ERR	0.26	0.09	0.19	0.41	3.58	0.02	0.43	0.08	0.91
Biomass %SPD	0.02	0.74	1.85	0.53	2.48	0.81	1.05	0.62	1.64
GLC %SPD	0.00	0.00	0.00	0.00	1.63	0.00	4.70	0.00	1.05
SA %SPD	0.63	0.26	0.00	0.93	0.75	0.00	3.42	0.00	6.33
Biomass $R^2$	0.996	0.997	0.998	0.997	0.950	0.993	0.988	0.992	0.962
GLC $R^2$	0.998	0.993	0.999	0.993	0.977	0.999	0.983	0.962	0.997
SA $R^2$	0.995	0.998	0.996	0.991	0.932	1.000	0.992	0.998	0.986
X $p$ - valor	3E-11	5E-09	3E-09	9E-12	1E-20	5E-10	3E-09	8E-09	2E-09
GLC $p$ - valor	8E-14	6E-09	1E-11	2E-12	2E-26	6E-13	2E-11	6E-08	4E-15
SA $p$ - valor	5E-12	7E-11	6E-10	1E-11	3E-19	4E-15	3E-13	1E-13	1.0E-11

Con respecto a los parámetros cinéticos, la velocidad de consumo de GLC en fase exponencial ( $q_{glc}^{exp}$ ) muestra una morfología de silla de montar con el punto crítico en  $\approx 96:37$  GLC:YE g/L (Figura 4.2 E). Este tipo de punto crítico está caracterizado por su condición de ser máximo para la variable de glucosa y simultáneamente mínimo para la variable YE. Estos resultados sugieren fuertemente la aparición de respuesta de control no lineales en el metabolismo celular de AR36, modificando la velocidad de consumo. De esta forma, en concentraciones de 75 g/L a 96g/L se incrementa el consumo y en concentraciones superiores a estas, se regula negativamente la velocidad de consumo, posiblemente derivado de una saturación metabólica en nodos glicolíticos importantes como PYR, PEP, FDP (Shimizu, 2014). Las superficies muestran que los valores más altos de  $q_{glc}^{exp}$  se encuentran en las condiciones



con concentraciones iniciales más bajas de YE y GLC. Mientras que los valores más bajos para este parámetro se encuentran en las condiciones de mayor concentración inicial de GLC. Por su parte, la superficie para la velocidad de producción específica de SA en fase exponencial ( $q_{sa}^{exp}$ ), muestra una tendencia a aumentar con respecto a condiciones menores de [YE] iniciales. En efecto, se observa para esta variable hasta un 50 % de disminución a valores mayores de 40 g/L de YE (Figura 4.2 E y F). Con respecto a la velocidad de crecimiento máxima ( $\mu_{max}$ ), el análisis de superficie mostró la existencia de un punto mínimo en las condiciones  $\approx 105:21$  g/L GLC:YE (Figura 4.2 K). Los valores más altos de velocidad de crecimiento se encontraron en la zona de menor [GLC] inicial en combinación con las [YE] más altas. Finalmente, la velocidad de producción de acetato en fase exponencial ( $q_{ac}^{exp}$ ) muestra una clara tendencia a tener valores más altos conforme las concentraciones iniciales de YE y GLC aumentan (Figura 4.2 I). Este aumento podría estar interrelacionado con la reducción de las velocidades de crecimiento y de producción de SA observadas en esta zona, específicamente en condiciones con concentraciones superiores a los  $\approx 40$  g/L [YE] y  $\approx 110$  g/L GLC. La morfología de todas las velocidades específicas en esta fase de crecimiento sugiere la localización de cuadrantes específicos de maximización de los diferentes productos de la fermentación, como se describe a continuación: en las zonas de altas [GLC] y [YE] iniciales, la producción de AC es predominante; en las condiciones de baja [GLC] y alta [YE] inicial, la producción de biomasa es predominante; a alta [GLC] pero con baja [YE] inicial, la producción de SA es predominante; finalmente, en condiciones bajas de ambos sustratos, el crecimiento y la producción de SA está más balanceado evitando también la producción de AC (Figura 4.2 E,F,K e I).

En la fase estacionaria, se observa una reducción generalizada de la actividad en todos los parámetros de velocidad de consumo y producción. La superficie para  $q_{glc}^{sta}$  (Figura 4.2 G) muestra la tendencia a presentar valores más altos en condiciones iniciales de sustrato más altas de los dos sustratos (GLC y YE). La velocidad de producción de SA en fase estacionaria ( $q_{sa}^{sta}$ ) presenta una superficie (Figura 4.2 H) que revela la tendencia a incrementar sus valores hacia concentraciones bajas de GLC combinadas con concentraciones iniciales altas de YE. Esta superficie muestra,

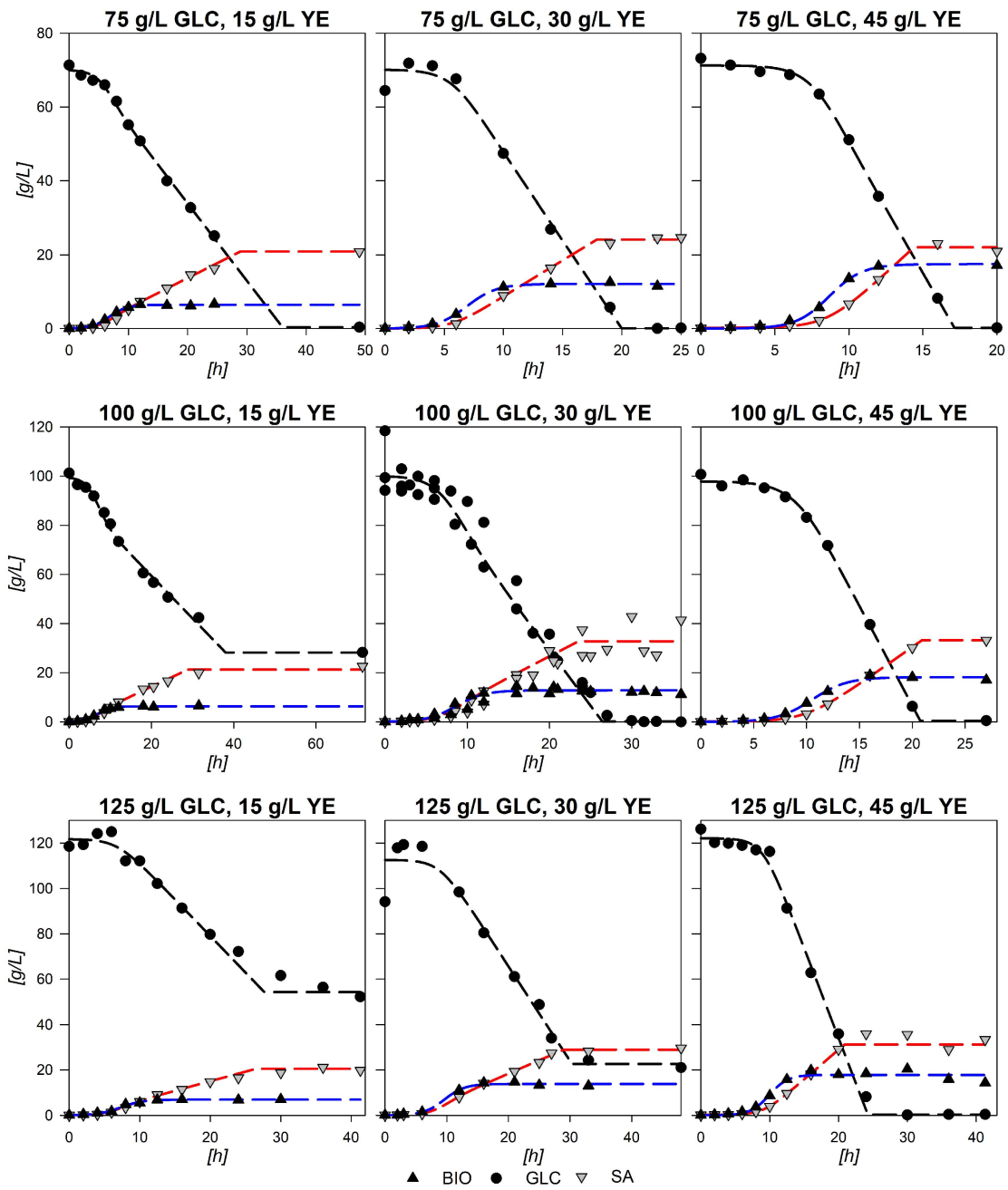


Figura 4.1: Resultados para las nueve condiciones de fermentación realizadas con la cepa AR36, aproximación de los modelos fisiológicos a los datos experimentales.

un comportamiento contrario a la superficie de la  $q_{ac}^{sta}$  (Figura 4.2 J). El análisis de las morfologías de superficie para estos parámetros muestra la aparición de regiones de predominancia para la formación de productos de la forma siguiente: la zona de preferencia de producción de SA se encuentra en la región superior a una línea imaginaria que corta el área del diseño experimental de bajas a altas concentraciones de ambos substratos. Por su parte la zona de preferencia de producción de AC se encuentra en el área debajo de esta misma diagonal. Es importante notar que las zonas de preferencia localizadas en esta fase se encuentran en lados opuestos con respecto a las zonas identificadas para las velocidades específicas en la fase exponencial.

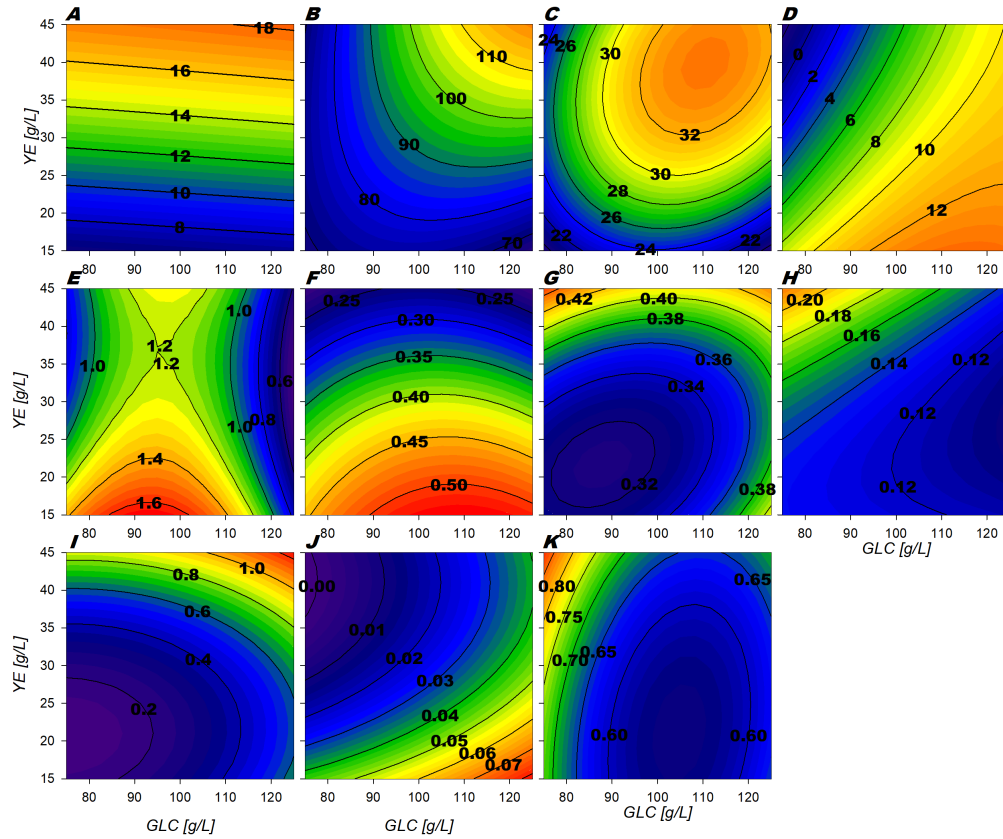


Figura 4.2: Contornos de las superficies de respuesta para los parámetros estimados por los modelos. A)  $X_{max}$  [g/L], B)  $\Delta$ [GLC] [g/L], C)  $[SA]_f$  [g/L], D)  $[AC]_f$  [g/L], E)  $q_{glc}^{exp}$  [g/Lh], F)  $q_{sa}^{exp}$  [g/Lh], G)  $q_{glc}^{sta}$  [g/Lh], H)  $q_{sa}^{sta}$  [g/Lh], I)  $q_{ac}^{exp}$  [g/Lh], J)  $q_{ac}^{sta}$  [g/Lh], K)  $\mu_{max}$  [ $h^{-1}$ ].

### 4.1.2. Validación de las superficies modeladas mediante la comparación de predicciones del comportamiento contra experimentos no contenidos en el diseño experimental.

La calidad de la capacidad descriptiva y predictiva de las superficies de respuesta construidas para caracterizar el comportamiento fisiológico de la cepa AR36, fue validada por medio de su comparación con el comportamiento experimental de esta en condiciones no incluidas en la matriz para el cálculo de los modelos. Las condiciones elegidas para la validación fueron 75:20, 80:40 y 115:45 GLC:YE g/L iniciales. La Figura 4.3 muestra los resultados obtenidos de los modelos logísticos para el crecimiento, el consumo de GLC y la producción de SA calculados con los parámetros obtenidos de las superficies para los parámetros  $X_{max}$ ,  $\mu$ ,  $q_{glc}^{exp}$ ,  $q_{sa}^{exp}$ ,  $q_{glc}^{sta}$ ,  $q_{sa}^{sta}$  y  $SA_{final}$ . Así mismo, se muestran en las mismas gráficas los puntos experimentales obtenidos para las condiciones de validación. Como se puede observar, todos los modelos siguieron adecuadamente el comportamiento experimental de las distintas fermentaciones realizadas. Las diferencias observables más grandes se encontraron en el máximo SA producido en la condición 80:40 GLC:YE, esto podría deberse a la contribución de YE para la producción de SA derivada del metabolismo de varios de sus componentes como aminoácidos y carbohidratos. Adicionalmente, se realizó la evaluación estadística de acuerdo con lo estipulado en el capítulo de materiales, modelos y métodos. El porcentaje de error calculado de la relación de SSE con el SSEP presentó valores de entre 0.14 a 0.91 para la biomasa, de 0.04 a 0.08 para la GLC y de 0.04 a 0.26 para el SA, sugiriendo desviaciones relativamente pequeñas entre los modelos y los puntos experimentales a lo largo de toda la fermentación. Los valores de  $R^2$  presentaron valores altos de entre 0.96 a 0.99 para las tres curvas, además de presentar desviaciones percentiles de la pendiente SDP menores al 1% en todos los casos con  $p$ -valores menores a 0.05. Estos valores estadísticos y la observación de los comportamientos que se muestran en la Figura 4.3 señalan que los modelos y las superficies construidas pueden describir y predecir el comportamiento de la cepa AR36 razonable y comprensivamente bien dentro del área del diseño experimental.

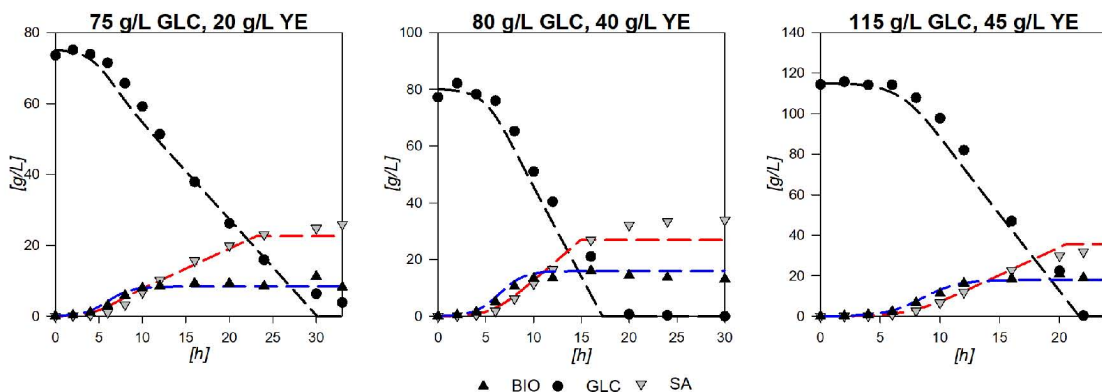


Figura 4.3: Predicción fisiológica de los modelos a tres experimentos de validación y puntos experimentales.

Los parámetros calculados de las superficies se compararon con los valores calculados directamente de los valores experimentales. En el Cuadro 4.3 se pueden observar todos los parámetros y el promedio del porcentaje de error calculado para los tres experimentos de validación analizados. El acetato presentó el error más alto, probablemente porque en los experimentos con concentraciones iniciales bajas de [GLC] y altas de [YE] no fue producido acetato en la fase estacionaria y las superficies calculadas de la ecuación polinomial no puede predecir adecuadamente el comportamiento limitado a estos valores. La  $q_{glc}^{exp}$  y por lo tanto el  $Y_{p/s}$  también mostraron errores superiores al 10%. Esto puede estar relacionado con la contribución del YE, ya que este no contiene solamente aminoácidos aromáticos, sino también otros aminoácidos y carbohidratos que pudieran contribuir a modificar el comportamiento y los efectos previamente observados. Sin embargo, se realizó una prueba estadística de *t*-Student para la comparación de los datos, utilizando la varianza obtenida para el punto central por triplicado. Los resultados indican que los valores comprobados fueron estadísticamente similares, rechazando la hipótesis de diferencia de medias con *p*-valores superiores a 0.05. Lo anterior indica que los valores de los parámetros predichos por las superficies se encuentran razonablemente en rango y pueden ser utilizados para comparar, caracterizar y estudiar el comportamiento de la cepa AR36 dentro de los límites del diseño experimental y que las inferencias a realizar en el análisis comparativo de las superficies podrán ser tomados como válidos.

Cuadro 4.3: Valores experimentales, modelados y error promedio para la validación de las superficies.

GLC:YE g/L	75:20		80:40		115:45		%Error Promedio
	Exp.	Mod.	Exp.	Mod.	Exp.	Mod.	
$X_{max}$ [g/L]	9.30	8.40	14.30	16.03	19.76	18.05	7.18
$\Delta$ GLC [g/L]	70.03	70.70	80.11	78.19	115.76	113.84	1.18
$[SA]_f$ [g/L]	24.62	22.13	33.19	26.62	31.84	32.96	7.88
$[AC]_f$ [g/L]	5.25	6.45	0.00	0.85	4.90	8.73	35.70
$\mu_{max}$ [h <sup>-1</sup> ]	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

### 4.1.3. Caracterización del comportamiento de las distribuciones de flujo mediante la construcción y análisis de superficies de respuesta de los flujos modelados en diferentes etapas de la fermentación.

Para obtener una mejor comprensión con respecto a las zonas encontradas para las salidas metabólicas observadas para la cepa AR36 en el área experimental, se construyeron modelos de flujo metabólico dinámicos de acuerdo con lo presentado en el capítulo de materiales, modelos y métodos. Esto se debió a que fue evidente que la cepa AR36 mostraba una regulación no lineal dentro del área experimental que generaban zonas definidas entre salidas preferenciales. Los cálculos de los flujos internos por medio de la aproximación cibernética usada permitieron determinar el comportamiento de salida biomasa, GLC, SA y AC. Como se puede observar en la Figura 4.4, el comportamiento modelado para las entradas y salidas metabólicas siguió los puntos experimentales de buena forma. Es importante notar que, estos modelos metabólicos pudieron a diferencia de los fisiológicos presentados en la sección anterior, describir adecuadamente el comportamiento del AC. La principal característica encontrada en los perfiles de AC para las fermentaciones fue una producción inicial

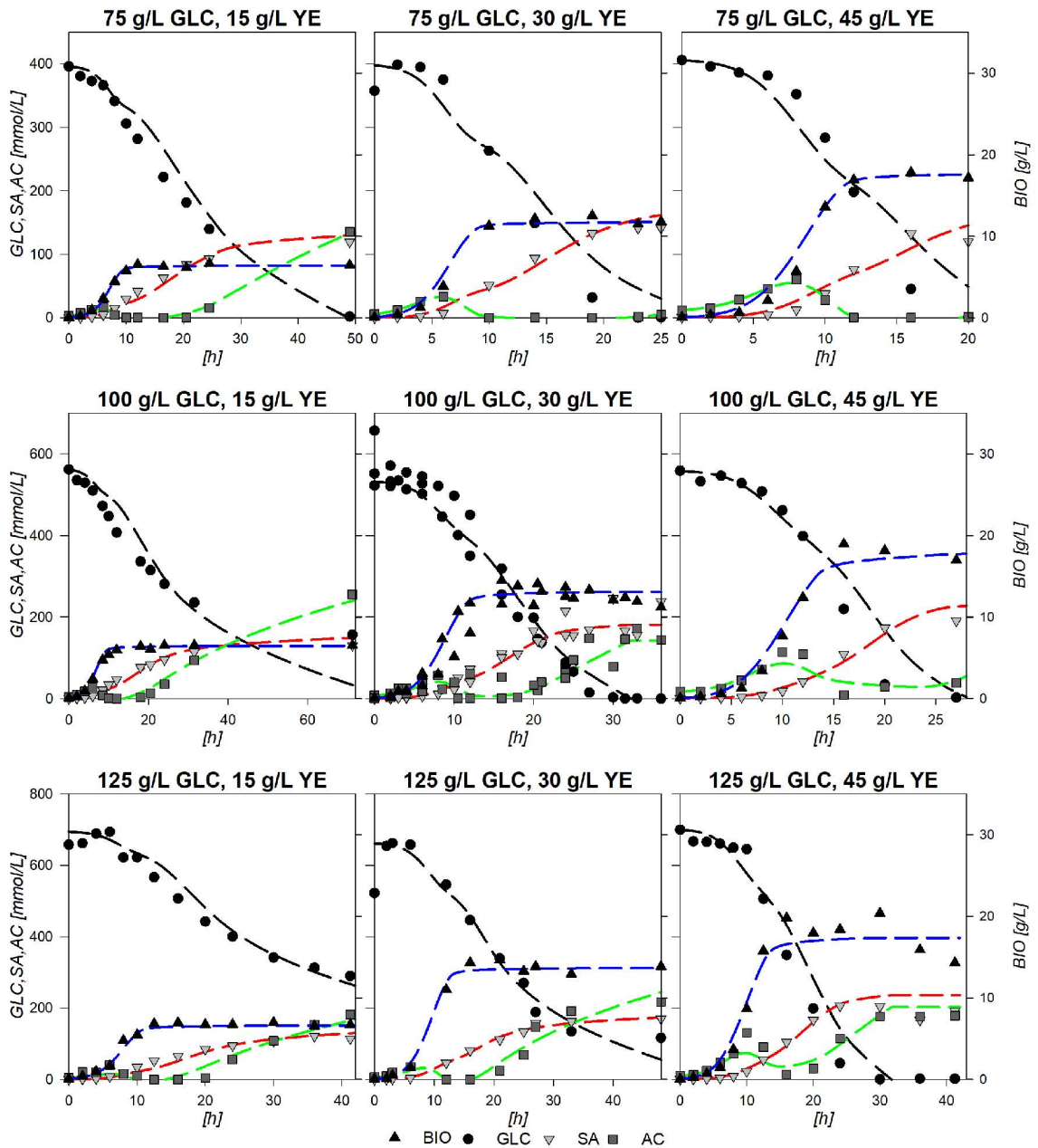


Figura 4.4: Comparación de los datos experimentales con los perfiles de metabolitos externos resultantes de la modelación de las distribuciones de flujos para cada punto del diseño experimental.

de AC en fase exponencial hasta un pico máximo local durante esta fase. Después de este punto el AC es consumido casi en su totalidad en todas las fermentaciones hasta el punto de entrada de la fase estacionaria. Posteriormente, un segundo proceso de producción de acetato comienza en esta sección en todos los experimentos exceptuando los realizados en las condiciones 75:30 y 75:45 GLC:YE, posiblemente debido al agotamiento de la GLC en tiempos mas cortos para estas últimas condiciones.

Los modelos metabólicos presentaron valores de entre 0.15 y 2.28 % de error para las aproximaciones de biomasa, de 0.56 a 2.3 % de error para la GLC, de 0.25 a 4.19 % de error para el SA y de 0.35 a 10.53 % de error para el AC. Los valores de  $R^2$  de Pearson para las regresiones lineales entre los valores del modelo y los valores experimentales resultaron todas arriba de 0.9 y el valor de significancia para la linealización fue menor a 0.05 para todas las comparaciones. Con respecto al parámetro de comparación SPD, los valores más altos de desviación se encontraron para los perfiles de GLC y de AC. Específicamente, un 21 % de SDP fue encontrado para el perfil de GLC en la condición 100:15. Con relación a este valor, el modelo presenta un mayor consumo de GLC en la parte final de la fermentación comparada con los valores experimentales. En este sentido, se puede observar en la Figura 4.4, que en esta condición particular el modelo casi agota la GLC mientras que el valor experimental de punto final para este compuesto es de 156 mM, indicando que el modelo sobrestima el consumo de GLC. En comparación, en todas las otras fermentaciones, el modelo se observa subestimando el consumo de GLC en la última fase de fermentación, generando valores de SDP de entre 1.07 a 19 %, donde las desviaciones más altas corresponden a las fermentaciones que contienen concentraciones iniciales de YE más altas. Para el caso específico de la condición 75:15 g/L, esta subestimación del consumo (encontrada de forma casi generalizada), podría significar que la GLC fuera agotada en un tiempo menor al estimado por el modelo.

Con respecto a las desviaciones en los perfiles de AC, los valores de SDP variaron de 1.31 a 9 % en todos los casos, excepto para las fermentaciones en las condiciones de 100:30 y 100:35 donde se estimaron del orden de 21 % y 42 % respectivamente. Para la condición 100:45, la desviación fue derivada de la diferencia en el pico de



AC en la mitad de la fase exponencial y a la falta de producción de AC en la fase estacionaria experimentalmente. Estos grandes valores de SDP pueden deberse a que la contribución de YE fue simplificada a sólo considerarlo como fuente de precursor de biomasa (BIOMp), glutamato (GLU), alanina (ALA) y aminoácidos aromáticos de forma acoplada, lo que podría generar incrementos no deseados en el consumo de fuentes de carbono y la acumulación exacerbada de AC. A pesar de estos valores de SDP, los demás valores estadísticos y la observación del seguimiento de los modelos con respecto a la comparación de Datos experimentales con los perfiles experimentales mostrados en la Figura 4.4, sugieren que los modelos construidos son aproximaciones suficientemente viables para describir el comportamiento metabólico de la cepa dentro del área de diseño experimental.

Para realizar la caracterización del comportamiento metabólico de la cepa AR36, los flujos calculados fueron normalizados con respecto al flujo de consumo de GLC y utilizados para construir superficies de respuesta en las diferentes etapas de cultivo. De esta forma se generaron superficies de respuesta de distribuciones de flujo metabólico al inicio de la fase exponencial (IEx), mitad de la fase exponencial (MEx) y mitad de la fase estacionaria (MST). Las superficies calculadas para los puntos IEx y MEx presentaron comportamientos y valores similares, por lo que el análisis y caracterización fue realizado sólo sobre IEx.

### **Caracterización del comportamiento de las distribuciones de flujo en el metabolismo central de carbono durante el crecimiento**

Se presentan superficies de distribución de flujo de reacciones selectas del CCM en el tiempo IEx en las Figuras 4.5 y 4.6. La Figura 3.1 muestra las vías del metabolismo central utilizadas y los nombres presentados en esta serán empleados como referencia en las siguientes secciones. Las superficies para las reacciones glicolíticas del EMP durante el crecimiento, presentan la misma morfología desde la entrada de la glucosa-6-fosfato (G6P) hasta su conversión en PEP. Estas reacciones presentan superficies con estructura de silla de montar, teniendo los mayores valores de flujo relativo en condiciones iniciales bajas de [GLC]. A pesar de tener la misma morfología, los valores entre estas presentan diferencias importantes. En específico, las

reacciones de Pgi (Figura 4.5 A), Pfk y FBA presentan valores relativamente bajos a los esperados, acumulando sólo entre  $\approx 6$  a 25 % del flujo relativo al consumo de glucosa por parte de las reacciones GalP y Glk. Esto provoca que, en los modelos, la mayor parte del flujo glicolítico pase a través de las reacciones oxidativas de la vía de PPP, entrando a este por medio de la reacción catalizada por G6Pdh codificada por el gen *zwf* y la 6-fosfogluconolactonasa (Pgl). G6Pdh presenta valores de  $\approx 75$  a 95 % de flujo relativo (Figura 4.5 E) y su morfología presenta las características inversas morfológicas a la de Pgi, teniendo los valores más altos entre mayores sean las concentraciones iniciales de [GLC] y [YE]. A pesar del bajo flujo sobre Pgi, se observan flujos relativos de entre el 75 y 88 % para las reacciones catalizadas por la gliceraldehido-3-fosfato deshidrogenasa-A (GapA) (Figura 4.5 B), y las reacciones lineales subsiguientes del EMP hasta el PEP. Esto es posible ya que en el nodo de las PPP de las reacciones catalizadas por las enzimas 6-fosfogluconato deshidrogenasa(Gnd)/fosfogluconato deshidratasa(PGdh), el flujo relativo revela valores hacia la vía de Entner-Doudoroff (EDP) de entre  $\approx 80$  a 95 %, y sólo entre ( $\approx 6$  a 8 %) hacia la ruta no oxidativa de las PPP (Figura 4.5 F y G respectivamente). Esto indica que la mayoría del carbono redirigido hacia la sección oxidativa de la PPP tiene por destino el glicerol-3-fosfato (G3P) y el piruvato (PYR). Lo anterior sugiere que, aún con bajos valores de flujo sobre Pgi, la glicólisis puede tener valores altos de distribución de flujo con el nodo de G3P y que la enzima GapA podría resultar importante para el control para la velocidad de flujo de G6P a PEP.

Las superficies de flujo sobre las reacciones catalizadas por Pgl, Gnd, Pgdh y la 2-ceto-3-deoxy-6-fosfogluconato aldolasa (KDPG1) muestran la misma morfología que la de G6Pdh. Esta está caracterizada por una zona de tendencia a la maximización hacia valores iniciales más altos de [YE] (Figuras 4.5 E a G). En estas condiciones, en las superficies de descripción de los parámetros fisiológicos se había observado la zona de maximización de la producción de biomasa, lo cual podría estar relacionada con esta observación de maximización de flujo a través de la vía oxidativa de PPP, la cual sintetiza NADPH, además de R5P para la síntesis de RNA y DNA, y por lo tanto podría estar estimulando la formación de bloques biosintéticos para la biomasa. Con respecto a las reacciones de la sección no oxidativa de la ruta de las PPP, la transcetolasa I (TktA) y la transaldolasa (Tal) presentan superficies con

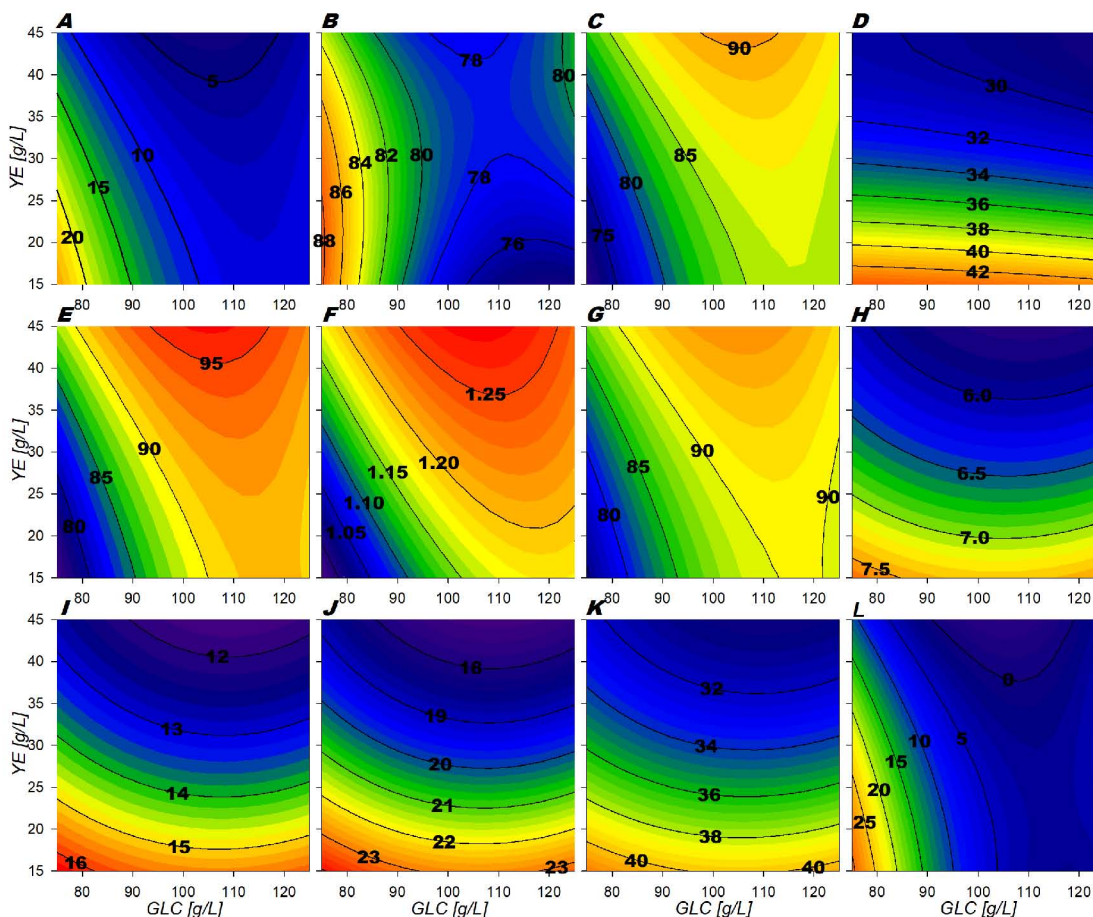


Figura 4.5: Contornos de superficies de respuesta para la fase de crecimiento pt.1 (% Flujo relativo al consumo de GLC). A) Pgi, B) GapA, C) PykA, D) LpdA, E) G6Pdh, F) Gnd, G) PDdh, H) TktA 1, I) TktA 2, J) DAHPs, K) PckA, L) Ppc

flujos relativos bajos,  $\approx 6$  a  $8\%$  (Figura 4.5 H) con una tendencia de incremento en condiciones iniciales bajas de [YE]. Esto indica que la mayor proporción de flujo de carbono hacia E4P está siendo dirigido desde el EMP a través de la reacción secundaria de TktA, la cual convierte F6P y G3P en E4P y X5P. Esta reacción presenta flujos relativos de  $\approx 12$  a  $16\%$  (Figura 4.5 I). Estos últimos, sumados al flujo de la vía no oxidativa de las PPP, permite una distribución de flujo relativo de  $\approx 18$  a  $24\%$  hacia la ruta del SA por medio de la reacción catalizada por la enzima 2-dehidro-3-deoxifosfoheponato aldolasa (DAHPs) durante la fase de crecimiento (Figura 4.5 J).

Una consecuencia interesante de este bajo flujo, aun con sobreexpresiones constitutivas y fuertes de los genes *zwf* y *tktA*, es que la mayoría del flujo en el modelo está siendo redirigido hacia la formación de PYR por medio de la enzima piruvato cinasa II (PykA), con valores de  $\approx 72$  a  $90\%$  (Figura 4.5 C). Lo anterior significa que el substrato limitante para la producción de SA es en efecto la E4P, provocando en la cepa un sobreflujo sobre PEP. En este sentido la fosfoenolpiruvato carboxilasa (Ppc) muestra valores de flujo relativo de  $\approx 31$  a  $41\%$  para la conversión de PEP hacia OA (Figura 4.5 L), presentando la misma morfología de los genes glicolíticos. Mientras tanto, la reacción catalizada por la fosfoenolpiruvato carboxicinas (PckA) (Figura 4.5 K) presenta valores de  $\approx 0$  a  $30\%$  de flujo con la tendencia a incrementar hacia condiciones iniciales de baja [YE], lo que indica la aparición de un ciclo de consumo de ATP.

En esta etapa de fermentación, se observa una gran proporción de flujo hacia PYR, causado por el flujo a través de EDP y PykA principalmente. Además, también se observan contribuciones realizadas por una alta cantidad de flujo relativo de las rutas málicas, reincorporando esqueletos de carbono desde el ciclo de los ácidos tricarbóxicos (TCA). Estas últimas presentan valores de  $\approx 28-38\%$  para la enzima dependiente de NADPH (MaeB) (Figura 4.6 L) y de  $\approx 0-30\%$  para la enzima dependiente de NADH (MaeA). En el modelo, el PYR también puede ser producido de la reacción de consumo de YE derivada de la conversión de ALA a este compuesto por la alanina-D-aminoácido deshidrogenasa (DadA). Todo lo anterior sugiere que el PYR, que es un nodo recolector y de distribución de carbono en el CCM, se encuentra posiblemente saturado en la cepa AR36 en las condiciones experimentales. En este sentido, la reacción de conversión de PYR hacia acetyl coenzima-A (ACCOA), presenta un flujo relativo con valores de  $\approx 28-42\%$  los cuales se esperarían fueran más altos dada la cantidad de carbono llegando a PYR. Esto último se debe a que, en el modelo, se observa que la mayoría del PYR está siendo convertido a ACCOA a través de un ciclo extendido a la producción y consumo de AC iniciado por la piruvato oxidasa (PoxB). Esta reacción presenta flujos relativos de  $\approx 176-186\%$  (Figura 4.6 E). El AC producido es posteriormente consumido de vuelta ya sea por la acetyl-CoA sintetasa (Acs) o la acetato cinasa (AckA). Es importante notar que para las

reacciones de AckA y Acs ninguna restricción fue impuesta en el modelo, por lo que la diferenciación es realizada por la producción y consumo de ATP. Debido a que la reacción de AckA es más favorable energéticamente (menos requerimiento de ATP) el modelo parece preferir esta reacción para la reincorporación del AC al CCM. Por tanto, la velocidad de consumo de Acs presenta valores más bajos con respecto a AckA (Figuras 4.6 C a E). Este ciclo de producción/consumo está relacionado con la observación de la aparición del pico de AC en fase exponencial descrito anteriormente. Relacionado a esto, los flujos de transporte extracelular de AC muestran una mayor proporción de flujo en zonas de alta concentración inicial de YE. Esto puede ser atribuido a la introducción de esqueletos de carbono al CCM a través del consumo de ALA y GLU derivados del catabolismo de los componentes del YE (Figura 4.6 A y B). Esta inferencia se podría extender a otros aminoácidos no tomados en cuenta en el modelo. Por otro lado, el importe de AC presenta una tendencia a maximizarse hacia concentraciones iniciales menores de YE, presentando flujos en toda el área de diseño experimental de entre  $\approx 90$  a  $115\%$ .

Para validar y clarificar la distribución de flujo alrededor de la reacción de PoxB, se construyó una mutante de AR36 carente del gen *poxB*. Esta cepa AR36 $\Delta$ *poxB* fue cultivada en condiciones iniciales de alta concentración de GLC y YE, las cuales fueron determinadas como condiciones de maximización para la producción de AC, tanto por las superficies de flujo (Figura 4.6 E) como por las superficies de comportamiento fisiológico (Figura 4.5). Interesantemente, el perfil de productos de fermentación para esta cepa en las condiciones 145:45 GLC:YE g/L, mostró la pérdida del pico de concentración de AC observado en todas las fermentaciones realizadas con la cepa parental AR36 (Figura 4.7). Estos resultados apoyan a la distribución de flujo encontrada por los modelos metabólicos que sugieren a PoxB como la principal enzima contribuidora a la producción de AC en la cepa AR36. Aunado a lo anterior, la cepa AR36 $\Delta$ *poxB* mostró una disminución de la velocidad de crecimiento hasta valores de  $0.21 \text{ h}^{-1}$ , relacionada a una disminución de la velocidad de consumo de GLC en fase exponencial hasta valores de  $0.61 \text{ g/gh}$ . Esto resultó, además, en una disminución de la producción de SA cercana al  $50\%$ . Esto podría indicar que su inactivación causa una acumulación en la concentración intracelular de PYR, lo

cual provoca cascadas de señales de regulación (ej. reguladores globales como *Crp*) que derivan en una menor velocidad de consumo (Shimizu, 2014). Además, podría ocasionar una concomitante reducción de la producción de ATP por la cadena de transferencia de electrones, debido a que uno de los subproductos de la reacción de PoxB es una molécula oxidada de hidroquinona utilizada por los complejos de esta cadena (Causey et al., 2004).

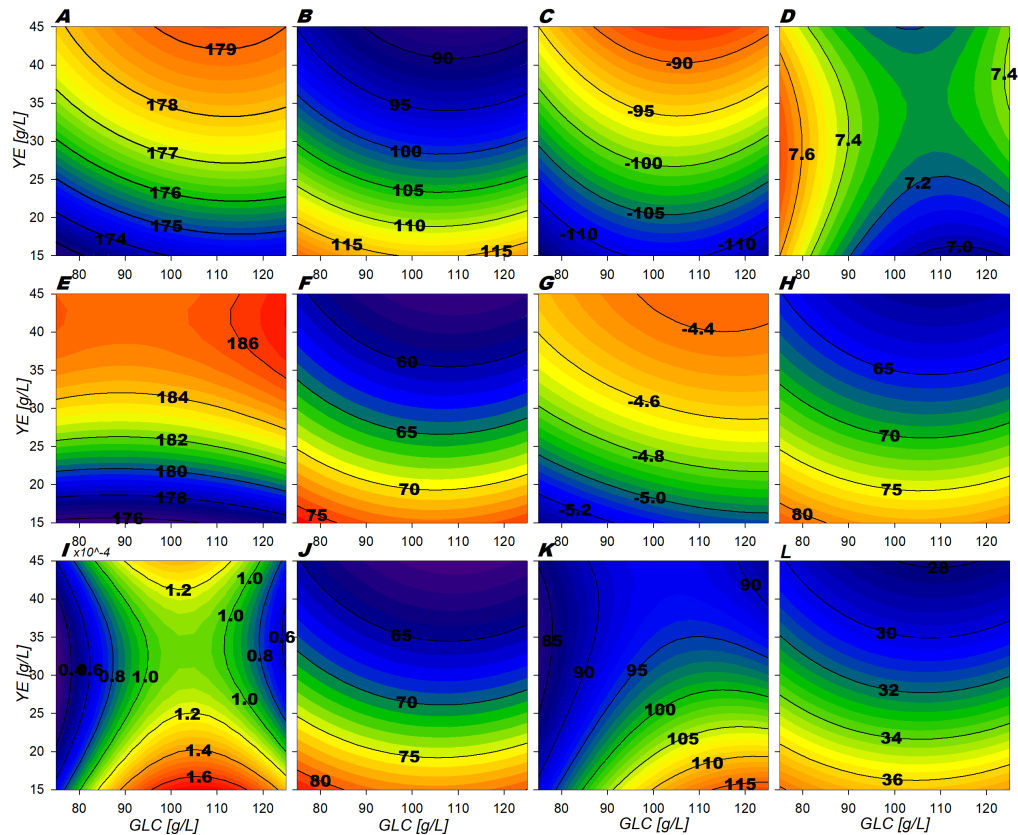


Figura 4.6: Contornos de superficies de respuesta para la fase de crecimiento pt.2 (% Flujo relativo al consumo de GLC). A)ActPout, B)ActPin, C)AckA, D)Acs, E)PoxB, F)Csyn, G)IcdA, H) Icl, I)KGdh, J)SdhABCD, K)Mdh, L)MaeB

Con respecto a las distribuciones de flujo en el TCA, la reacción catalizada por la enzima citrato sintasa (Csyn) presenta valores de  $\approx 56$  a 75 % de flujo (Figura 4.6 F). De forma esperada, la reacción de la aconitasa (Acn) presenta el mismo comportamiento de la reacción anterior, donde ambas tienen características morfológicas de

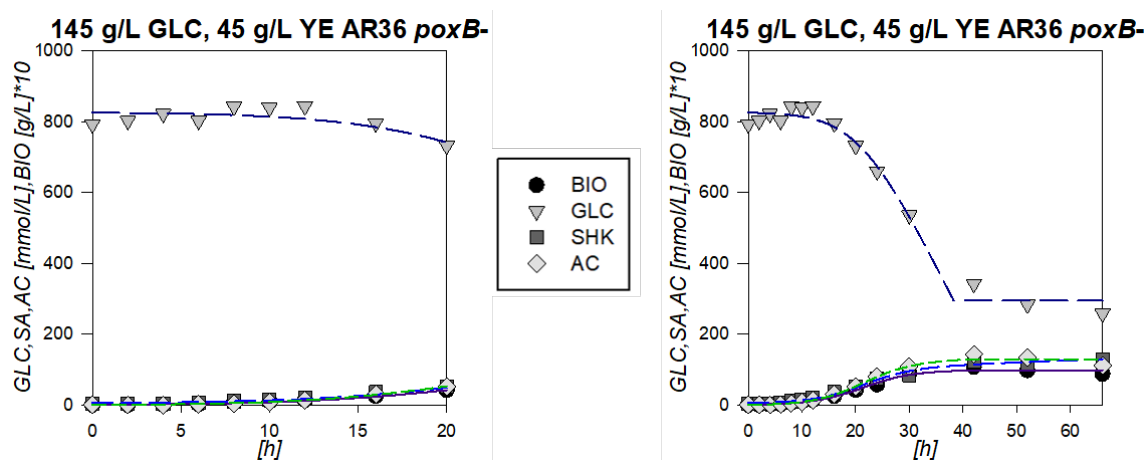


Figura 4.7: Perfiles de fermentación de la cepa  $AR36\Delta poxB$  cultivada en condiciones iniciales 145:45 GLC:YE g/L izq. sección hasta las 20 horas (fase de crecimiento). der. Perfil completo de fermentación. Las líneas punteadas representan el resultado del cálculo de los modelos logísticos obtenidos para esta cepa.

superficie inversas a las relacionadas con la producción de AC, indicando un comportamiento competitivo entre estas dos secciones del metabolismo. Un resultado no intuitivo encontrado en el comportamiento del TCA, fue que la reacción siguiente catalizada por la isocitrato deshidrogenasa (IcdA), parece no estar enviando flujo en la dirección oxidativa del TCA. De forma contraria, el modelo presenta la reacción en el sentido reversa, transportando el exceso de GLU posiblemente derivado del consumo de YE hacia isocitrato (ICIT) (Figura 4.6 G). Los cálculos de este flujo en sí mismo son relativamente pequeños, ya que la enzima glutamato deshidrogenasa (Gdh) presenta distribuciones de entre  $\approx 3.1$  a  $3.7\%$ , sugiriendo una aportación menor de los componentes del YE hacia la vía de TCA. Derivado de esto, el flujo se distribuye preferencialmente a través de las reacciones catalizadas por las enzimas isocitrato liasa (Icl) y malato sintasa (Msn) de la vía de la lanzadera de glioxilato, presentando valores de flujo relativo contabilizando del  $\approx 60$  al  $82\%$  del flujo relativo (Figura 4.6 H). Estas superficies muestran las mismas características morfológicas que la reacción de entrada de flujo de carbono hacia el TCA (Csyn). Por otro lado, las enzimas 2-cetoglutarato deshidrogenasa (KGdh) y el complejo de las succinil-CoA sintetasas (SucCD) presentan distribuciones de flujo bajas o casi nulas hacia

la producción de succinato (SUC) (Figura 4.6 I), indicando que la mayoría del SUC está siendo producido por la vía GSP. Subsecuentemente, el SUC es transformado en malato (MAL) por medio del complejo de succinato deshidrogenasas (SdhABCD) (Figura 4.6 J) y la fumarasa (Fum). Estas dos reacciones comparten por lo tanto las características morfológicas de las reacciones del GSP. En consecuencia, la malato deshidrogenasa presenta flujos más altos, de entre  $\approx 85$  a 115 % (Figura 4.6 K) dado que asimila otra molécula de ACCOA derivada de las reacciones de GSP.

La observación de las características morfológicas de las reacciones del CCM en la fase de crecimiento exponencial sugieren también, la localización de diferentes flujos predominantes en diferentes zonas del diseño experimental. De forma tal que: la producción de SA predomina en zonas de baja concentración inicial de YE. Además, se observa una disminución del flujo hacia AC producido con respecto a la disminución de la concentración inicial de GLC, indicando que en estas regiones se favorece un crecimiento más balanceado. Esto genera, una zona de preferencia de flujo hacia la producción de biomasa en condiciones de baja concentración inicial de GLC que aumenta conforme se incrementa la concentración inicial de YE. Finalmente, se aprecia una zona de mayor velocidad de producción y consumo de AC en zonas de alta concentración inicial de ambos substratos, GLC y YE. Esta distribución de zonas concuerda con las superficies generadas a partir de los parámetros calculados con los modelos logísticos.

### **Caracterización del comportamiento de las distribuciones de flujo en el metabolismo central de carbono después del crecimiento**

En la fase estacionaria de la fermentación también se observó una distribución de flujo hacia la vía de las PPP, con valores de entre el 97 al 108 % a través de las reacciones catalizadas por las enzimas G6Pdh y Pgl (Figura 4.8 E). Estas dos enzimas de la sección oxidativa de la PPP presentan morfologías de superficie que indican una tendencia a la minimización en forma circunyacente hacia valores bajos de concentración inicial de GLC y altos de YE. En contraste con la fase de crecimiento, la reacción catalizada por Gnd presenta valores de flujo relativo más altos con  $\approx 40$  a 100 % (Figura 4.8 F). Mientras que la reacción competidora catalizada por



PGdh, presenta valores de entre  $\approx 0$  a 70 % de flujo hacia la vía EDP. Estas superficies presentan morfologías de comportamiento invertidas como se puede observar en las Figuras 4.8 E y F. Interesantemente, la superficie de la PGdh tiene su zona de maximización en la misma región que la superficie de la superficie de G6Pdh, lo que sugiere que el exceso de flujo es lo que promueve el uso de la vía EDP. Por su parte, la morfología de Gnd presenta un comportamiento anillado con la tendencia a un máximo hacia la región de baja [GLC] y alta [YE] inicial, esquina superior izquierda del área del diseño experimental. Debido a esta distribución, la reacción de TktA y Tal hacia la producción de R5P, E4P y F6P en la sección no oxidativa de la PPP exhibe el mismo comportamiento que la Gnd.

El F6P producido por estas reacciones entra entonces a la vía de EMP donde es dirigido hacia la formación de PEP a través de las reacciones glicolíticas, entre ellas las catalizadas por las enzimas Pfk y Fba, cuyos valores de flujo relativo fueron estimados en  $\approx 10-55$  %. La morfología de superficies para estas últimas enzimas, también muestran una superficie anillada con la tendencia a un máximo hacia condiciones iniciales de baja [GLC] y alta [YE]. De hecho, estas características morfológicas son observadas a través de todas las reacciones glicolíticas subsiguientes del EMP hasta la formación de PEP, pero con valores de flujo relativo más altos de  $\approx 90$  a 99 % (Figura 4.8 B). De forma poco intuitiva, el modelo describe un flujo bajo a través de la enzima Pgi de  $\approx 0$  a 9 % (Figura 4.8 A) en la dirección de la F6P a G6P, el cual es consumido por la G6Pdf, en un ciclo que podría estar perdiendo un carbono a través de dicha reacción, pero ganando poder reductor en forma de NADPH, en casi toda el área del diseño experimental. Consecuentemente, Pgi presenta sólo un flujo bajo en la dirección oxidativa estándar de G6P a F6P en condiciones iniciales de baja [GLC] y alta [YE], representando sólo entre el  $\approx 0-2$  % del flujo relativo. Con respecto a las reacciones del nodo de PEP, la reacción de la fosfoenolpiruvato sintasa (PpsA) es poco significativa, sugiriendo que no existe flujo gluconeogénico del PYR hacia el PEP en ninguna de las condiciones evaluadas. En la misma medida, la reacción catalizada por la PckA presenta valores de flujo relativa bajos, de entre  $\approx 0-5$  % donde sólo en condiciones de alta concentración inicial de YE existe flujo de oxaloacetato (OA) hacia PEP (Figura 4.8 K). Por el otro lado, la reacción

de Ppc muestra valores más altos ( $\approx 0-50\%$ ), con valores mayores hacia la zona de baja [YE] inicial y en la dirección de aumento de la concentración inicial de GLC (Figura 4.8 L). Estas dos superficies muestran también morfologías anilladas, pero con tendencias de maximización/minimización invertidas.

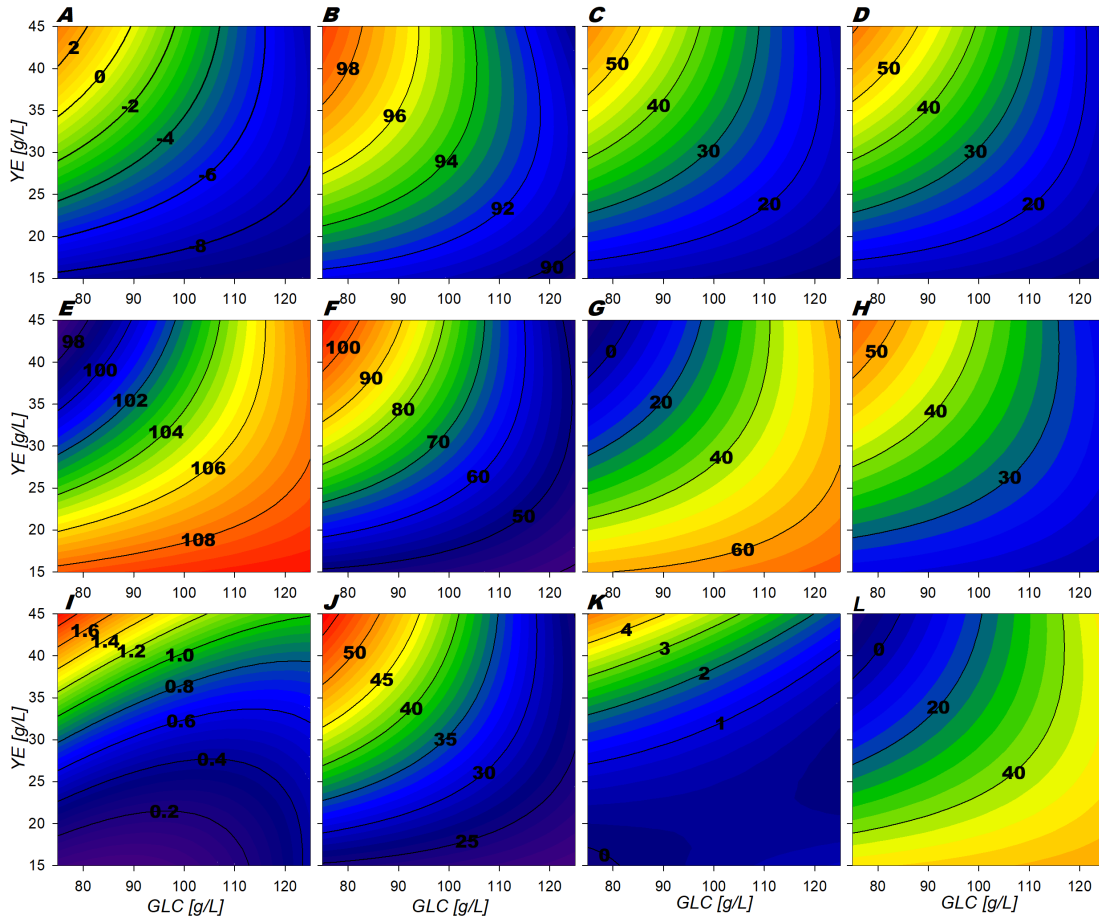


Figura 4.8: Contornos de superficies de respuesta para la fase estacionaria pt.1 (% Flujo relativo al consumo de GLC). A) Pgi, B) GapA, C) PykA, D) LpdA, E) G6Pdh, F) Gnd, G) PDdh, H) TktA 1, I) TktA 2, J) DAHPs, K) PckA, L) Ppc

Subsecuentemente, el consumo de PEP por PykA y la DAHPs presentan la misma morfología, pero con la tendencia de maximización hacia baja [GLC] inicial y altas [YE] iniciales, similares a los genes glicolíticos. En contraste con las observaciones hechas en la fase exponencial de crecimiento, estas dos superficies presentan

distribuciones de flujo similares en toda el área de diseño experimental con valores de entre  $\approx 15-55\%$  (Figuras 4.8 C y J), debido a una mayor proporción de flujo pasando por la vía no oxidativa de la PPP y su concomitante aumento en la producción de E4P. La zona de mayores flujos relativos a través de la DAHPs hacia SA se encuentra efectivamente sobre la misma zona caracterizada por las mayores velocidades específicas de producción de SA por los modelos logísticos. Las reacciones de PykA y DAHPs compiten con las reacciones de producción de acetato, en específico con PoxB, la cual exhibe una tendencia a la maximización de su flujo relativo hacia valores altos [GLC] inicial y bajas [YE] iniciales en las fermentaciones (Figura 4.9 E). Por estos motivos la superficie de PoxB muestra características de superficies inversas a estas últimas. La reacción de exportación de AC sigue el comportamiento de la superficie de PoxB, por lo que esta es la principal vía de producción de AC también en esta fase de fermentación (Figura 4.9 A). Por otro lado, la reacción de importación de AC presenta valores de casi  $\approx 0\%$  de flujo relativo en condiciones iniciales de baja [YE]. En efecto sólo se observan flujos de consumo extracelular de AC en concentraciones bajas de GLC y concentraciones iniciales altas de YE, esquina superior izquierda del área de diseño experimental (Figura 4.9 B). Esto podría estar relacionado con el hecho de que, en concentraciones iniciales más altas de YE, más biomasa es producida y por lo tanto más GLC ha sido consumida, lo que significa que a la entrada de la fase estacionaria se esperan concentraciones menores de GLC y consecuentemente una disminución de la producción de acetato. Esto sugiere que concentraciones altas de YE podrían estar contribuyendo a la asimilación de AC en fase estacionaria. Además, se ha reportado que algunos de sus aminoácidos podrían provocar la sobreexpresión del metabolismo gluconeogénico y la concomitante fijación del ciclo de AC en fases de crecimiento exponencial (Aguilar et al., 2012; Flores et al., 2004, 2005, 2002; Rodriguez et al., 2013; Sigala et al., 2008).

Con respecto al comportamiento del TCA y del GSP, sus superficies también exhiben morfologías de superficie anilladas con tendencias particulares de maximización o minimización hacia la esquina superior izquierda del diseño experimental. Específicamente, las reacciones catalizadas por Csyn y Acn presentan la tendencia a un mínimo en esta zona y valores de flujo relativo de  $\approx 54$  a  $70\%$  (Figura 4.9 F).

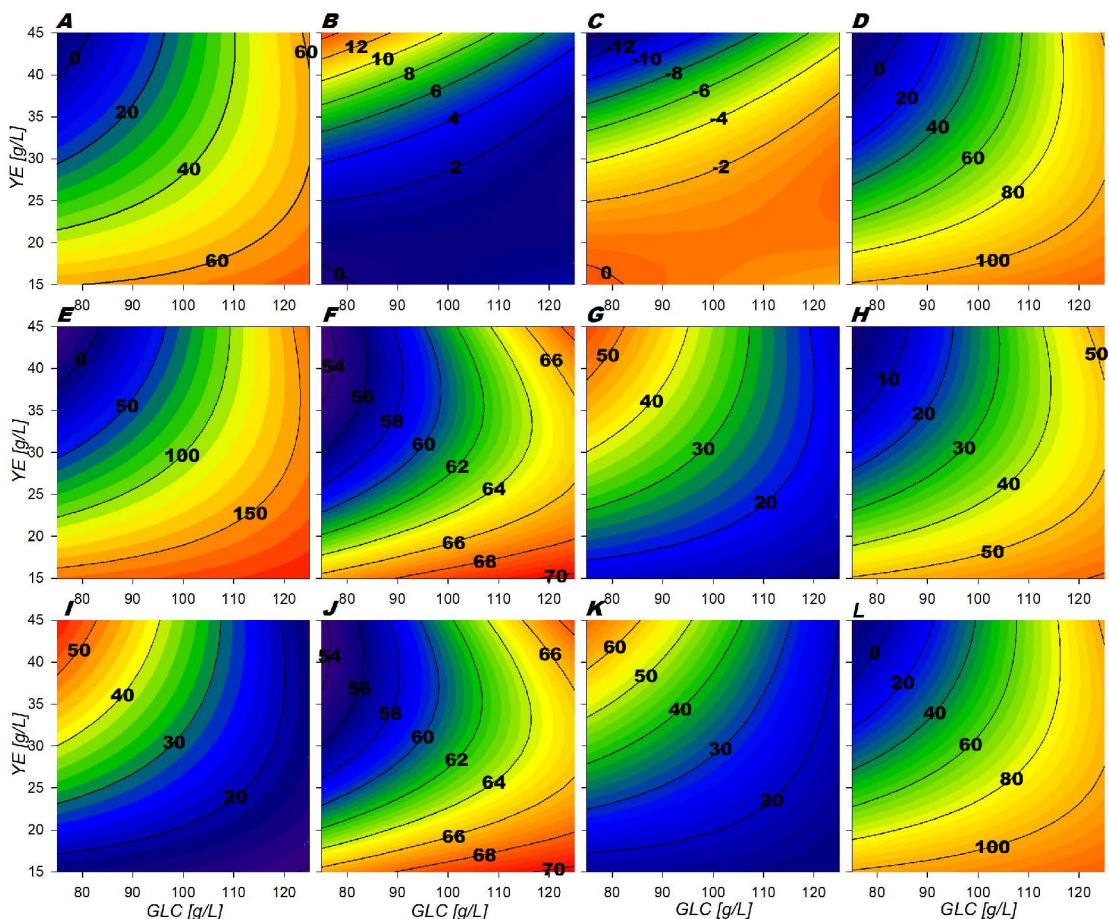


Figura 4.9: Contornos de superficies de respuesta para la fase estacionaria pt.2 (% Flujo relativo al consumo de GLC). A)ActPout, B)ActPin, C)AckA, D)Acs, E)PoxB, F)Csyn, G)IcdA, H) Icl, I)KGdh, J)SdhABCD, K)Mdh, L)MaeB

De la misma forma, las enzimas del GSP presentan este comportamiento (Figura 4.9 H). En contraste con la fase de crecimiento, las reacciones de IcdA, KGdh y SucCD presentan flujos oxidativos en el TCA en toda el área del diseño experimental, presentando flujos entre  $\approx 10$  al 55% (Figuras 4.9 G, I, J). Sus superficies presentan una morfología de maximización hacia la región de condiciones iniciales de baja [GLC] y alta [YE]. Las reacciones catalizadas por la Sdh y Fum siguen el mismo comportamiento que Csyn. Adicionalmente, conforme mayor porcentaje de flujo es jalado hacia TCA, mayor es el flujo de la reacción málica catalizada por MaeB, la cual compite con Mdh en la formación de OA (Figuras 4.9 L y K respectivamente).

Esto está relacionado con el aumento de flujo hacia PYR y la concomitante expresión del ciclo de reciclaje de carbono a través del AC. En efecto en esta zona se observa que el ciclo de reciclaje de carbono tiene un recorrido desde el PYR pasando por AC hacia TCA y las vías del GSP, conforme más saturación de carbono existe en el ambiente extracelular.

Las superficies de las distribuciones de flujo generadas por los modelos dinámicos en la zona de la mitad de la fase estacionaria de crecimiento sugieren la localización de dos zonas primordiales de salida de flujo en el área del diseño experimental. De esta forma se puede nuevamente trazar una diagonal que corte el área experimental de las condiciones iniciales bajas a altas de concentración de ambos sustratos. Donde la zona de producción predominante de SA se encuentra en la parte superior a esta diagonal imaginaria, mientras que la zona de predominancia de producción de AC se encuentra debajo de la misma. Esta localización de zonas se relaciona de buena forma con las establecidas por los parámetros de fermentación de la caracterización de la cepa AR36 (Rodriguez et al., 2013, 2017).

## 4.2. DISCUSIÓN

### 4.2.1. Discusión, análisis y comparación de los resultados obtenidos de los modelos fisiológicos y metabólicos para la determinación de blancos genéticos y metabólicos para el mejoramiento de la producción de ácido shikímico.

Como se observa en las secciones anteriores, las superficies de respuesta construidas de la ecuación polinomial fueron capaces de caracterizar de forma adecuada el comportamiento fisiológico de la cepa AR36. El aumento observado en  $X_{max}$  principalmente con respecto a la [YE] está relacionado con el hecho de que éste último es la única fuente de aminoácidos aromáticos (Figura 4.2 A). Esto indica que, en toda el área del diseño experimental, el YE puede tomarse como el substrato limitante para la producción de biomasa. Esta inferencia también se puede observar en los perfiles de fermentación mostrados en la Figura 4.1, donde a la entrada de la fase estacionaria existen concentraciones de GLC superiores a limitantes en todos los experimentos. Es un resultado esperado que la producción de SA siga el perfil de superficie de la del consumo de GLC, debido a que es la principal fuente de carbono para la producción de PEP y E4P, como se puede observar en las Figuras 4.2 B y C. Sin embargo, la producción de SA muestra un punto máximo antes de las condiciones de consumo total de GLC. Esta diferencia puede ser explicada por la producción final de AC ( $[AC]_f$ ), la cual incrementa conforme las concentraciones iniciales de GLC aumentan. Este último proceso se ve exacerbado además conforme las condiciones iniciales de [YE] son menores (Figura 4.2 D). Esto sugiere que el extracto de levadura promueve una menor producción de AC. Esto puede ser debido posiblemente a una mayor cantidad de biomasa en los fermentadores consumiendo mayores cantidades de GLC, al estrés de producción de AC por exceso de fuente de carbono, así como a que algunos componentes del extracto de levadura pueden fomentar un consumo mayor de las diferentes fuentes de carbono, inclusive de AC (Aguilar et al., 2012; Areense et al., 2010; Flores et al., 2004, 2005, 2002; Matsuoka y Shimizu, 2013; Rodriguez et al., 2017; Sigala et al., 2008).

El aumento de la cantidad de AC en las condiciones de alta [GLC] y baja [YE], podría también estar inhibiendo la capacidad de consumo de glucosa y de producción de ácido shikímico. En efecto, altas [AC] pueden desajustar el balance de iones  $H^+$  en la membrana plasmática de las bacterias, y dado el caso de que la cepa AR36 usa el simportador GalP, el transporte de GLC podría verse comprometido (Rodríguez et al., 2013). Además, se ha reportado que altas concentraciones de AC hacen más costosa la producción de ATP en términos energéticos debido a la pérdida de potencial electroquímico a través de la membrana, y en consecuencia también la fosforilación ATP-dependiente de GLC por medio de la glucocinasa GLK (Rodríguez et al., 2013). Finalmente, el desbalance electroquímico en la cadena transportadora de electrones podría estar modificando el balance en los cofactores redox NADH/NAD y NADPH/NADP lo cual impactaría sobre la velocidad de reacciones dependientes de éstos últimos como el SA. La producción de AC se relaciona comúnmente con la saturación cinética de las rutas metabólicas, debido al desbalance entre la velocidad de las reacciones glucolíticas y la capacidad de oxidación de los intermediarios por TCA (Matsuoka y Shimizu, 2013). En este sentido, se ha reportado que los metabolitos PEP y PYR son sensores celulares para regular la velocidad de consumo de glucosa y los procesos de reciclaje de carbono y producción de AC (Matsuoka y Shimizu, 2013; Shimizu, 2014). En la cepa AR36, el gen de la piruvato cinasa *pykF* se encuentra inactivo lo que modifica el balance de las concentraciones de PEP/PYR. La disminución del flujo de PEP a PYR envía señales de regulación para la disminución de las reacciones glucolíticas y sobreexpresa enzimas de producción de AC que a su vez compiten por esqueletos de carbono y recursos energéticos para la producción de SA (Aguilar et al., 2012; Flores et al., 2004, 2005; Martínez et al., 2008; Matsuoka y Shimizu, 2013; Rodríguez et al., 2017).

De hecho, se observa que en las zonas donde la producción de AC se encuentra exacerbada, disminuye el consumo de GLC y la producción de SA (alta [GLC] y baja [YE] inicial). Las superficies permiten delimitar un límite virtual entre la preferencia de producción de SA con respecto a AC en valores cercanos a los 100 g/L de GLC, donde a concentraciones superiores, la producción de AC se torna perjudicial para la

eficiencia metabólica celular. Es importante notar que en todo el diseño experimental las concentraciones de GLC utilizadas típicamente resultarían en producciones más elevadas de AC y menores velocidades de crecimiento, consumo y producción de metabolitos en *E. coli* (Lara et al., 2007; Luli y Stohl, 1990). Sin embargo, se ha reportado que la cepa AR36 puede crecer con altas velocidades de crecimiento y mantener la producción de SA en concentraciones elevadas de GLC en el orden de los 100 g/L (Rodríguez et al., 2013, 2017). Esta característica le fue atribuida a la sobreexpresión constitutiva del operón sintético en el plásmido de alto número de copias para la producción de SA, y a la falta de represión catabólica derivada de su condición  $PTS^-$ , que permite un consumo simultáneo de otras fuentes de carbono incluyendo el AC (Aguilar et al., 2018; Rodríguez et al., 2017).

Con respecto a las velocidades específicas, se observa también una disminución de la velocidad de consumo con respecto al aumento de la [YE] inicial. Esta disminución podría estar relacionada a la necesidad de energía de transporte, cofactores y/o la asignación de aminoácidos para la síntesis de los diversos transportadores para poder consumir los componentes de YE. Esto puede ser posible debido a que la inactivación del operón *crrHI* en la cepa AR36, no sólo evita la represión catabólica, sino también aumenta la expresión de reguladores globales que inducen la síntesis de varios transportadores de carbono alternativos a GLC, entre otros mecanismos de búsqueda de substratos (Aguilar et al., 2012, 2018; Flores et al., 2005; Martínez et al., 2008; Rodríguez et al., 2017). Aunado a esto, en las condiciones altas de [YE] inicial en la fase de crecimiento, el consumo de PYR podría estar siendo saturado por la inclusión de carbono a través del consumo de alanina en el medio y por lo tanto señalizando la reducción del consumo de glucosa. En contraste, en la fase estacionaria la velocidad específica de consumo de glucosa presentó valores más altos en concentraciones iniciales altas para ambos substratos, GLC y YE. Este comportamiento podría ser esperado para la GLC, dado que concentraciones más altas de este substrato existen en fase estacionaria en estas condiciones iniciales, por lo que la velocidad de consumo se espera sea superior en comparación con condiciones iniciales más bajas de GLC. Sin embargo, permanece poco clara la razón de por qué altas concentraciones iniciales de YE podrían causar velocidades específicas más altas de



consumo de GLC en fase estacionaria. Una posibilidad es que, entre más altas sean las concentraciones iniciales de [YE] más grande podría ser la concentración intracelular de oxaloacetato derivado del consumo de algunos aminoácidos como el GLU y el aspartato (ASP), lo que podría aumentar la actividad del TCA y por lo tanto aumentar el consumo de GLC (Ahn et al., 2008). Esta observación está relacionada con el comportamiento del consumo total de glucosa al final de la fermentación, en la que sólo en concentraciones altas de [YE] se agotan completamente concentraciones superiores a los 110 g/L de GLC.

Con respecto a los perfiles de producción y consumo de SA y AC, respectivamente, se puede observar que son los principales productos metabólicos, junto con el dióxido de carbono, el cual no fue medido experimentalmente. Es posible observar tanto en la fase exponencial de crecimiento como en la fase estacionaria, zonas de maximización o de preferencia inversamente posicionadas en el área experimental (Figuras 4.2 F,H,I y J). Esta característica, en conjunto con la producción de biomasa, permitió establecer las zonas descritas en las secciones anteriores. Sin embargo, es interesante notar, que es posible encontrar velocidades similares de consumo de glucosa dentro del espacio experimental, con salidas fisiológicas completamente distintas, indicando que la regulación presenta un comportamiento no lineal. Este efecto se presenta aun modificando sólo la concentración inicial de las fuentes de carbono en concentraciones superiores a las limitantes. Más aún, las superficies de respuesta presentan contornos dependientes de la interacción de las dos fuentes de substratos, inclusive generando puntos críticos de punto de silla de montar como se observa en la superficie de velocidad de consumo de glucosa en fase de crecimiento. Este tipo de comportamiento sugiere la existencia de multiplicidad de estados metabólicos de forma similar a los descritos por Namjoshi y colaboradores (2001) para estados pseudo-estacionarios en reactores continuos (Namjoshi y Ramkrishna, 2001). Las diferencias en las respuestas de estas superficies de respuesta, deriva por tanto en la naturaleza dinámica de la regulación, lo que produce diferentes salidas dependientes de las condiciones intracelulares que modifican el comportamiento metabólico de la cepa de acuerdo a la dinámica extracelular, distinta para cada condición inicial (Martínez et al., 2008; Matsuoka y Shimizu, 2013; Namjoshi y Ramkrishna, 2001).

Para poder describir de mejor forma esta relación y algunas características del sistema difíciles de abordar sólo con la información obtenida de las superficies del comportamiento de los metabolitos externos, se construyeron modelos metabólicos dinámicos por medio de la aproximación cibernética. Esta aproximación fue utilizada debido a que incluye términos relacionados con la regulación de la síntesis y actividad enzimática sin la necesidad de describir el mecanismo o las condiciones específicas de su funcionamiento (Kompala et al., 1986; Ramkrishna y Song, 2012; Varner y Ramkrishna, 1999). Es importante notar que el medio de cultivo utilizado contiene un substrato no completamente descrito (YE) y por lo tanto fue utilizada una simplificación para la modelación, la cual consiste en que el catabolismo de este substrato (tomado como un solo compuesto) resulta en el consumo de GLU, ALA y precursores de biomasa (BIOMp) como se describe en la red metabólica descrita en materiales, modelos y métodos. Esta simplificación podría no sólo impactar en la velocidad de crecimiento, sino que también significa que el consumo de otros componentes del YE no tomados en cuenta podrían modificar los rendimientos y las velocidades de consumo y producción de los metabolitos incluidos en el modelo. Sin embargo, de acuerdo con lo observado en la caracterización del modelo metabólico en las secciones anteriores, la simplificación resultó suficiente para describir razonablemente algunas de las características de los flujos en la cepa AR36, y modelar el comportamiento extracelular de los metabolitos principales con suficiente precisión. De la misma forma, es importante notar que el algoritmo genético utilizado permitió encontrar valores para los parámetros del modelo adecuados y en relativamente pocas generaciones. El uso de este tipo de algoritmos para precisar el valor de parámetros puede causar la sobrestimación de modelos y por lo tanto de la pérdida de su capacidad descriptiva y/o predictiva para diferentes condiciones. Sin embargo, los resultados que se observan en las secciones anteriores son la respuesta de la selección de modos elementales y la construcción de ecuaciones con respecto a sólo los rendimientos del punto central 100:30 g/L de GLC:YE. Esto significa que, en esencia, el modelo es el mismo para los 9 experimentos y que la aproximación fue realizada sólo en los parámetros que regulan la combinación de dichos los EMs a través del tiempo, lo que sugiere que estos EMs son suficientes para describir el comportamiento metabólico dentro del rango de variación de las condiciones exploradas. Más aún,

las inferencias y la descripción metabólica no están basadas en los datos de flujo individuales para cada uno de los modelos, sino en el análisis y caracterización de las superficies de respuesta construidas para el flujo relativo al consumo de GLC. Lo anterior significa que, aun cuando uno de los experimentos pudiera contener algún rango de error o sobrestimación, la combinación de los 9 experimentos con el punto central por triplicado ayudaría a evitar su propagación a la descripción del metabolismo realizada. Finalmente, el análisis y la comparación de las superficies de los flujos metabólicos se relacionan de buena forma con las observaciones de los modelos fisiológicos presentados anteriormente y con el conocimiento acumulado para la cepa AR36, sus cepas parentales y el metabolismo de *Escherichia coli* en la literatura, indicando que dichas superficies describen adecuadamente el comportamiento de la cepa.

Con respecto al análisis de los modelos metabólicos, se encontró una inusual distribución de flujo en el nodo Pgi/G6Pdh, redirigiendo la mayoría del flujo de carbono a través de la vía de las PPP. Esta distribución es posible en la cepa AR36 debido a la alta sobreexpresión del gen *zwf* debida al promotor fuerte en el plásmido de alto número de copias reportada anteriormente (Rodríguez et al., 2013). Además, en los modelos metabólicos basados en análisis del balance de flujos (FBA) de Chen y colaboradores (2011) (Chen et al., 2011), se estableció a la enzima G6Pdh como la limitante para el flujo sobre la PPP y el control sobre este nodo. A pesar de esta distribución, se encontró un alto flujo glicolítico en las reacciones posteriores a la catalizada por GapA como fue descrito por Rodríguez y colaboradores en el 2017 para esta cepa, pero en condiciones de células en reposo (Rodríguez et al., 2017). En específico, el modelo predice un flujo relativo de entre el 75 al 88 % del carbono hacia PEP y hacia su conversión a PYR. En los resultados generados por los modelos, esta condición es posible debido a que la mayoría del flujo de carbono distribuido hacia la vía de las PPP es dirigido a través de la vía EDP hacia G3P y PYR, indicando que estos metabolitos son receptores de la mayoría del flujo glicolítico en la cepa, y podrían ser posibles nodos de control por IVM.

En relación a esta distribución de flujo, se ha reportado anteriormente que en cepas mutantes en *pykF* (como lo es AR36), los flujos sobre la vía de las PPP se incrementan hasta en un 79 % debido a la sobreexpresión de los genes *zwf*, *gnd* y *edd* y a la subsecuente subexpresión de los genes *pgi*, *pfkA* y *tpiA* (Escalante et al., 2010; Kabir y Shimizu, 2003; Rodriguez et al., 2017; Siddiquee et al., 2004). Además, en la cepa AR36, han sido reportadas concentraciones relativamente bajas de fructosa-1,6-difosfato (FDP) por medio de metabolómica comparativa (Martínez et al., 2008; Rodriguez et al., 2017). Esta baja concentración intracelular de FDP había sido explicada como consecuencia de la actividad de la enzima TktA. Sin embargo, los resultados obtenidos por los modelos metabólicos en este trabajo sugieren que podría deberse a la distribución de flujo encontrada hacia la vía de las PPP por la G6Pdh. Profundizando en la distribución alta sobre la PPP, mutantes sobre *pgi* han sido reportadas como cepas con velocidades de crecimiento bajas derivadas de la acumulación del cofactor redox NADPH. Sin embargo, también se ha demostrado que la sobreexpresión de rutas metabólicas consumidoras de esta molécula provoca que dichas cepas recobren su velocidad de crecimiento (Kabir y Shimizu, 2003). Por lo tanto, es posible que en AR36 al producir SA se pueda estar aliviando el desbalance redox sobre NADPH debido al consumo de este cofactor por la enzima shikimato deshidrogenasa codificada por *aroE*, promoviendo incluso velocidades de crecimiento altas con distribuciones de flujo de carbono elevadas para la vía de las PPP (Rodriguez et al., 2017). Esto sugiere que la producción de SA por el operón sintético expresado en esta cepa podría estar actuando como una ventaja de selección para obtener velocidades de crecimiento y de consumo altas (Rodriguez et al., 2017). La distribución de flujo sobre la vía de las PPP persiste para la fase estacionaria de las fermentaciones, pero en contraste con la fase de crecimiento se favorece el flujo de carbono hacia la vía no oxidativa de las pentosas por la enzima Gnd, incluso de forma casi total para algunas condiciones. Esto último permite a la cepa AR26 obtener rendimientos cercanos al 50 % para la producción de SA (Rodriguez et al., 2013).

A pesar de esto, los resultados indican que la E4P es el substrato limitante para la producción de SA incluso en la condición de alta sobreexpresión de *zwf* y distribución de flujo hacia PPP. Sin embargo, la sola sobreexpresión de este gen resulta

insuficiente debido a la salida de flujo a través de la vía EDP. Por este motivo, se propone a los genes *edd* y/o *eda* como blancos importantes para su eliminación, en conjunto con la sobreexpresión del gen *gnd*, para permitir una mayor proporción del flujo hacia E4P en condiciones de sobreexpresión de *zwf*. Esto podría evitar también la formación de G3P y la disminución del flujo hacia PEP por lo que un mejor control sobre la sobreexpresión de *zwf* resulta vital para tener distribuciones de flujo adecuadas sobre el nodo Pgi/G6Pdh. Continuando con la producción de SA en AR36, considerando la mayor abundancia de PEP con respecto a la E4P, la sobreexpresión del gen *ppsA* previamente reportada como útil para el aumento de la producción de este compuesto pensamos que no aumentaría la producción en esta cepa (Chen et al., 2014; Cui et al., 2014; Flores et al., 2004, 2005; Martínez et al., 2015). Más aún la reacción de la enzima codificada por este gen presentó valores de flujo cercanos a 0 en todas las condiciones experimentales y su posible sobreexpresión podría derivar en una velocidad de crecimiento y de consumo de GLC más bajas, al reducir el flujo de carbono hacia el TCA y otras vías biosintéticas derivadas del PYR.

Los análisis también revelaron una inusual distribución de flujo sobre PoxB en fase exponencial de crecimiento, consecuencia de los flujos incrementados hacia el PYR. Algunos estudios en *E. coli* han reportado a esta enzima como la principal productora de AC en velocidades altas de crecimiento en acelerostatos (Nahku et al., 2010). Además, esta distribución de flujo sobre PoxB y la importancia de esta enzima en células carentes del sistema PTS, también ha sido previamente propuesta en otros trabajos publicados para el linaje celular de la AR36 (Flores et al., 2004, 2005, 2002). Entre éstos se ha reportado que, para estas líneas celulares, la mutación existente en el sistema *arcA/arcB* podría ser la responsable de la sobreexpresión del gen *poxB* reportada durante su crecimiento exponencial. También, se ha reportado previamente en la literatura, que estas cepas sin el sistema PTS sobreexpresan los genes *acs* y *poxB* y por lo tanto sugirieron la aparición de un ciclo de reciclaje de carbono a través de las enzimas codificadas por estos genes (Aguilar et al., 2012; Flores et al., 2004, 2005; Sigala et al., 2008). Lo anterior concuerda con los resultados encontrados por el modelo, en el que se observa claramente dicho ciclo de reciclaje de carbono durante el crecimiento y en algunas condiciones experimentales en la fase estacionaria.

Finalmente, esta característica del metabolismo de AC en *E. coli* ha sido previamente propuesta por diversas investigaciones como respuesta a la saturación cinética del metabolismo celular y como reguladora de desbalances de flujo entre la glicólisis y el TCA (Flores et al., 2005, 2002; Matsuoka y Shimizu, 2013; Sigala et al., 2008).

Es importante hacer notar que el modelo no predice el consumo de AC a través de Acs durante el crecimiento, y que el modelo podría estar prediciendo la distribución incorrecta entre AckA-Pta y Acs en esta fase, tomando en cuenta la sobreexpresión de esta para líneas celulares derivadas de PB12 (Aguilar et al., 2012; Flores et al., 2004, 2005; Sigala et al., 2008). Esto es resultado de que en este nodo no se realizó ninguna restricción matemática o de rendimientos en el modelo, por lo que la restricción principal es el balance de la producción y consumo de ATP. En este sentido, en el modelo la reacción de Acs consume 2 moléculas de ATP, esto fue realizado como simplificación energética derivada de producción de AMP y no ADP por esta enzima. Sin embargo, esta simplificación podría no ser completamente adecuada para la correcta selección de la reacción de consumo por el modelo. Por otro lado, las cepas del linaje de AR36, no presentan disminución en las concentraciones intracelulares de AMP cíclico de mutantes carentes de PTS, debido a la sobreexpresión encontrada en estas cepas de la adenilato ciclasa y el ciclo de reciclaje de AC a través de Acs (Aguilar et al., 2012; Flores et al., 2004, 2005; Sigala et al., 2008). A pesar de esto, es interesante que los EMs utilizados para la fase estacionaria sí utilizan a Acs como la principal responsable del consumo de AC (Figuras 4.9 C y D), probablemente debido a una menor demanda de ATP durante esta fase por el modelo al no tener activa la reacción de producción de biomasa. Finalmente, como se observa en las secciones anteriores, fue posible validar en este trabajo, el ciclo reciclaje de carbono a través del AC producido por PoxB a través de la construcción de una mutante carente de PoxB. En esta cepa observamos la reducción de la producción de AC durante toda la fermentación y una disminución de la velocidad de consumo de GLC y de la velocidad de crecimiento. Más aún, la combinación de este ciclo de AC con las reacciones de exporte e importe del compuesto muestran las mismas zonas de aumento de las velocidades de producción de AC de las superficies de caracterización fisiológica para los metabolitos externos (Figura 4.2 I).

En cuanto a las distribuciones de flujo en el TCA, se encontró un inesperado flujo a través de la reacción reversa de la *IcdA* en la fase exponencial de crecimiento. Esto puede ser atribuido a la contribución del GLU entrando al TCA a través del  $\alpha$ -cetoglutarato (AKG) por la enzima Gdh. Sin embargo, este compuesto es también usado para la formación de precursores de biomasa, por lo que sólo si existe un exceso del consumo, este entraría al CCM. Con respecto a lo anterior, en el modelo se observa sólo una pequeña contribución de esta reacción hacia la producción de AKG. A pesar de esto, reportes en medios complejos han demostrado que *E. coli* favorece el consumo de los aminoácidos extracelulares y su catabolismo a través de AKG con la concomitante sobreexpresión de las vías de producción de bloque biosintéticos y la vía de la GSP (Arense et al., 2010; Lyubetskaya et al., 2006), efectos que observamos en las distribuciones de flujo modeladas. Además, la alta proporción de flujo sobre la GSP y las enzimas málicas sugieren que la cepa AR36 está tratando de contender con la saturación del metabolismo central y el flujo alto sobre PYR mediante el reciclaje de AC, y la mayor asimilación de ACCOA a través de TCA y GSP. Este tipo de metabolismo ha sido reportado como respuesta de contingencia para altas presiones osmóticas, de forma que *E. coli* pueda asimilar rápidamente substratos en condiciones de alta concentración y recuperar la velocidad de crecimiento (Arense et al., 2010). De hecho, en condiciones de estrés osmótico se ha comprobado el incremento de la actividad de las vías GSP a través de la reducción en la proporción de actividades de las enzimas *IcdA/Icl*, favoreciendo la producción de bloques para la producción de biomasa (Arense et al., 2010). De esta forma, el flujo alto a través de las reacciones anapleróticas y la vía GSP podría estar contribuyendo a mantener las altas producciones de biomasa y SA en las condiciones experimentales, mediante la reducción de la concentración de AC en la cepa AR36 y el balance de transportadores de electrones como el NADH y NADPH (Aguilar et al., 2012; Arense et al., 2010; Flores et al., 2004, 2005, 2002; Matsuoka y Shimizu, 2013; Rodriguez et al., 2017; Sigala et al., 2008).

De forma global, tanto las superficies de caracterización fisiológica como los modelos de flujo arrojaron resultados compatibles, que incluso permiten localizar las mismas zonas de preferencia de salidas de metabolitos en las diferentes etapas de la fermentación. Por ejemplo, para la fase de crecimiento en la zona de altas [GLC] y [YE] iniciales, predomina la producción de AC; en las condiciones de baja [GLC] y alta [YE] inicial se favorece la producción de biomasa; a alta [GLC] pero baja [YE] inicial, predomina la producción de SA; finalmente, en condiciones bajas de ambos substratos el crecimiento y la producción de SA es más balanceada evitando la producción de AC. En contraste, para la fase estacionaria las zonas de SA y AC se invierten ya que la producción favorable de SA se encuentra en la parte superior izquierda del área experimental y la zona de predominancia de producción de AC en el área inferior derecha. Particularmente, encontramos en la mayoría de las superficies puntos críticos, de silla, máximos y mínimos en las condiciones de entre 110-115 g/L iniciales de GLC y entre 35-15 g/L iniciales de YE, desde los cuales se delimitan los bordes del comportamiento metabólico como se describe en las secciones anteriores. Estos puntos críticos podrían estar delimitando la participación de estados metabólicos múltiples derivados de los cambios en las condiciones extracelulares (Namjoshi y Ramkrishna, 2001). Aunado a esto, los modelos dinámicos construidos por medio de la aproximación cibernética permitieron revelar comportamientos concuerdan con los datos fisiológicos observados y con el conocimiento disponible de esta cepa y sus precursoras. Es importante notar que los 9 EMs seleccionados fueron suficientes para describir todos los patrones conformados por las superficies mediante la modificación de pocos parámetros en las ecuaciones de velocidad (18 en total). Estos cambios, a pesar de que no pueden ser utilizados para describir mecánicamente los procesos regulatorios, sí resultan útiles para revelar características importantes de dichos procesos y su efecto en el metabolismo.



### 4.3. APLICACIÓN: PRODUCCIÓN DE ÁCIDO SHIKIMICO

#### 4.3.1. Mejoramiento de la producción de ácido shikímico mediante el uso de la información obtenida por modelación matemática.

Para evaluar la utilidad de los modelos previamente descritos con respecto al mejoramiento de la producción de SA, se diseñó un proceso de producción en modo lote alimentado. Las condiciones elegidas para este proceso conforme a los datos obtenidos fueron con 80 g/L de GLC y 40 g/L de YE de acuerdo con las consideraciones que se presentan a continuación. Las superficies revelaron las más altas velocidades de crecimiento en condiciones de alta [YE] y baja [GLC] inicial (Figura 4.2 K). Bajo estas condiciones también se encuentran las zonas con más bajas concentraciones finales de AC así como producciones media-altas de SA y altas concentraciones de biomasa (Figuras 4.2 C D y A) como productos finales de fermentación. A pesar de que el máximo título de SA fue encontrado cerca de la condición 110:40 GLC:YE en las superficies (Figura 4.2 C), esta condición también resulta en concentraciones más altas de producción de AC y menores velocidades de consumo y rendimiento en la fase estacionaria, las cuales resultan en un consumo incompleto de sustrato (Figuras 4.2 C D H y G). En contraste con esto, las velocidades más altas de producción de SA, consumo de glucosa y velocidades más bajas de producción de acetato fueron encontradas en condiciones iniciales cercanas a los 80:40 g/L GLC:YE (Figuras 4.2 H, G y J). Por lo que en estas condiciones la acumulación de biomasa a mayores velocidades de crecimiento podría ser encontradas sin comprometer las velocidades y rendimientos de producción de SA y consumo de GLC en fase estacionaria. Estas inferencias también son apoyadas por las superficies de respuesta metabólica, ya que en las condiciones iniciales de 80:40 GLC:YE se encuentra la zona de maximización para la producción de SA y consumo de glucosa marcada debido a las reacciones catalizadas por las enzimas Pgi, GapA, PykA, Gnd, TktA, DAHPs y PckA (Figuras 4.8 A,B,C,F,H,I,J y K). Además, bajo las condiciones seleccionadas, se presentan menores flujos relativos para las reacciones catalizadas por las enzi-

mas G6Pdh, PGdh y Ppc (Figuras 4.8 E,G y L). Lo anterior sugiere que, en estas condiciones, menos flujo de carbono es enviado a través de la ruta de EDP y por lo tanto es redirigido hacia E4P a través de Gnd y TktA en la fase estacionaria. De hecho, los datos de distribución de flujo hacia SA en estas condiciones indican una proporción de flujo cercana al 50% (Figura 4.10 C), mientras que el resto del flujo es redirigido por PykA y LpdA evitando la disminución de las velocidades de consumo. Más aún, los flujos de consumo de AC en estas condiciones se encuentran maximizados durante esta fase de fermentación con una concomitante reducción del flujo sobre PoxB, lo que disminuye la producción de AC (Figuras 4.9 B,C y E), lo que también se observa en las velocidades específicas de producción de acetato de los modelos fisiológicos (Figuras 4.2 J) . Por lo tanto, estas condiciones fueron escogidas a pesar de no encontrarse en la zona de maximización de la producción de SA en fase exponencial de crecimiento, donde se presenta una preferencia de la producción de biomasa, consumo de glucosa y un metabolismo glicolítico marcado por la zona de mayor flujo a través de GalP (Figura 4.10 A y B). A pesar de esto las velocidades de flujo a través de la reacción catalizada por la DAHPs hacia SA presentan valores de rango medio en las condiciones elegidas para el bioproceso (Figura 4.10 C).

Como se menciona en el párrafo anterior, la operación del proceso en modo lote alimentado, fue diseñado para dar preferencia a la producción de biomasa en la fase de crecimiento exponencial y a la producción de SA en fase estacionaria, lo que permite maximizar la productividad. La hipótesis de diseño utilizada, se basó en que el mantener las concentraciones extracelulares iniciales permitiría que los procesos de regulación derivados de la interacción (en el modelo las variables cibernéticas), mantuvieran constantes las concentraciones intracelulares y por tanto los perfiles de flujo metabólicos de acuerdo con las superficies construidas con los modelos. Además, al suspender la alimentación y entrar a la fase estacionaria, los perfiles de flujo derivarían de forma similar a las reportadas en las superficies metabólicas para esta fase en estas condiciones iniciales. Lo anterior significaría que se podría aumentar la concentración final de SA, manteniendo los rendimientos y evitando la producción de AC en la parte final del proceso de fermentación. El control fue realizado mediante el diseño de una alimentación pseudo-exponencial desde el inicio del cultivo. Dicha

alimentación contenía GLC y YE con la intención de mantener las concentraciones iniciales de estos sustratos el mayor tiempo posible. Sin embargo, es importante notar que como se ha señalado en todos los modelos, el YE fue considerado de forma simplificada por lo que el balance del flujo para este sustrato complejo fue sólo una aproximación.

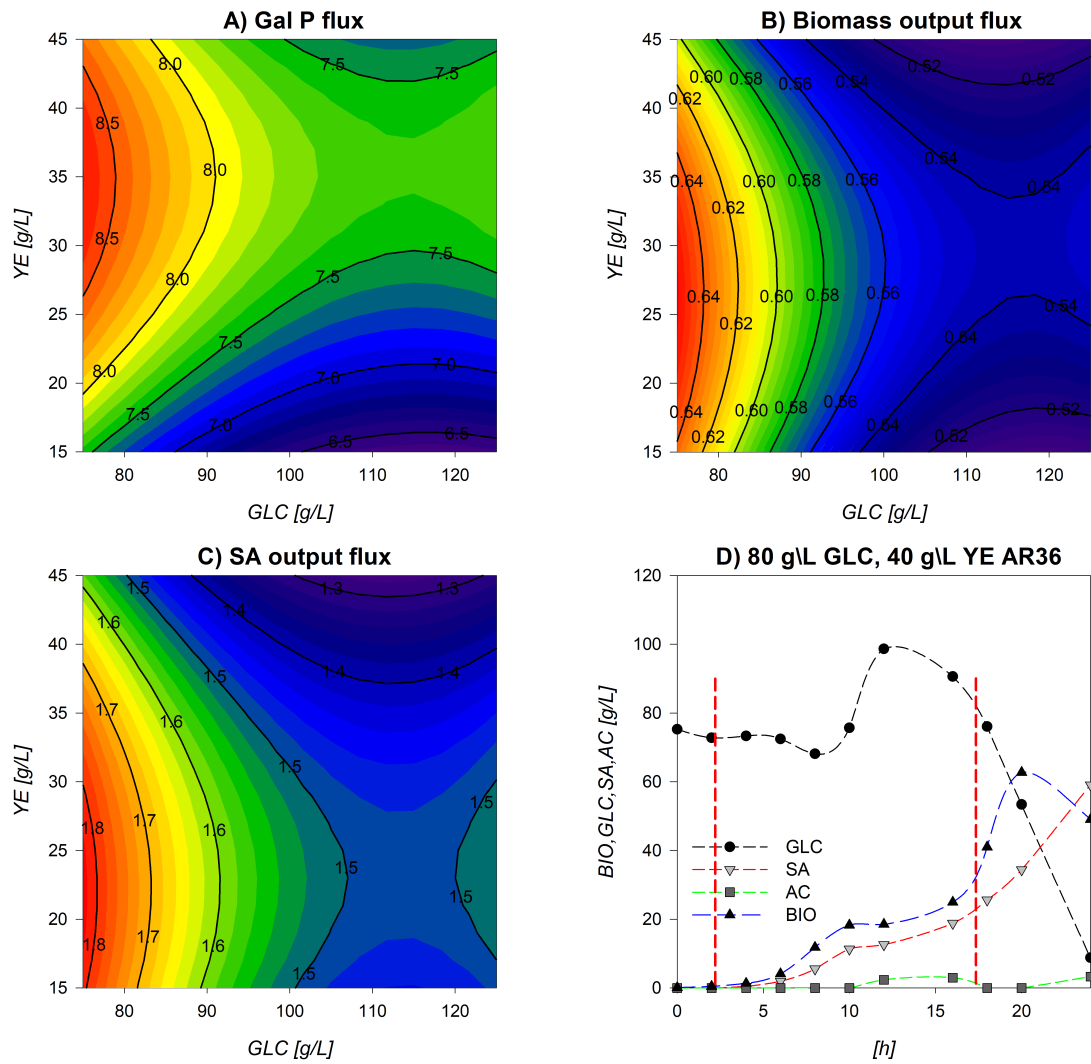


Figura 4.10: Perfil de fermentación del lote alimentado en condiciones iniciales de 80 g/L GLC y 40 g/L YE.

En la Figura 4.10 D se pueden observar los perfiles de fermentación para el proceso en modo lote alimentado diseñado para mejorar la producción de SA. La alimentación fue realizada desde la hora 3 debido a que en tiempos menores a este los flujos calculados de alimentación eran más bajos que la capacidad de la bomba peristáltica. La alimentación continuó desde este punto hasta la hora 18 controlada cada 15 minutos de forma manual para ajustarse a la velocidad de crecimiento y consumo de GLC calculadas. Se observa que la concentración de GLC fue mantenida en un rango de entre  $\approx 75$ -80 g/L durante las primeras 8 a 10 horas de alimentación, donde un incremento de la concentración de este compuesto fue observado aproximadamente a la hora 12 alcanzando los 100 g/L. Durante este proceso de acumulación, se observó también una disminución de la velocidad de crecimiento que pudiera ser posiblemente la causa de la disminución del consumo y acumulación de GLC. Este efecto podría ser atribuido en principio al control manual de la alimentación y a la simplificación del YE, lo que podría llevar a la sobreestimación o subestimación del flujo de algunos compuestos contenidos en el YE necesarios para el crecimiento. Lo anterior generaría desbalances entre la alimentación de GLC y YE pudiendo llevar a la disminución de la velocidad de crecimiento y el consumo de substratos. La primera velocidad de crecimiento presentó valores de  $0.90 \text{ h}^{-1}$ , las cuales están en el rango de las predichas por los modelos ( $0.8$ - $0.85 \text{ h}^{-1}$ ), mientras que la segunda velocidad de crecimiento fue de sólo  $0.18 \text{ h}^{-1}$  y presentó un comportamiento más lineal, sugiriendo la limitación por un substrato desconocido (probablemente derivado del YE). A pesar de esto, no queda clara la razón particular de estas disminuciones, por lo que futuras mejoras a los procesos de medición y control de la alimentación y de los metabolitos extracelulares en tiempo real podrían ser de utilidad para revelar la naturaleza de este fenómeno y mejorar la eficiencia del proceso.

Con respecto a la producción de SA, su velocidad específica también mostró una disminución en las horas 8-16 relacionadas con la disminución del consumo de GLC. A pesar de esto, las velocidades de consumo de GLC y producción de SA en fase estacionaria fueron mantenidas cercanas a las esperadas agotando la glucosa en tan sólo 24h. Durante el cultivo, un total de 180.5 g de GLC fueron consumidos para alcanzar un título total de 59.1 g/L de SA. Este título representa el más alto

reportado para esta cepa. Sin embargo, aún es 30 % menor al máximo reportado para producciones con *E. coli* por Chandran et al. (2003). Por otro lado, la concentración de AC en el fermentador nunca presentó valores superiores a los 5 g/L, probando que el diseño logró limitar su producción incluso en estas condiciones atípicamente altas de concentración y alimentación de substratos, comparados con fermentaciones en modo lote, donde >15 g/L AC pueden ser acumulados. Más aún, el proceso presentó productividades volumétricas del orden de los 2.45 g SA/L\*h lo que representa un incremento del 70 % comparado con la previamente reportada para esta cepa por Rodríguez y colaboradores (1.43 gSA/L\*h (Rodriguez et al., 2013) y 20 % mayor a la cepa industrial reportada (2.04 gSA/L\*h Chandran y colaboradores (Chandran et al., 2003)). Crucialmente, los rendimientos calculados por medio de aproximación lineal presentaron valores de 0.40 para  $Y_{p/s}$  y 0.74 para  $Y_{p/x}$ , lo que significa que ambos rendimientos fueron mantenidos a lo largo de la fermentación relativos a los obtenidos en fermentadores en modo lote. Todo esto sugiere que la distribución de flujos de carbono, en los nodos relevantes para la producción de SA, a lo largo del metabolismo se mantuvieron conforme a las superficies calculadas de forma razonable y de acuerdo con los rendimientos encontrados en trabajos previos para cepas carentes del sistema PTS (Flores et al., 2004, 2005, 2002; Martínez et al., 2015, 2008; Rodriguez et al., 2013, 2014).

---

## Capítulo 5

# CONCLUSIONES, PERSPECTIVAS y PRODUCTOS

## 5.1. CONCLUSIONES

En este reporte describimos el estudio de una cepa de *Escherichia coli* PTS<sup>-</sup> carente del sistema PTS y modificada por IVM para la sobreproducción de SA, bajo una aproximación de modelación dinámica de su comportamiento fisiológico y metabólico en diferentes condiciones iniciales de sustratos (Aguilar et al., 2012; Flores et al., 2004, 2005, 1996, 2002; Rodriguez et al., 2013, 2017). Los modelos construidos fueron capaces de seguir con buena precisión los datos experimentales de los perfiles de fermentación obtenidos con esta cepa. De esta forma se pudieron obtener parámetros descriptores que simulan y predicen el comportamiento del consumo y producción final de biomasa, GLC, SA y AC, así como del metabolismo central de carbono basados en porcentajes de flujo normalizados al consumo de GLC. Con estos parámetros fue posible construir superficies de respuesta basadas en una ecuación polinomial, las cuales resultaron útiles para caracterizar de forma morfológica el comportamiento de la cepa en variaciones de las concentraciones iniciales de sustratos en las diferentes etapas de las fermentaciones. Además, permitieron revelar características importantes del comportamiento metabólico de AR36. Se encontró en este sentido, que la cepa AR36 responde de forma distinta a las diferentes concentraciones de sustrato variando la designación de recursos y flujos de carbono aun cuando presenta velocidades de consumo similares. Esta inferencia además se puede observar en las superficies de respuesta donde la morfología presenta en la mayoría de los casos contribuciones no lineales a cada una de las salidas metabólicas.

Los modelos dinámicos construidos tanto para la descripción del comportamiento de los metabolitos extracelulares basados en ecuaciones logísticas, como los modelos cibernéticos para las distribuciones de flujo intracelular presentaron resultados similares y concordantes. Además, sus inferencias y análisis estuvieron de acuerdo con la información y datos previamente reportados para la cepa AR36 y su linaje carente del sistema PTS (Aguilar et al., 2012, 2018; Flores et al., 2004, 2005, 1996; Rodriguez et al., 2013, 2017; Sigala et al., 2008). De esta forma las superficies construidas resultaron útiles para localizar zonas preferenciales de salidas y flujos metabólicos dentro del área experimental caracterizada.

Las superficies de flujo metabólico ayudaron a explicar comportamientos previamente observados para la cepa AR36, como la baja concentración intracelular de fructosa-1.6-bifosfato (FDP) en condiciones de fermentación de altas cantidades de sustrato (Rodríguez et al., 2013), donde los modelos revelaron una posible distribución de flujo alto sobre la vía de las PPP durante todo el proceso de fermentación. Además, determinaron que el desbalance de NADPH producido por esta distribución de flujos se podría ver compensado por la producción de SA, contribuyendo a mantener la velocidad de crecimiento y el consumo de glucosa (Rodríguez et al., 2013). El análisis de las superficies de flujo también reveló el uso de ciclos extendidos del metabolismo central de carbono a través de las vías de producción y consumo de AC, GSP y demás vías anapleróticas en la cepa AR36 para contender con el estrés metabólico derivado de las altas concentraciones de sustrato en el medio. La reacción catalizada por PoxB se determinó como la principal para la producción de AC en la cepa AR36 durante toda la fermentación. Se construyó una cepa derivada carente de esta enzima, y se cultivó en condiciones de producción preferente de AC, mostrando que su inactivación reduce la producción de acetato en conjunto con la velocidad de consumo de glucosa y el crecimiento, probablemente debido a una sobre acumulación del PYR (Flores et al., 2004, 2005, 1996, 2002; Rodríguez et al., 2013, 2017; Sigala et al., 2008).

El análisis de las superficies de respuesta también resultó útil para la determinación de blancos de modificación genética para mejorar la producción de SA en AR36. En este sentido se determinó que los genes *edd* y/o *eda* podrían ser eliminados en conjunto con un mejor control de la sobreexpresión de los genes *zwf* y *gnd* para obtener mejores distribuciones de flujo hacia la producción de E4P y su equilibrio con PEP para la producción de SA. Además, se determinó que en esta cepa particular la sobreexpresión del gen *ppsA* y otras modificaciones dirigidas a la acumulación de PEP podrían no resultar útiles para la producción de SA, y que inclusive podrían resultar perjudiciales para el crecimiento y la aptitud celular.



Finalmente, para evaluar la utilidad de los modelos y superficies construidas en la presente investigación con respecto a la producción de SA, se diseñó un proceso de producción en modo lote alimentado. Dicho proceso fue diseñado para mejorar la productividad y título final del proceso, minimizar la producción de AC y mantener los rendimientos reportados hasta el momento para esta cepa. Un formato de alimentación inusual fue diseñado para mantener las condiciones iniciales de fermentación, lo que en teoría ayudaría a mantener constantes las propiedades metabólicas y regulativas del sistema. Este proceso resultó en un aumento del 40 % del título y 70 % de la productividad limitando el AC a valores sorprendentemente menores a 5 g/L. Los títulos y rendimientos obtenidos en este trabajo son los más altos reportados para esta cepa. Los modelos implementados aquí, representan la primera aproximación para establecer distribuciones de flujo metabólico para la cepa AR36 en condiciones de alta concentración de los substratos GLC y YE. Así mismo representa uno de los pocos esfuerzos reportados hacia la modelación para ahondar en el entendimiento del metabolismo de *E. coli* en medios complejos y en condiciones dinámicas.

## 5.2. PERSPECTIVAS

Las contribuciones realizadas en el presente trabajo de investigación derivaron además en diferentes puntos necesarios a investigar para que se pueda incrementar de forma significativa el conocimiento sobre la regulación y comportamiento metabólico de la cepa *E. coli* AR36 y otras derivadas similares a esta en el laboratorio del Dr. Francisco Bolivar. Por tanto, dentro de las perspectivas del trabajo se encuentran:

- Realizar las modificaciones genéticas en los blancos genéticos propuestos. Inactivación de los genes *edd* y *eda* y sobreexpresión de *gnd* en AR36 buscando aumentar el rendimiento de SA. Modular la expresión del gen *poxB* con un promotor controlable que permita reducir el nivel de la enzima PoxB.
- Estudio transcriptómico en las diferentes secciones de interés del diseño experimental para encontrar información sobre las respuestas regulatorias a nivel de DNA/RNA

- Estudio proteómico en las diferentes secciones de interés del diseño experimental para encontrar información sobre las respuestas regulatorias a nivel RNA/Proteína
- Extensión de la red metabólica a escala genómica y establecimiento de protocolos de reducción de EMs eficientes para el uso de modelación cibernética
- Extensión del diseño experimental a otras variables de interés, como velocidad de transferencia de oxígeno, pH y temperatura, para la generación de superficies de respuesta utilizando los modelos fisiológicos.
- Estudio de los componentes y perfiles de consumo de aminoácidos por UPLC provenientes del YE
- Optimización de los perfiles de alimentación y formulaciones de soluciones para el aumento de la producción de SA en modo lote alimentado con la cepa AR36.

## 5.3. PRODUCTOS DEL TRABAJO

### 5.3.1. Publicaciones científicas

De este trabajo se publicaron 5 artículos en los cuales el M.C. Juan A. Martínez es primer autor de dos y segundo autor de los otros 3. El artículo principal resultante de la investigación se titula “**Metabolic modeling and response surface analysis of an *Escherichia coli* strain engineered for shikimic acid production**”, publicado en la revista BMC System biology. Dos de los cuatro artículos restantes son ligados directamente al trabajo de investigación doctoral y resultaron además en la invitación de escritura de los otros 2, siendo éstos revisiones relacionadas con el área de investigación, donde uno fue elegido para una sección de un libro electrónico de concentración de tópicos de investigación. Los artículos se listan a continuación y se adjuntan como anexos al final del documento.

- Juan A. Martínez, Alberto Rodríguez, Fabian Moreno, Noemí Flores, Alvaro R. Lara, Octavio T. Ramírez, Guillermo Gosset and Francisco Bolívar. **Metabolic modeling and response surface analysis of an *Escherichia coli* strain**

**engineered for shikimic acid production.** BMC Systems Biology, 12:102, 2018.

- Alberto Rodríguez, Juan A. Martínez, Pierre Millard, Guillermo Gosset, Jean-Charles Portais, Fabien Létisse and Francisco Bolívar. **Plasmid-encoded biosynthetic genes alleviate metabolic disadvantages while increasing glucose conversion to shikimate in an engineered *Escherichia coli* strain.** Biotechnology and Bioengineering, 114:6, 2017.
- Juan A. Martínez, Francisco Bolívar and Adelfo Escalante. **Shikimic acid production in *Escherichia coli*: from classical metabolic engineering strategies to omics applied to improve its production.** Frontiers in Bioengineering and Biotechnology , 3:145, 2015.
- El artículo anterior es escogido para formar parte del libro electrónico “**Research topic: Quantitative Systems Biology for Engineering Organisms y Pathways**” que contiene 11 artículos y 34 autores. Frontiers, 2016.
- Alberto Rodríguez, Juan A. Martínez, Noemí Flores, Adelfo Escalante, Guillermo Gosset and Francisco Bolívar. **Engineering *Escherichia coli* to overproduce aromatic amino acids and derived compounds.** Microbial Cell Factories, 13:126, 2014.
- Alberto Rodríguez, Juan A. Martínez, José L. Báez-Viveros, Noemí Flores, Georgina Hernández-Chávez, Octavio T. Ramírez, Guillermo Gosset and Francisco Bolívar. **Constitutive expression of selected genes from the pentose phosphate y aromatic pathways increases the shikimic acid yield in high-glucose batch cultures of an *Escherichia coli* strain lacking PTS and *pykF*.** Microbial Cell Factories, 12:86, 2013.

### 5.3.2. Trabajo de apoyo a docencia y desarrollo profesional de estudiantes.

Durante el tiempo de este trabajo también se realizaron apoyos a docencia mediante la coordinación y la impartición de parte del temario del curso de “**Introducción a la Ingeniería de vías metabólicas**”. Además, se apoyó a la conceptualización, desarrollo y realización de la tesis de licenciatura del alumno David Rodríguez Lozada bajo el título: “**Efecto de la expresión del gen de hemoglobina de *Vitreoscilla stercoraria* en el metabolismo y producción de shikimato de la cepa sobreproductora *Escherichia coli* PB12.AR36**” durante el 2015.

### 5.3.3. Colaboraciones con revistas indexadas, industria y apoyo a actividades de investigación.

Durante el tiempo del trabajo de doctorado se participó además en un proyecto CONACYT de desarrollo y transferencia de tecnología con una empresa instalada en el valle de México bajo un proyecto PROINNOVA, dentro del cual participó el grupo del Dr. Agustín López-Munguía en el año 2016. Dentro de este proyecto el M.C. Juan Andrés Martínez participó como responsable de investigación por parte de la compañía y siendo el enlace para el desarrollo de las secciones correspondientes al proyecto en el Instituto de Biotecnología de la UNAM.

Además, el M.C. Juan Andrés Martínez participa como revisor activo de la revista internacional indexada “Ciencias y Tecnologías del Agua” perteneciente al “Instituto Mexicano de Tecnologías del Agua” y fue invitado a pertenecer al comité editorial durante el periodo 2015-2018.

---

# Bibliografía

Aguilar, C., Escalante, A., Flores, N., de Anda, R., Riveros-Mckay, F., Gosset, G., Morett, E., y 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*, 13:1–17. doi:10.1186/1471-2164-13-385.

Aguilar, C., Martínez-Batallar, G., Flores, N., Moreno-Avitia, F., Encarnación, S., y Escalante, A. (2018). Analysis of differentially upregulated proteins in *ptshicr*<sup>-</sup> and *rp-ph*<sup>-</sup> mutants in *Escherichia coli* during an adaptive laboratory evolution experiment. *Applied Microbiology and Biotechnology*, 102(23):10193–10208.

Ahn, J. O., Lee, H. W. L., Saha, R., Park, M. S. P., Jung, J. K., y Lee, D. Y. (2008). Exploring the effects of carbon sources on the metabolic capacity for shikimic acid production in *Escherichia coli* using in silico metabolic predictions. *Journal of Microbiology and Biotechnology*, 18:1773–1784. doi:10.4014/jmb.0700.705.

Akesson, M., Forster, J., y Nielsen, J. (2004). Integration of gene expression data into genome-scale metabolic models. *Metabolic Engineering*, 6:284–293. doi:10.1016/j.ymben.2003.12.002.

Arense, P., vicente Bernal, Iborra, J. L., y Cánovas, M. (2010). Metabolic adaptation of *Escherichia coli* to long-term exposure to salt stress. *Process Biochemistry*, 45:1495–1467. doi:10.1016/j.procbio.2010.05.022.

Báez, J. L., Bolivar, F., y Gosset, G. (2001). Determination of 3-deoxy-darabinoheptulosonate 7-phosphate productivity and yield from glucose in *Escherichia coli*

- devoid of the glucose phosphotransferase transport system. *Biotechnology and Bioengineering*, 73:530–535.
- Báez-Viveros, J. L., Osuna, J., Hernández-Chávez, G., Soberon, X., Bolivar, F., y Gosset, G. (2004). Metabolic engineering and protein directed evolution increase the yield of l-phenylalanine synthesized from glucose in *Escherichia coli*. *Biotechnology and Bioengineering*, 87:516–524. doi:10.1002/bit.20159.
- Biggs, B. W., Paepe, B. D., Santos, C. N. S., Mey, M. D., y Ajikumar, P. K. (2014). Multivariate modular metabolic engineering for pathway and strain optimization. *Current Opinion in Biotechnology*, 29:156–162. doi:10.1016/j.copbio.2014.05.005.
- Bongaerts, J., Krämer, M., Müller, U., Raeven, L., y Wubbolts, M. (2001). Metabolic engineering for microbial production of aromatic amino acids and derived compounds. *Metabolic Engineering*, 3:289–300. doi:10.1006/mben.2001.0196.
- Bradley, D. (2005). Star role for bacteria in controlling flu pandemic? *Nature Reviews Drug Discovery*, 4(12):945–946.
- Causey, T. B., Shanmugam, K. T., Yomano, L. P., y Ingram, L. O. (2004). Engineering *Escherichia coli* for efficient conversion of glucose to pyruvate. *Proceedings of the National Academy of Sciences of the United States of America*, 8:2235–2240. doi:10.1073/pnas.0308171100.
- Chandran, S. S., Yi, J., Draths, K., von Daeniken, R., Weber, W., y Frost, J. (2003). Phosphoenolpyruvate availability and the biosynthesis of shikimic acid. *Biotechnology Progress*, 19:808–814. doi:10.1021/bp025769p.
- Chen, K., Dou, J., Tang, S., Yang, Y., Wang, H., Fang, H., y Zhou, C. (2012). Deletion of the *aroK* is essential for high shikimic acid accumulation through the shikimate pathway in *E. coli*. *Bioresource Technology*, 119:141–147. doi:10.1016/j.biortech.2012.05.100.
- Chen, P. T., Chiang, C.-J., Wang, J.-Y., Lee, M.-Z., y Chao, Y.-P. (2011). Genomic engineering of *Escherichia coli* for production of intermediate metabolites in the aromatic pathway. *Journal of the Taiwan Institute of Chemical Engineers*, 42:34–40. doi:10.1016/j.jtice.2010.03.010.

- Chen, X., Li, M., Zhou, L., Shen, W., Algasan, G., Fan, Y., y Wang, Z. (2014). Metabolic engineering of *Escherichia coli* for improving shikimate synthesis from glucose. *Bioresource Technology*, 166:64–71. doi:10.1016/j.biortech.2014.05.035.
- Chen, X., Zhou, L., Tian, K., Kumar, A., Singh, S., Prior, B. A., y Wang, Z. (2013). Metabolic engineering of *Escherichia coli*: a sustainable industrial platform for bio-based chemical production. *Biotechnology Advances*, 31:1200–1223. doi:10.1016/j.biotechadv.2013.02.009.
- Chávez, M. I., Martínez, A., Bolívar, F., y Gosset, G. (2005). Metabolic pathway engineering for microbial production of aromatic amino acids. *Res. Adv. in Food Science*, 5:11–20.
- Cloots, L. y Marchal, K. (2011). Network-based functional modeling of genomics, transcriptomics and metabolism in bacteria. *Current Opinion in Microbiology*, 14:599–607. doi:10.1016/j.mib.2011.09.003.
- Covert, M. W., Schilling, C. H., Famili, I., Edwards, J. S., Goryanin, I. I., Selkov, E., y Palsson, B. O. (2001a). Metabolic modeling of microbial strains *in silico*. *Trends in Biochemical Sciences*, 26:179–186. doi:10.1016/S0968-0004(00)01754-0.
- Covert, M. W., Schilling, C. H., y Palsson, B. (2001b). Regulation of gene expression in flux balance models of metabolism. *Journal of Theoretical Biology*, 213(1):73–88.
- Cui, Y.-Y., Ling, C., Zhang, Y.-Y., Huang, J., y Liu, J.-Z. (2014). Production of shikimic acid from *Escherichia coli* through chemically inducible chromosomal evolution and cofactor metabolic engineering. *Microbial Cell Factories*, 13:1–11. doi:10.1186/1475-2859-13-21.
- Díaz-Quiroz, D. C., Carmona, S. B., Bolívar, F., y Escalante, A. (2014). Current perspectives on applications of shikimic and aminoshikimic acids in pharmaceutical chemistry. *Research and Reports in Medicinal Chemistry*, 4:35–46. doi:10.2147/RRMC.S46560.

- Draths, K., Knop, D. R., y Frost, J. (1999). Shikimic acid and quinic acid: Replacing isolation from plant sources with recombinant microbial biocatalysis. *Journal Of American Chemical Society*, 121:1603–1604.
- Escalante, A., Calderón, R., Araceli Valdivia, R. d. A., Hernández, G., Ramírez, O. T., Gosset, G., y Bolívar, F. (2010). Metabolic engineering for the production of shikimic acid in an evolved *Escherichia coli* strain lacking the phosphoenolpyruvate: carbohydrate phosphotransferase system. *Microbial Cell Factories*, 9:1–12. doi:10.1186/1475-2859-9-21.
- Estévez, A. M. y Estévez, R. J. (2012). A short overview on the medicinal chemistry of (-)shikimic acid. *Mini-Reviews in Medicinal Chemistry*, 12:1443–1454. doi:10.2174/138955712803832735.
- Ferreira, S. L., Dos Santos, W. N., Quintella, C. M., Neto, B. B., y Bosque-Sendra, J. M. (2004). Doehlert matrix: A chemometric tool for analytical chemistry - Review. *Talanta*, 63(4):1061–1067.
- Flores, N., de Anda, R., Flores, S., Escalante, A., Hernández, G., Martínez, A., Ramírez, O. T., Gosset, G., y Bolívar, F. (2004). Role of pyruvate oxidase in *Escherichia coli* strains lacking the phosphoenolpyruvate: carbohydrate phosphotransferase system. *Journal of Molecular Microbiology and Biotechnology*, 8:209–221. doi:10.1159/000086702.
- Flores, N., Escalante, A., De Anda, R., Báez-Viveros, J. L., Merino, E., Franco, B., Georgellis, D., Gosset, G., y Bolívar, F. (2008). New insights into the role of sigma factor rpos as revealed in escherichia coli strains lacking the phosphoenolpyruvate:carbohydrate phosphotransferase system. *Journal of Molecular Microbiology and Biotechnology*, 14(4):176–192.
- Flores, N., Flores, S., Escalante, A., de Anda, R., Leal, L., Malpica, R., Georgellis, D., Gosset, G., y 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. *Metabolic Engineering*, 7:70–87. doi:10.1016/j.ymben.2004.10.002.



- Flores, N., Leal, L., Sigala, J. C., De Anda, R., Escalante, A., Martínez, A., Ramírez, O. T., Gosset, G., y Bolívar, 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. *Journal of Molecular Microbiology and Biotechnology*, 13(1-3):105–116.
- Flores, N., Xiao, J., Berry, A., Bolívar, F., y Valle, F. (1996). Pathway engineering for the production of aromatic compounds in *Escherichia coli*. *Nature Biotechnology*, 14:620–623.
- Flores, S., Gosset, G., Flores, N., de Graaf, A., y Bolívar, F. (2002). Analysis of carbon metabolism in *Escherichia coli* strains with an inactive phosphotransferase system by  $^{13}\text{C}$  labeling and nmr spectroscopy. *Metabolic Engineering*, 4:124–137. doi:10.1006/mben.2001.0209.
- Fong, S. S. (2014). Computational approaches to metabolic engineering utilizing systems biology and synthetic biology. *Computational and Structural Biotechnology Journal*, 11:28–34. doi:10.1016/j.csbj.2014.08.005.
- Frost, J. W., Frost, K. M., y Knop, D. R. (2002). Biocatalytic synthesis of shikimic acid. *United States Patent*, (20026472169B1).
- Ghosh, S., Chisti, Y., y Banerjee, U. C. (2012). Production of shikimic acid. *Biotechnology Advances*, 30:1425–1431. doi:10.1016/j.biotechadv.2012.03.001.
- Gosset, G. (2005). Improvement of *Escherichia coli* production strains by modification of the phosphoenolpyruvate:sugar phosphotransferase system. *Microbial Cell Factories*, 4:1–11.
- Gosset, G. (2009). Production of aromatic compounds in bacteria. *Current Opinion in Biotechnology*, 20:651–658. doi:10.1016/j.copbio.2009.09.012.
- Hernandez-Montalvo, V., Valle, F., Bolívar, F., y Gosset, G. (2001). Characterization of sugar mixtures utilization by and *Escherichia coli* mutant devoid of the phosphoenolpyruvate system. *Appl Microbiol Biotechnol*, 57:186–191.

- Herrmann, K. M. y Weaver, L. M. (1999). The shikimate pathway. *Annual Review of Plant Physiology and Plant Molecular Biology*, 50(1):473–503.
- Huang, F., Xiu, Q., Sun, J., y Hong, E. (2002). Anti-platelet and anti-thrombotic effects of triacetylshikimic acid in rats. *Journal of Cardiovascular Pharmacology*, 39(2):262–270.
- Johansson, L. y Lidén, G. (2006). Transcriptome analysis of a shikimic acid producing strain of *Escherichia coli* w3110 grown under carbon- and phosphate-limited conditions. *Journal of Biotechnology*, 126(4):528–545.
- Jouhten, P. (2012). Metabolic modelling in the development of cell factories by synthetic biology. *Computational and Structural Biotechnology Journal*, 3:e201210009. doi:10.5936/csbj.201210009.
- Kabir, M. M. y Shimizu, K. (2003). Gene expression patterns for metabolic pathway in *pgi* knockout *Escherichia coli* with and without *phb* genes based on rt-pcr. *Journal of Biotechnology*, 105:11–31. doi:10.1016/S0168-1656(03)00170-6.
- Kao, R. Y., Yang, D., Lau, L. S., Tsui, W. H., Hu, L., Dai, J., Chan, M. P., Chan, C. M., Wang, P., Zheng, B. J., Sun, J., Huang, J. D., Madar, J., Chen, G., Chen, H., Guan, Y., y Yuen, K. Y. (2010). Identification of influenza A nucleoprotein as an antiviral target. *Nature Biotechnology*, 28(6):600–605.
- Keseler, I., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martínez, C., Fulcher, C., Huerta, A., Kothari, A., y Krummenacker, M. (2013). Ecocyc: fusing model organism databases with systems biology. *Nucleic Acids Res.*, 41(Database issue):D605–D612.
- Kim, T. Y., Sohn, S. B., Kim, Y. B., Kim, W. J., y Lee, S. Y. (2012). Recent advances in reconstruction and applications of genome-scale metabolic models. *Current Opinion in Biotechnology*, 23:617–623. doi:10.1016/j.copbio.2011.10.007.
- Knop, D. R., Draths, K., Chandran, S. S., Baker, J. L., von Daeniken, R., Weber, W., y Frost, J. (2001). Hydroaromatic equilibration during biosynthesis of shikimic acid. *Journal of American Chemical Society*, 123:10173–10182.

- Kompala, D. S., Ramkrishna, D., Jansen, N. B., y Tsao, G. T. (1986). Investigation of bacterial growth on mixed substrates: Experimental evaluation of cybernetic models. *Biotechnology and Bioengineering*, 28:1044–1055. doi:10.1002/bit.260280715.
- Krämer, M., Bongaerts, J., Bovenberg, R., Kremer, S., Müller, U., Org, S., Wubbolts, M., y Raeven, L. (2003). Metabolic engineering for microbial production of shikimic acid. *Metabolic Engineering*, 5:277–283. doi:10.1016/j.ymben.2003.09.001.
- Lara, A. R., Caspeta, L., Gosset, G., Bolivar, F., y Ramírez, O. T. (2007). 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. *Biotechnology and Bioengineering*, 99:893–901. doi:10.1002/bit.21664.
- Leib, T. M., Pereira, C. J., y Villadsen, J. (2001). Bioreactors: A chemical engineering perspective. *Chemical Engineering Science*, 56(19):5485–5497.
- Li, S., Yuan, W., Wang, P., Zhang, Z., Zhang, W., y Ownby, S. (2007). Method for the extraction and purification of shikimic acid.
- Lin, S., Liang, R., Menx, X., OuYang, H., Yan, H., Wang, Y., y Jones, G. S. (2014). Construction and expression of mutagenesis strain of aroG gene from *Escherichia coli* k-12. *International Journal of Biological Macromolecules*, 68:173–177. doi:10.1016/j.ijbiomac.2014.04.034.
- Liu, Q., Yongsong Cheng, X. X., Xu, Q., y Chen, N. (2012). Modification of tryptophan transport system and its impact on production of l-tryptophan in *Escherichia coli*. *Bioresource Technology*, 114:549–554. doi:10.1016/j.biortech.2012.02.088.
- Long, M. R., Ong, W. K., y Reed, J. L. (2015). Computational methods in metabolic engineering for strain design. *Current Opinion in Biotechnology*, 34:135–141. doi:10.1016/j.copbio.2014.12.019.
- Luli, G. y Stohl, W. (1990). Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations. *Applied and Environmental Microbiology*, 4:1004–1011. doi:0099-2240/90/041004-08\$02.00/0.

- Lyubetskaya, A. V., Rubanov, L. I., y Gelfand, M. S. (2006). Use of the flux model of amino acid metabolism of *Escherichia coli*. *Biochemistry*, 71:1256–1260.
- Ma, Y., Sun, J., Xu, Q., y Guo, Y. (2004). Inhibitory effects of shikimic acid on platelet aggregation and blood coagulation. *Acta Pharmaceutica Sinica*, 35(1):1–3.
- Machado, D. y Herrgard, M. (2014). Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLOS Computational Biology*, 10:e1003580. doi:10.1371/journal.pcbi.1003580.
- Martínez, J. A., Bolívar, F., y Escalante, A. (2015). Shikimic acid production in *Escherichia coli*: from classical metabolic engineering strategies to omics applied to improve its production. *Frontiers in Bioengineering and Biotechnology*, 3:1–16. doi:10.3389/fbioe.2015.00145.
- Martínez, K., de Anda, R., Hernández, G., Escalante, A., Ramírez, O., y Bolívar, F. (2008). Couitilization of glucose and glycerol enhances the production of aromatic compounds in an *Escherichia coli* strain lacking the phosphoenolpyruvate: carbohydrate phosphotransferase system. *Microbial Cell Factories*, 22:1–12. doi:10.1186/1475-2859-7-1.
- Matsuoka, Y. y Shimizu, K. (2013). A new insight into the main metabolic regulation of *Escherichia coli* based on systems biology approach. *12th IFAC Symposium on Computer Applications in Biotechnology*, 12:16–18.
- Nahku, R., Valgepea, K., Lahtve, P.-J., Erm, S., Abner, K., Adamberg, K., y Vilu, R. (2010). Specific growth rate dependent transcriptome profiling of *Escherichia coli* k12 mg16555 in accelerostat cultures. *Journal of Biotechnology*, 145:60–65. doi:10.1016/j.jbiotec.2009.10.007.
- Namjoshi, A. A. y Ramkrishna, D. (2001). Multiplicity and stability of steady states in continuous bioreactors: Dissection of cybernetic models. *Chemical Engineering Science*, 56:5593–5607. doi:10.1016/S0009-2509(01)00166-X.
- Neumann, G., Noda, T., y Kawaoka, Y. (2009). Emergence and pandemic potential of swine-origin h1n1 influenza virus. *Nature*, 459(7249):931–939.

- Nie, L.-d., Shi, X.-x., Ko, K. H., y Lu, W.-d. (2009). A short and practical synthesis of oseltamivir phosphate ( tamiflu ) from ( - ) -shikimic acid acid through a short and practical synthetic route via eight steps in 47 % overall yield . in addition , the highly regiose- lective and stereoselective nucle. *J. Org. Chem*, 74(10):3970–3973.
- O'Brien, E. J. y Palsson, B. O. (2015). Computing the functional proteome: recent progress and future prospects for genome-scale models. *Current Opinion in Biotechnology*, 34:125–134. doi:10.1016/j.copbio.2014.12.017.
- Patil, K. R., Akesson, M., y Nielsen, J. (2004). Use of genome-scale microbial models for metabolic engineering. *Current Opinion in Biotechnology*, 15:64–69. doi:10.1016/j.copbio.2003.11.003.
- Patnaik, R. y Liao, J. C. (1994). Engineering of escherichia coli central metabolism for aromatic metabolite production with near theoretical yield. *Applied and Environmental Microbiology*, 60(11):3903–3908.
- Patnaik, R., Spitzer, R. G., y Liao, J. C. (1995). Pathway engineering for production of aromatic in *Escherichia coli*: Confirmation of stoichiometric analysis by independent modulation of arog, tkta, and pps activities. *Biotechnology and Bioengineering*, 46:361–370. doi:10.1002/bit.260460409.
- Polen, T., Krämer, M., Bongaerts, J., Wubbolts, M., y Wendisch, V. F. (2005). The global gene expression response of escherichia coli to l-phenylalanine. *Journal of Biotechnology*, 115(3):221–237.
- Price, N. D., Papin, J. A., Schilling, C. H., y Palsson, B. O. (2003). Genome-scale in microbial *in silico* models: the constraints-based approach. *Trends in Biotechnology*, 21:162–169. doi:10.1016/S0167-7799(03)00030-1.
- Raghavendra, T. R., Vaidyanathan, P., Swathi, H., Ravikanth, G., Ganeshiah, K. N., Srikrishna, A., y Shaanker, R. U. (2009). Prospecting for alternate sources of shikimic acid, a precursor of tamiflu, a bird-flu drug. *Current Science*, 96:771–772. Scientific Correspondence.

- Ramkrishna, D. y Song, H.-S. (2012). Dynamic models of metabolism: Review of the cybernetic approach. *Bioengineering, Food, and Natural Products*, 58:986–997. doi:10.1002/aic.13734.
- Ramkrishna, D. y Song, H.-S. (2016). Analysis of bioprocesses. dynamic modeling is a must. *Materials Today Proceedings*, 3:3587–3599.
- Rawat, G., Tripathi, P., y Saxena, R. (2013). Expanding horizons of shikimic acid. *Applied Microbiology and Biotechnology*, 97(10):4277–4287.
- Rizk, M. L. y Liao, J. C. (2009). Ensemble modeling for aromatic production in *Escherichia coli*. *PLOS One*, 4:e6903. doi:10.1371/journal.pone.0006903.
- Roche, A. R. (2009). Excellence in science. *Annual Report*, (800):12–25.
- Rodriguez, A., Martínez, J. A., Báez-Viveros, J. L., Flores, N., Hernández-Chávez, G., Ramírez, O. T., Gosset, G., y Bolivar, F. (2013). 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*. *Microbial Cell Factories*, 12:1–16. doi:0.1186/1475-2859-12-86.
- Rodriguez, A., Martínez, J. A., Flores, N., Escalante, A., Gosset, G., y Bolivar, F. (2014). Engineering *Escherichia coli* to overproduce aromatic amino acids and derived compounds. *Microbial Cell Factories*, 13:1–15. doi:10.1186/s12934-014-0126-z.
- Rodriguez, A., Martínez, J. A., Millard, P., Gosset, G., Portais, J.-C., Létisse, F., y Bolívar, F. (2017). Plasmid-encoded biosynthetic genes alleviate metabolic disadvantages while increasing glucose conversion to shikimate in an engineered *Escherichia coli* strain. *Biotechnology and Bioengineering*, 114:1319–1330. doi:10.1002/bit.26264.
- Saha, R., Chowdhury, A., y Maranas, C. D. (2014). Recent advances in the reconstruction of metabolic models and integration of omics data. *Current Opinion in Biotechnology*, 29:39–45. doi:10.1016/j.copbio.2014.02.011.

- Schuetz, R., Kuepfer, L., y Sauer, U. (2007). Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Molecular Systems Biology*, 3:1–15. doi:10.1038/msb4100162.
- Shimizu, K. (2014). Regulation systems of bacteria such as *Escherichia coli* in response to nutrient limitation and environmental stresses. *Metabolites*, (4):1–35.
- Siddiquee, K. A. Z., Arauzo-bravo, M. J., y Shimizu, K. (2004). Effect of a pyruvate kinase (*pykF*-gene) knockout mutation on the control of gene expression and metabolic fluxes in *Escherichia coli*. *FEMS Microbiology Letters*, 235:25–33. doi:10.1016/j.femsle.2004.04.004.
- Sigala, J. C., Flores, S., Flores, N., Aguilar, C., de Anda, R., Gosset, G., y Bolívar, F. (2008). Acetate metabolism in *Escherichia coli* strains lacking phosphoenlpyruvate: carbohydrate phosphotransferase system; evidence of carbon recycling strategies and futile cycles. *Journal of Molecular Microbiology and Biotechnology*, 16:224–235. doi:10.1159/000151219.
- Song, H.-S. y Ramkrishna, D. (2009). Reduction of a set of elementary modes using yield analysis. *Biotechnology and Bioengineering*, 102:554–568. doi:10.1002/bit.22062.
- Song, H.-S. y Ramkrishna, D. (2012). Prediction of dynamic behavior of mutant strains from limited wild-type data. *Metabolic Engineering*, 14:69–80. doi:10.1016/j.ymben.2012.02.003.
- Stelling, J. (2004). Mathematical models in microbial systems biology. *Current Opinion in Microbiology*, 7:513–518. doi:10.1016/j.mib.2004.08.004.
- Stephanopoulos, G. y Sinskey, A. J. (1993). Metabolic engineering - methodologies and future prospects. *Trends in Biotechnology*, 11(9):392–396.
- Tang, L., Xiang, H., y Sun, Y. (2009). Monopalmitoyloxy shikimic acid: Enzymatic synthesis and anticoagulation activity evaluation. *Applied Biochemistry and Biotechnology*, 158(2):408–415.

- Terzer, M. y Stelling, J. (2008). Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics*, 24:2229–2235. doi:10.1093/bioinformatics/btn401.
- Varner, J. y Ramkrishna, D. (1999). The non-linear analysis of cybernetic models. guidelines for model formulation. *Journal of Biotechnology*, 71:67–104. doi:10.1016/S0168-1656(99)00016-4.
- Wang, G.-W., Hu, W.-T., Huang, B.-K., y Qin, L.-P. (2011). *Illicium verum*: A review on its botany, traditional use, chemistry and pharmacology. *Journal of Ethnopharmacology*, 136:10–20. doi:10.1016/j.jep.2011.04.051.
- WHO (2011). A proposal has been made for the use of the term “Guidelines” in place of “Framework” throughout this text. Proposals have also been made for the use of the terms “Multilateral Framework” or “International Framework” and/or “Global sharing”. pages 1–45.
- Yi, J., Draths, K. M., Li, K., y Frost, J. W. (2003). Altered Glucose Transport and Shikimate Pathway Product Yields in *E. coli*. *Biotechnology Progress*, 19(5):1450–1459.
- Yu, C., Cao, Y., Zou, H., y Xian, M. (2011). Metabolic engineering of *Escherichia coli* for biotechnological production of high-value organic acids and alcohols. *Applied Microbiology and Biotechnology*, 89(3):573–583.



---


# ANEXOS

RESEARCH ARTICLE

Open Access



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

Juan A. Martínez<sup>1</sup> , Alberto Rodríguez<sup>1</sup>, Fabian Moreno<sup>1</sup>, Noemí Flores<sup>1</sup>, Alvaro R. Lara<sup>2</sup>, Octavio T. Ramírez<sup>3</sup>, Guillermo Gosset<sup>1</sup> and Francisco Bolívar<sup>1\*</sup>

## 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](mailto:bolivar@ibt.unam.mx)

Francisco Bolívar is a member from El Colegio Nacional, México  
<sup>1</sup>Departamento de Ingeniería Celular y Biotecnología, 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



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

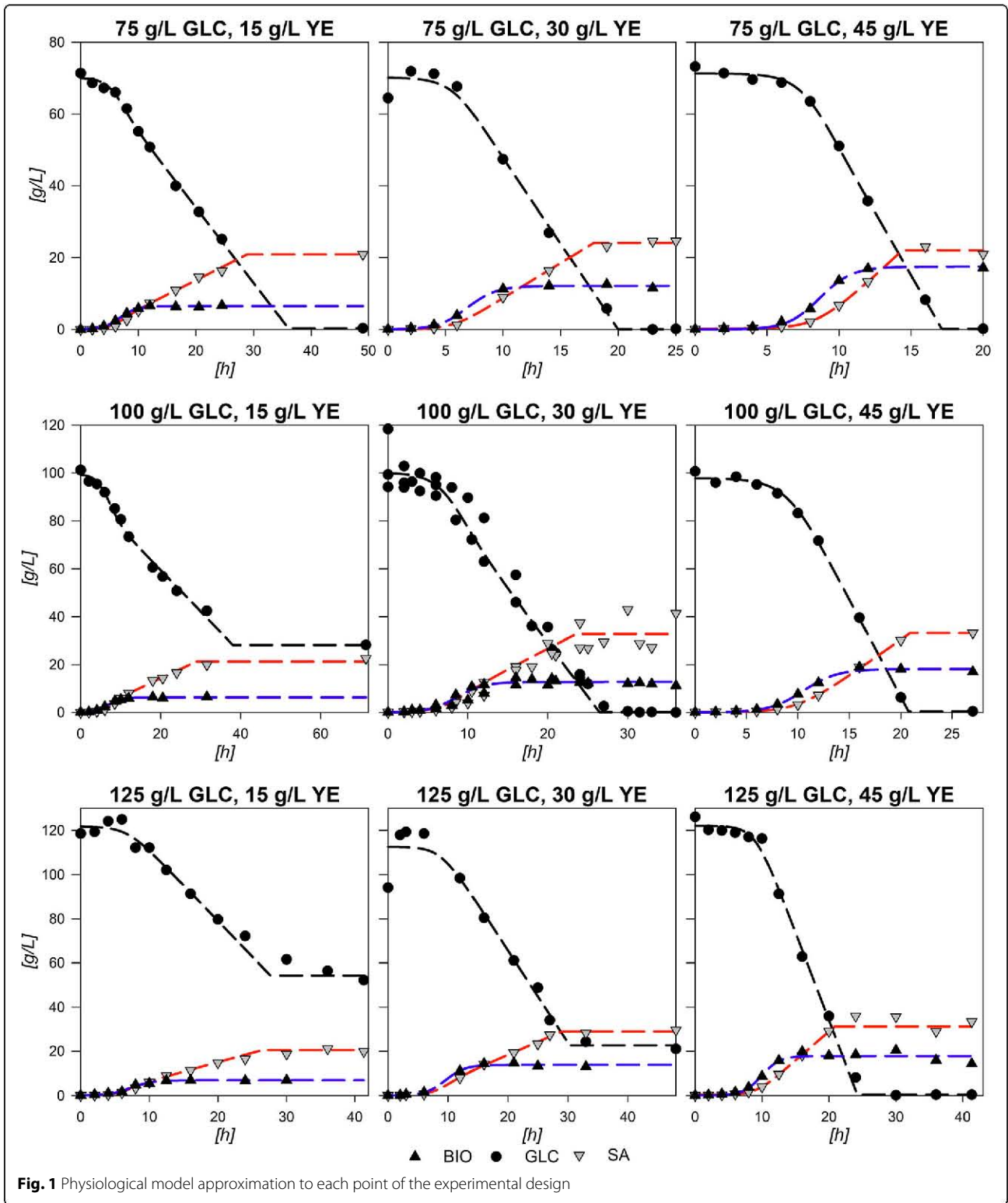
Figure 1 depicts the results from the 9 experimental design fermentations with the central point done by triplicate along with constructed physiological models (see “Methods” section). Central experimental condition (100:30 GLC:YE g/L) average parameters and deviations are summarized in Table 1. The 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  $\approx 40$  g/L of YE are required for complete exhaustion of more than  $\approx 110$  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  $\approx 75$  g/L (increasing consumption) or higher than  $\approx 100$  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 ( $\mu_{max}$ ) 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  $\approx 40$  g/L [YE] and  $\approx 110$  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:  $X_{max}$ ,  $\mu$ ,  $q_{glc}^{exp}$ ,  $q_{sa}^{exp}$ ,  $q_{glc}^{sta}$ ,  $q_{sa}^{sta}$  and  $SA_{final}$  parameters. As

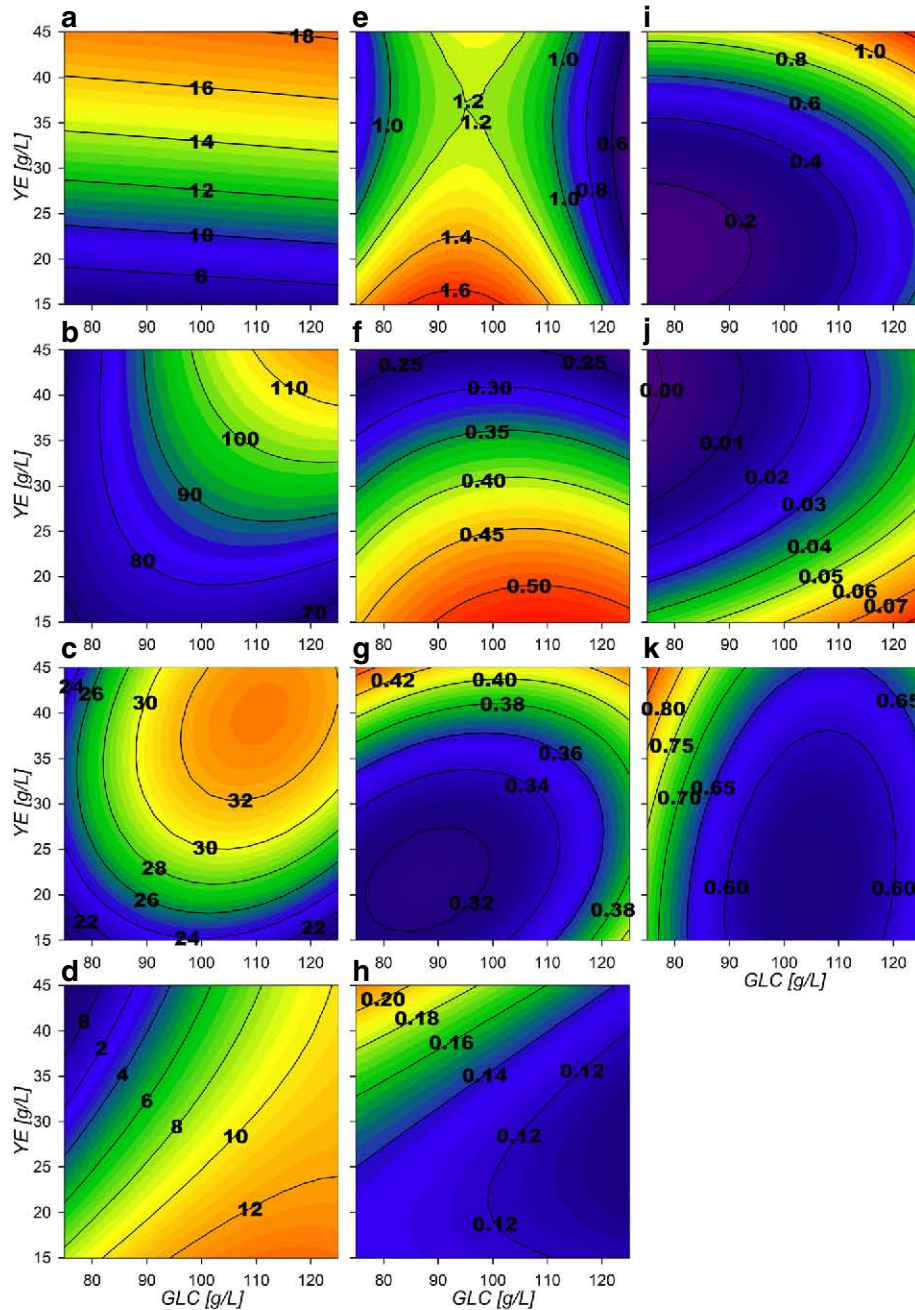


it can be seen, all models follow the experimental data with good agreement. The largest observable deviation is on the maximum SA achievable on the 80:40 GLC:YE

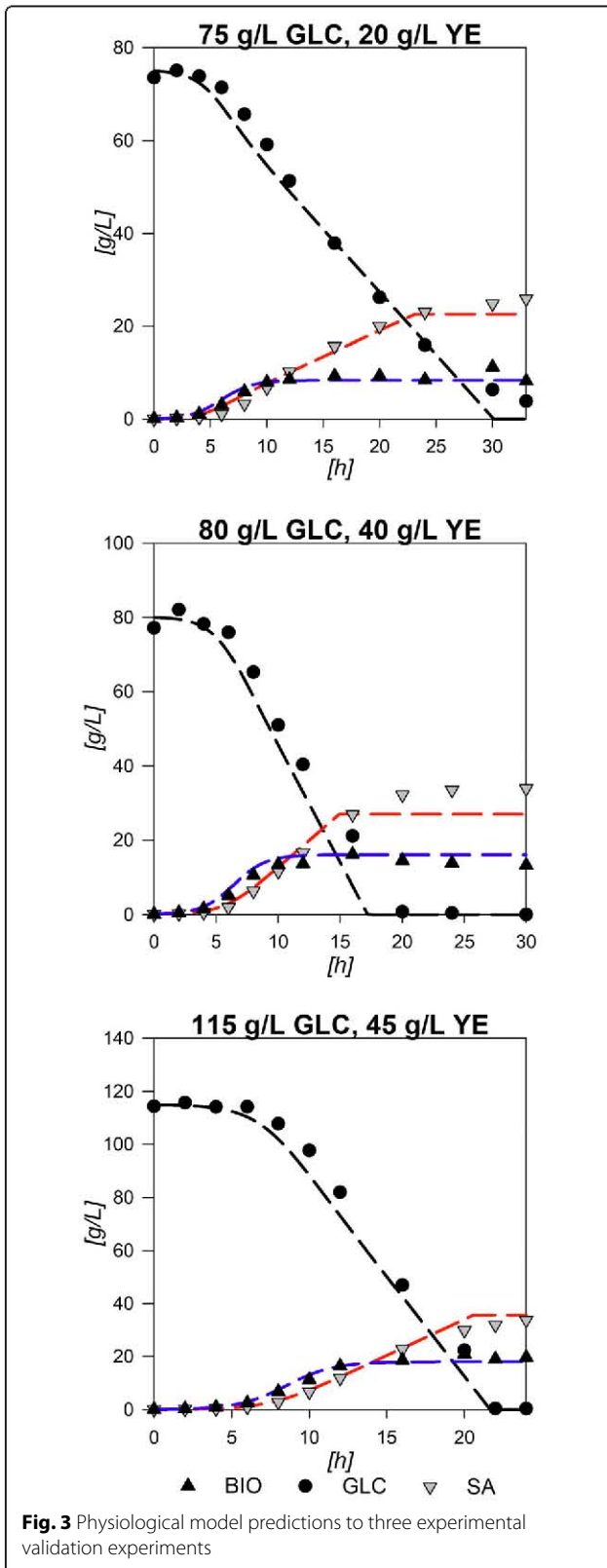
experiment, probably due to the contribution from YE to SA production. In addition, mathematical assessment of the validation was performed by a set of descriptive

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

	$X_{max}$ [g/L]	$\Delta$ [GLC] [g/L]	$[SA]_f$ [g/L]	$[AC]_f$ [g/L]	$\mu$ [h <sup>-1</sup> ]	$q_{glc}^{exp}$ [g/Lh]	$q_{sa}^{exp}$ [g/Lh]	$q_{glc}^{sta}$ [g/Lh]	$q_{sa}^{sta}$ [g/Lh]	$Y_{ps}$ [g/g]	$Y_{px}$ [g/g]	$q_{ac}^{exp}$ [g/Lh]	$q_{ac}^{sta}$ [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
$\sigma^2$	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**  $\mu_{max}$  [h<sup>-1</sup>]



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 Average
		Exp.	Model	Exp.	Model	Exp.	Model	
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
$\mu_{max}$	[h <sup>-1</sup> ]	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

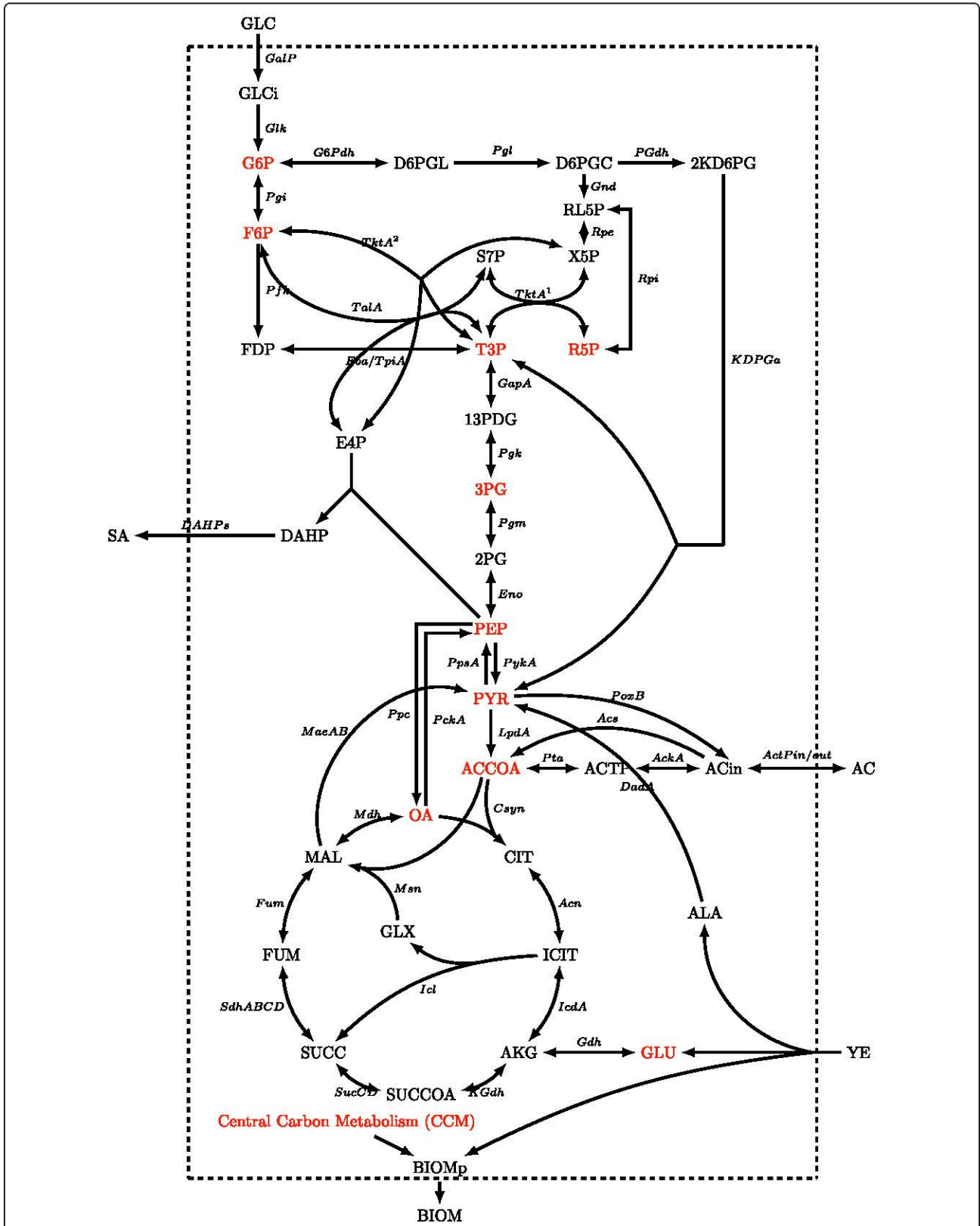
on a time prior to the model estimations. Regarding AC, SPD deviations ranged from 1.31 to 9% in all cases, except for 100:30 and 100:45 where values were 21 and 42% respectively. For the 100:45 condition, this overestimation is due to the error in the AC peak found on mid exponential phase and to the 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.

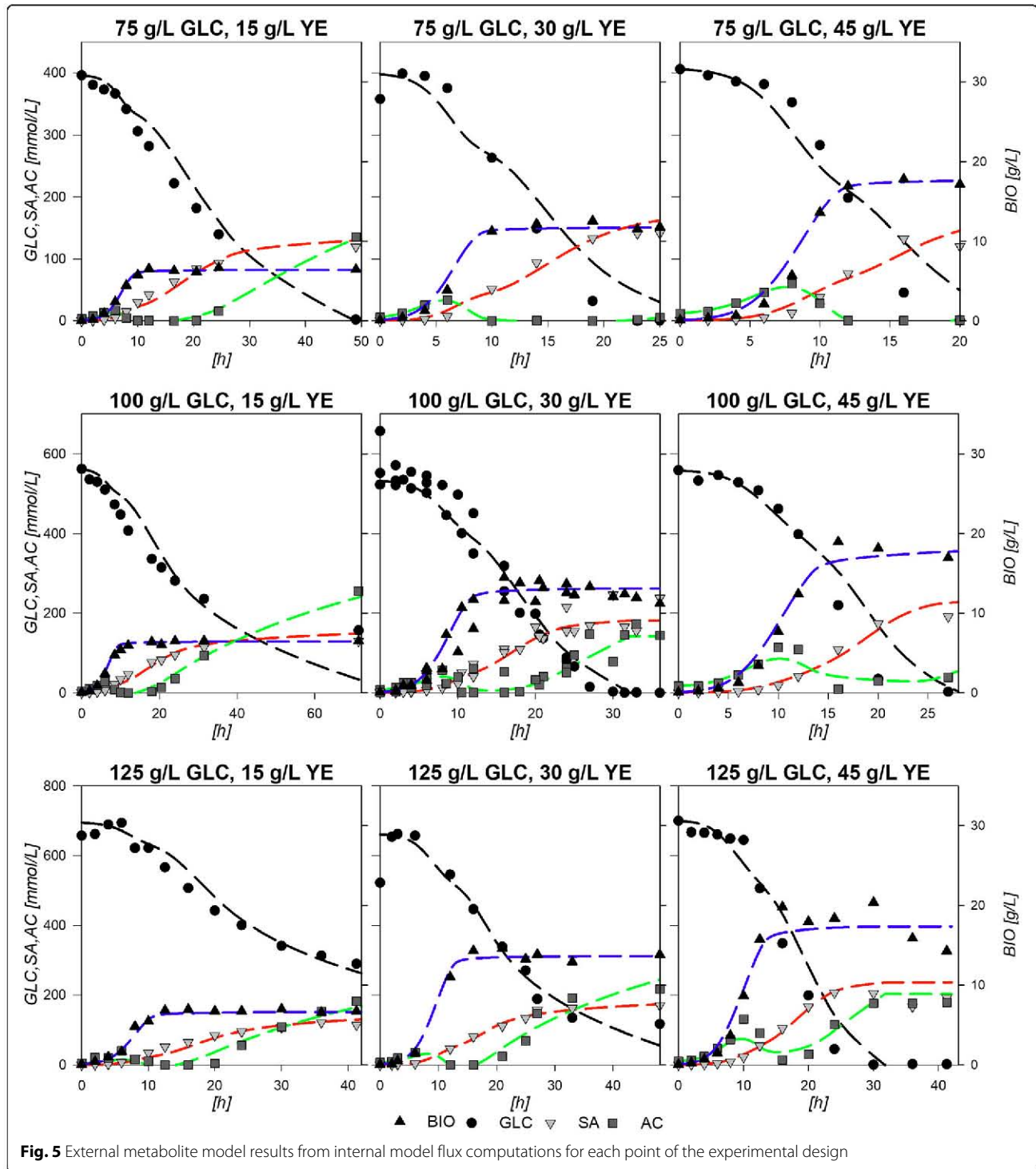
#### Central carbon metabolism flux distribution behavior during growth

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



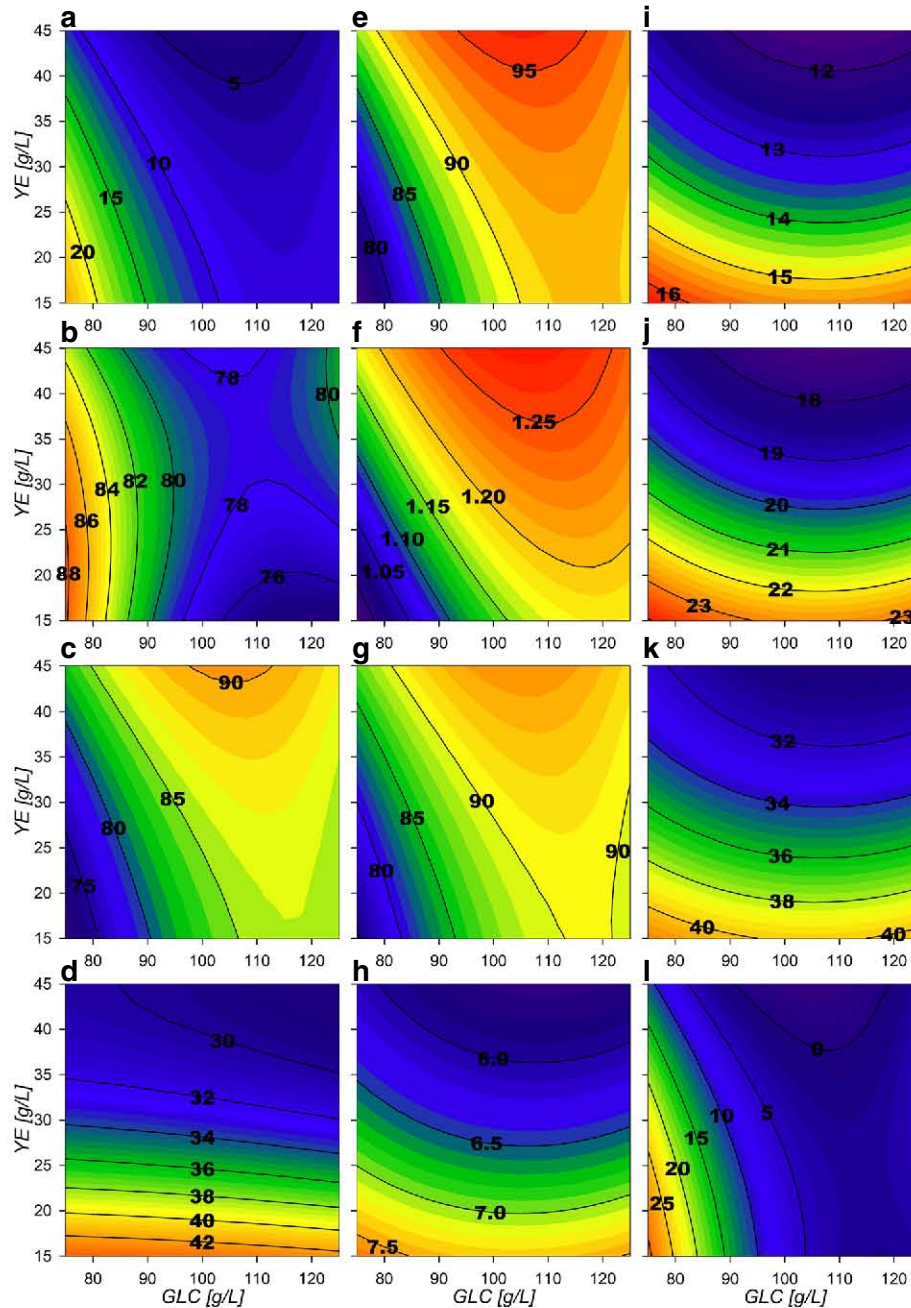


**Fig. 4** Central carbon metabolism constructed metabolic network for dynamic metabolic flux models. Metabolites on red refer to the CCM intermediaries used to produce biomass precursor (BIOMp)



Glk flux surfaces, but have smaller values than expected, accounting for only  $\approx 6$  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

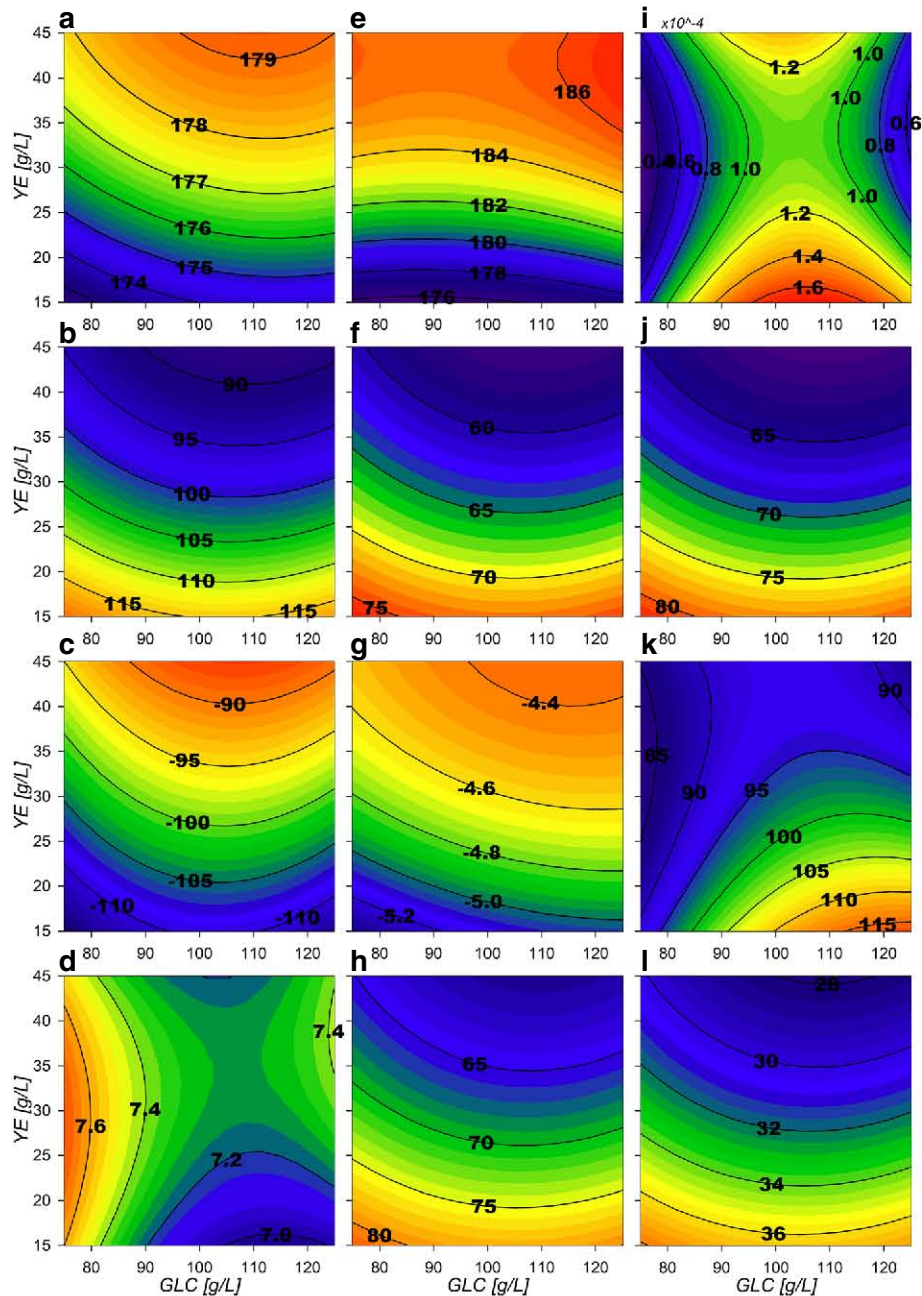
by *zwf* and the 6-phosphogluconolactonase (Pgl). G6Pdh presents relative flux values from  $\approx 75$  to 95% (Fig. 6e) and its morphology presents the inverse features than Pgi, which presents greater relative flux values at higher [GLC] and [YE] initial conditions. High relative flux values of



**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  $\approx 75$  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 ( $\approx 6$  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  $\approx 12$  to 16% (Fig. 6i). These non-oxidative branch reactions are responsible for the E4P formation and present combined flux values from  $\approx 18$  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  $\approx 72$  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  $\approx 31$  to 41% and  $\approx 0$  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  $\approx 28$ –38% from the NADPH dependent enzyme (MaeB) (Fig. 7l) and  $\approx 0$ –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  $\approx 28$ –42% of relative flux (Fig. 6d) compared to pyruvate oxidase (PoxB), which presented fluxes towards AC calculated to be between  $\approx 176$ –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  $\approx 90$  to 115%. To clarify the node distribution around PoxB, an AR36 $\Delta$ 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 suggests that PoxB could be indeed the main contributor to AC production in the AR36 PTS<sup>-</sup> strain [2, 33, 36–38]. The AR36 $\Delta$ poxB cultures also showed lower growth rates (0.21 h<sup>-1</sup>) 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  $\approx 56$  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  $\approx 60$  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  $\approx 3.1$  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  $\approx 85$  to 1125% (Fig. 7k) as it also assimilates ACCOA carbon derived from the Msn reaction on GSP.

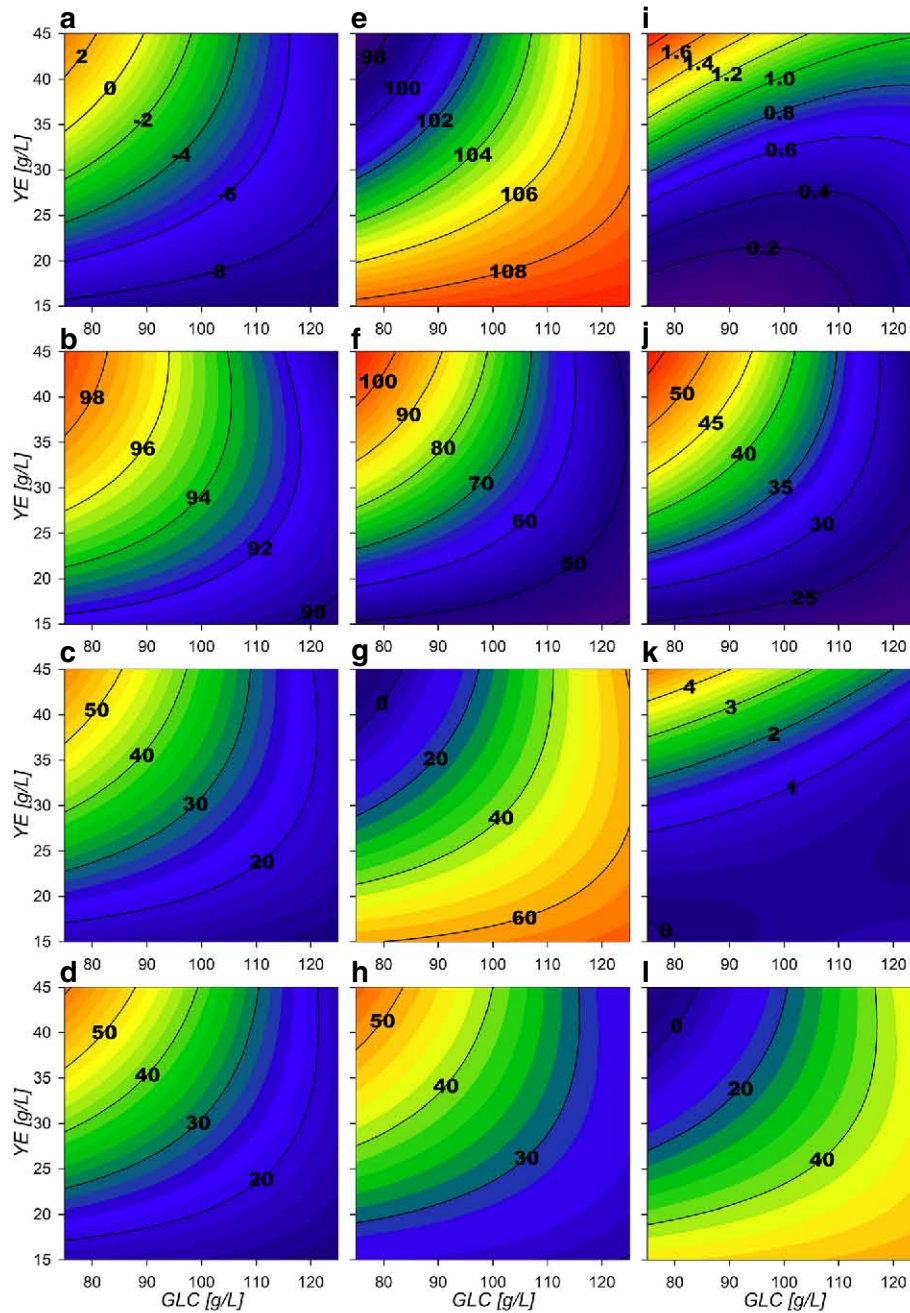
Surface morphologies suggest the allocation of different predominant extracellular production zones at the exponential growth phase. A SA production predominance flux zone is found at low initial [YE] and as the initial [GLC] diminishes a more balanced production towards biomass and SA is found. This 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  $\approx 40$  to 100% (Fig. 8f), whereas its competing PGdh accounts for  $\approx 0$ –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

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  $\approx 10$ –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  $\approx 90$  to 99% (Fig. 8b). Interestingly, the model renders a small ( $\approx 0$ –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  $\approx 0$ –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  $\approx 0$ –5% and only being present under high [YE] initial conditions (Fig. 8k). On the other hand, its counterpart reaction Ppc shows  $\approx 0$ –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  $\approx 15$ –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  $\approx 0$  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  $\approx 54$  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  $\approx 10$  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.

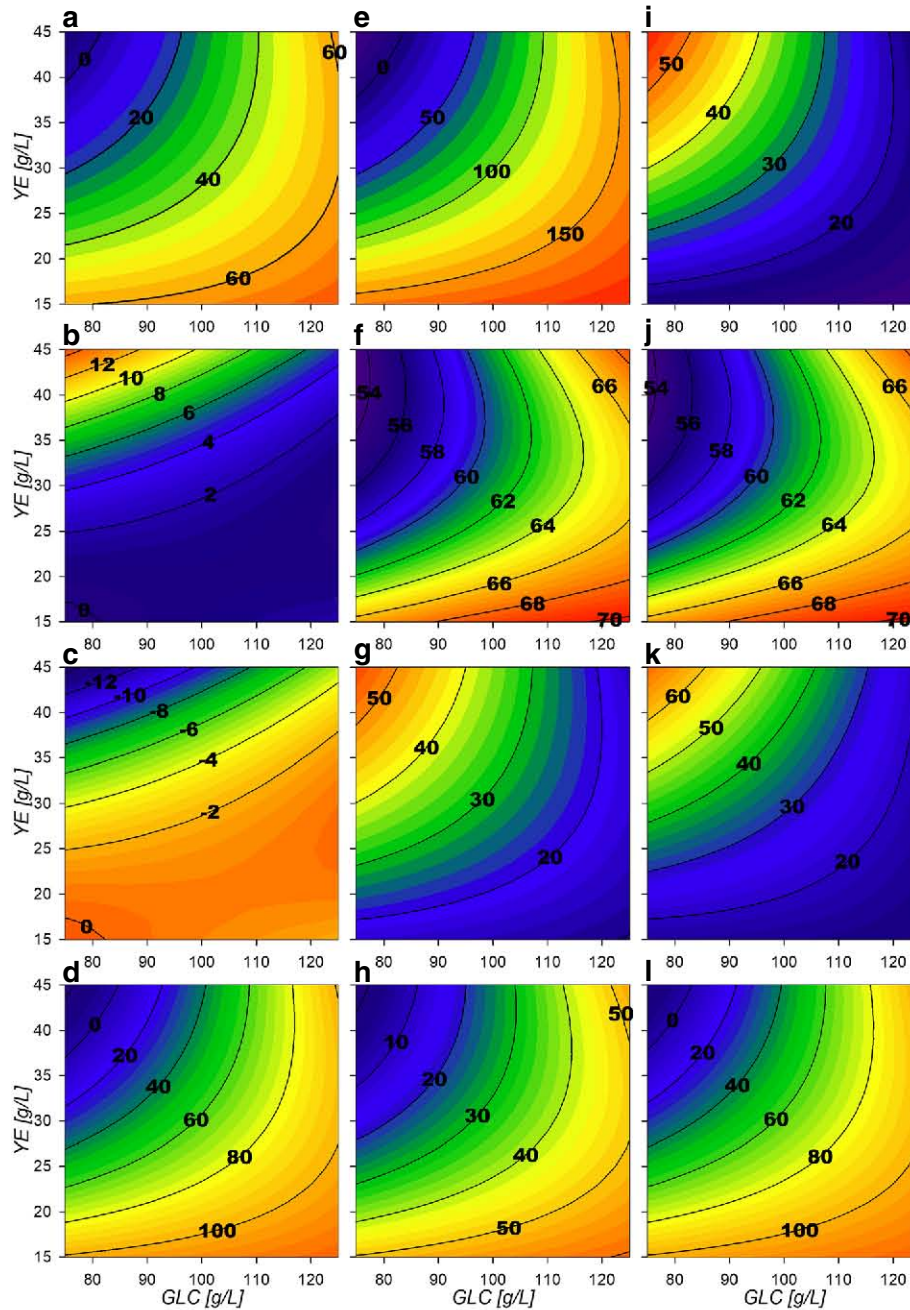
The dynamic cybernetic model on this phase showed two predominant production zones. The observations made by the dynamic flux models allocate a SA production zone above an imaginary diagonal line, cutting the experimental design area from low to high initial substrate concentration, and a predominantly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



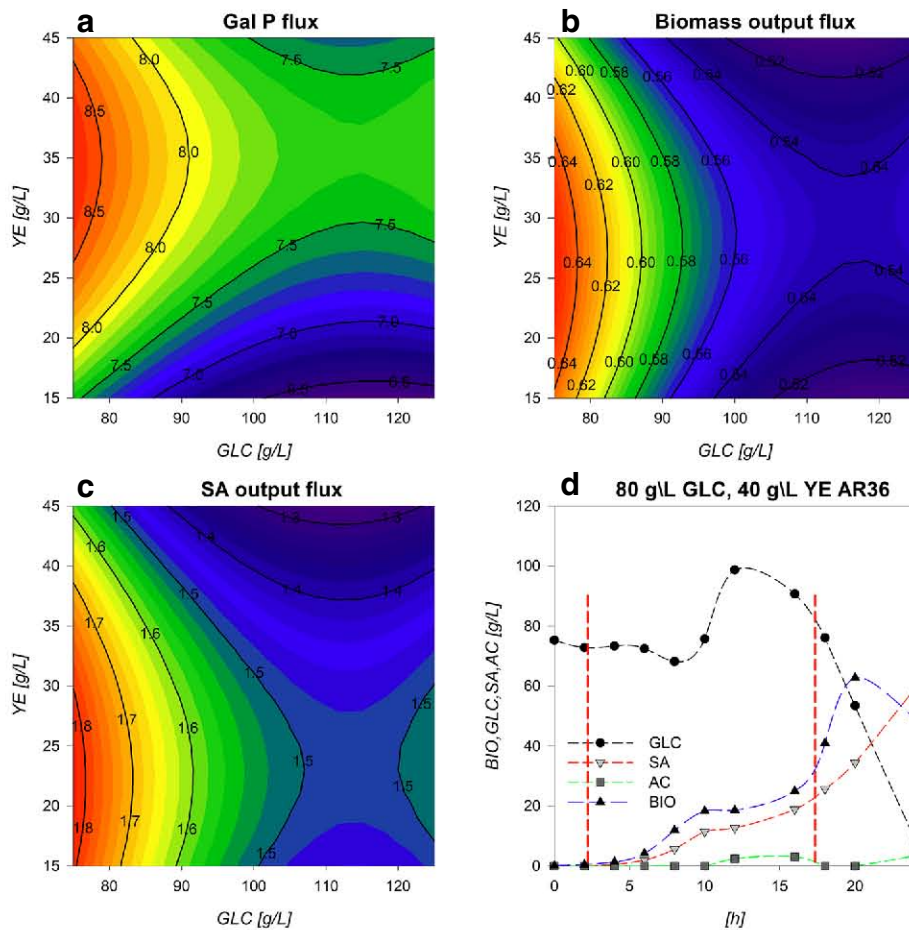
**Fig. 9** Response surface contour plots for the estimated internal fluxes at MSt pt2 (% Flux relative to GLC consumption). **a** ActPout, **b** ActPin, **c** AckA, **d** Acs, **e** PoxB, **f** Csyn, **g** IcdA, **h** Icl, **i** KGdh, **j** SdhABCD, **k** Mdh, **l** MaeB



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 Act-Pin 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 were 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  $\approx 75\text{--}80$  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  $0.90\text{ h}^{-1}$ , which is in range of the ones predicted by models ( $0.8\text{--}0.85\text{ h}^{-1}$ ), while the second growth rate is only about  $0.18\text{ h}^{-1}$ , 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  $2.45\text{ g SA/L}\cdot\text{h}$  representing a 70% increase from the  $1.43\text{ gSA/L}\cdot\text{h}$  reported by Rodríguez et al. (2013)[2] and which is 20% higher than the previously reported industrial *E. coli* strains ( $2.04\text{ gSA/L}\cdot\text{h}$  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  $\Delta[GLC]$  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 through 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 adenylylase, which in turn has been also found to be up-regulated on these PTS<sup>-</sup> 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  $\alpha$ -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<sup>-</sup> 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 AR36 $\Delta$ *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<sup>-</sup> 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<sup>-</sup> 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 phosphoenolpyruvate: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  $\mu_{max}$  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:

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

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

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

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 ( $q_b^{exp}$ ) and initial stationary ( $q_b^{sta}$ ) approximated production rates were calculated with the following equations:

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

$$q_b^{sta} = Q_b^{sta} / X_{max} \quad (5)$$

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  $u \equiv [u_1, u_2, u_3, \dots, u_m]$  and  $v \equiv [v_1, v_2, v_3, \dots, v_m]$  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 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:

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

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

where indexes  $M$  and  $G$  refer to the exponential (mixed consumption) family and stationary (glucose only consumption) family, respectively. Index  $i$  refers to each  $i^{th}$  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  $K_m$  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

**Additional file 1:** Model and validation data, parameters and statistical values. Description: Parameters and statistical value tables for physiological models. Response surface parameters and statistical values for the polynomial approximation. Response surface prediction validation data. Response surface critical points calculations. Dynamic Flux models parameters and statistics. (PDF 186 kb)



**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*<sup>-</sup> figures. Description: Table for all plasmids and oligonucleotides used. AR36 $\Delta$ *poxB* strain construction. Fermentation profiles for AR36 $\Delta$ *poxB* and AR36 on highly concentrated substrate media. (PDF 215 kb)

### Acknowledgments

We would like to thank Professor Doriswami Ramkrishna and Dr. Huyn-Seob Song for kindly providing the software and knowledge used in this work and for their support and fruitful discussions. Thanks to Dr. Adelfo Escalante for the help on the construction of this manuscript. Also, thanks to Georgina Hernández and Ramón de Anda for their technical support.

### Funding

This work was supported by Consejo Nacional de Ciencia y Tecnología (CONACYT, México) Ciencia Básica grant 240519 and DGAPA-PAPIIT UNAM, México grant IN209618.

### 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  $\Delta$ *poxB* 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

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

Received: 25 April 2018 Accepted: 12 October 2018

Published online: 12 November 2018

### References

- Patnaik R, Spitzer RG, Liao JC. Pathway engineering for production of aromatic in *Escherichia coli*: Confirmation of stoichiometric analysis by independent modulation of *arg*, *tka*, and *pps* activities. *Biotech Bioeng.* 1995;46:361–70. <https://doi.org/doi:10.1002/bit.260460409>.
- Rodríguez A, Martínez JA, Báez-Viveros JL, Flores N, Hernández-Chávez G, Ramírez OT, Gosset G, Bolívar 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:10.1186/1475-2859-12-86>.
- 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>.
- 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>.
- 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>.
- Báez-Viveros JL, Osuna J, Hernández-Chávez G, Soberon X, Bolívar 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>.
- Gosset G. Production of aromatic compounds in bacteria. *Curr Opin Biotechnol.* 2009;20:651–8. <https://doi.org/doi:10.1016/j.copbio.2009.09.012>.
- 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>.
- 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>.
- Li S, Yuan W, Wang P, Zhang Z, Zhang W, Ownby S. Method for Rhe Extraction and Purification of Shikimic Acid; 2007, p. 20070149805A1.
- 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.*
- 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>.
- 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>.
- 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>.
- Krämer M, Bongaerts J, Bovenberg R, Kremer S, Müller 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>.
- 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>.
- Cloots L, Marchal K. Network-based functional 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>.
- 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>.
- 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>.

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](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](https://doi.org/doi:10.1016/S0167-7799(03)00030-1).
24. Joutten 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:10.1016/j.copbio.2014.12.019>.
30. O'Brien 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 pathway. *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 <sup>13</sup>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 phosphoenolpyruvate: 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. Rodríguez A, Martínez JA, Flores N, Escalante A, Gosset G, Bolívar 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. Couitilization 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. Rodríguez A, Martínez JA, Millard P, 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, Bolívar 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:10.009-2240/90/041004-08/protectT1.txt\\$02.00/0](https://doi.org/doi:10.009-2240/90/041004-08/protectT1.txt$02.00/0).
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](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](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](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.jymben.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.jymben.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>.

**Ready to submit your research? Choose BMC and benefit from:**


- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)



# Plasmid-Encoded Biosynthetic Genes Alleviate Metabolic Disadvantages While Increasing Glucose Conversion to Shikimate in an Engineered *Escherichia coli* Strain

Alberto Rodriguez <sup>1</sup>, Juan A. Martínez,<sup>1</sup> Pierre Millard,<sup>2</sup> Guillermo Gosset,<sup>1</sup> Jean-Charles Portais,<sup>2</sup> Fabien Létisse,<sup>2</sup> Francisco Bolívar<sup>1</sup>

<sup>1</sup>Instituto de Biotecnología, Universidad Nacional Autónoma de México (UNAM), Cuernavaca, Morelos, Mexico

<sup>2</sup>LISBP, Université de Toulouse, CNRS, INRA, INSA, Toulouse, France; telephone: +33 (0) 561559407; fax: +33 (0) 561559689; e-mail: letisse@insa-toulouse.fr

**ABSTRACT:** Metabolic engineering strategies applied over the last two decades to produce shikimate (SA) in *Escherichia coli* have resulted in a battery of strains bearing many expression systems. However, the effects that these systems have on the host physiology and how they impact the production of SA are still not well understood. In this work we utilized an engineered *E. coli* strain to determine the consequences of carrying a vector that promotes SA production from glucose with a high-yield but that is also expected to impose a significant cellular burden. Kinetic comparisons in fermentors showed that instead of exerting a negative effect, the sole presence of the plasmid increased glucose consumption without diminishing the growth rate. By constitutively expressing a biosynthetic operon from this vector, the more active glycolytic metabolism was exploited to redirect intermediates toward the production of SA, which further increased the glucose consumption rate and avoided excess acetate production. Fluxomics and metabolomics experiments revealed a global remodeling of the carbon and energy metabolism in the production strain, where the increased SA production reduced the carbon available for oxidative and fermentative pathways. Moreover, the results showed that the production of SA relies on a specific setup of the pentose phosphate pathway, where both its oxidative and non-oxidative branches are strongly activated to supply erythrose-4-phosphate and balance the NADPH requirements. This work improves our understanding of the metabolic reorganization observed in *E. coli* in response to the plasmid-based expression of the SA biosynthetic pathway.

Biotechnol. Bioeng. 2017;114: 1319–1330.

© 2017 Wiley Periodicals, Inc.

**KEYWORDS:** shikimate; plasmid burden; *Escherichia coli*; metabolism

## Introduction

Shikimate (SA) is an intermediate metabolite of the aromatic amino acids biosynthetic pathway, which has gained attention for its role as starting material in the chemical synthesis of the anti-influenza drug oseltamivir phosphate (Kim and Park, 2012; Nie et al., 2012; Ward et al., 2005). SA also displays antioxidant, anticoagulant, and anti-inflammatory activities and can be used to synthesize derivatives with interesting pharmaceutical properties like anticancer drugs, hormones, and pigments (Díaz Quiroz et al., 2014; Estévez and Estévez, 2012). Several species of microorganisms have been modified to overproduce SA from different carbon sources, although most works have focused on engineered *Escherichia coli* strains fed with glucose (Glc) (Ghosh et al., 2012; Martínez et al., 2015; Rawat et al., 2013). Such interest has fueled the comprehensive testing and improvement of expression platforms, genetic backgrounds, and culture strategies to increase the production of this compound (Ahn et al., 2011; Bogosian and Frantz, 2013; Chandran et al., 2003; Chen et al., 2012, 2014; Cui et al., 2014; Escalante et al., 2010; Frost and Knop, 2002; Johansson et al., 2005; Rodriguez et al., 2014).

In particular, Rodriguez et al. reported the construction and initial characterization of an *E. coli* strain capable of producing SA in batch cultures with high concentrations of Glc and yeast extract (YE). This strain, named AR36, strongly expressed a six-gene synthetic operon in a constitutive manner from a plasmid present at about 50 copies per cell, allowing it to achieve high SA titers and yields in complex media (Rodriguez et al., 2013). One noteworthy characteristic of strain AR36 is its ability to withstand and consume more than 100 g/L of Glc in batch-mode cultivations without any evident affectations to its growth and production capabilities. In fact, this strain displayed similar specific growth, consumption and production rates, and constant SA yields on Glc, regardless of the fermentation volume or substrate concentrations tested. However, the use of a high-copy plasmid with a strong promoter to constitutively produce proteins of a biosynthetic pathway is considered to be very resource demanding, causing low biomass

Alberto Rodriguez present address is Joint BioEnergy Institute, Emeryville, CA 94608.

Conflicts of interest: The authors have no conflict of interest to declare.

Correspondence to: F. Létisse

Received 18 October 2016; Revision received 17 January 2017; Accepted 8 February 2017

Accepted manuscript online 10 February 2017;

Article first published online 9 March 2017 in Wiley Online Library (<http://onlinelibrary.wiley.com/doi/10.1002/bit.26264/abstract>).

DOI 10.1002/bit.26264

and product yields, and thus it is generally avoided as part of a metabolic engineering strategy (Jones et al., 2000). Recombinant protein production is considered to be the main contributor to such metabolic burden, although other factors like DNA replication and activation of stress responses also seem to play an important role (Glick, 1995; Hoffmann and Rinas, 2004; Silva et al., 2012). It is still difficult to predict the effects that plasmids can impose to the bacterial physiology and metabolism since this imbalance of energy and precursors can make the cells respond to the environment in a variety of ways (Carneiro et al., 2013; Ow et al., 2009). Interestingly, in some cases the presence of one or more plasmids, or the heterologous expression of genes, can increase the Glc consumption rate (Diaz-Ricci et al., 1995; Wang et al., 2006) or biomass production (Diaz-Ricci et al., 1992; Ingram and Conway, 1988). Aside from the properties of the expression system itself, the genetic background and cultivation conditions can also impact strain productivity by altering the availability of specific intracellular intermediates required for growth and plasmid maintenance (Carneiro et al., 2013; Diaz Ricci and Hernández, 2000; Donovan et al., 1996; Wu et al., 2016).

In this work, we were interested in getting an insight of the physiological state of the engineered strain AR36, since this information may allow for further optimization of the SA production process. Therefore, we investigated some of the changes imposed to its central metabolism by the presence of a high-copy plasmid and by the constitutive expression of six proteins encoded in a synthetic operon that promote SA production from the start of the cultivations. After comparing the kinetic, metabolomic, and fluxomic profiles of the production strain against the parental strain carrying no plasmid or an “empty” plasmid, key metabolic responses were identified and discussed.

## Materials and Methods

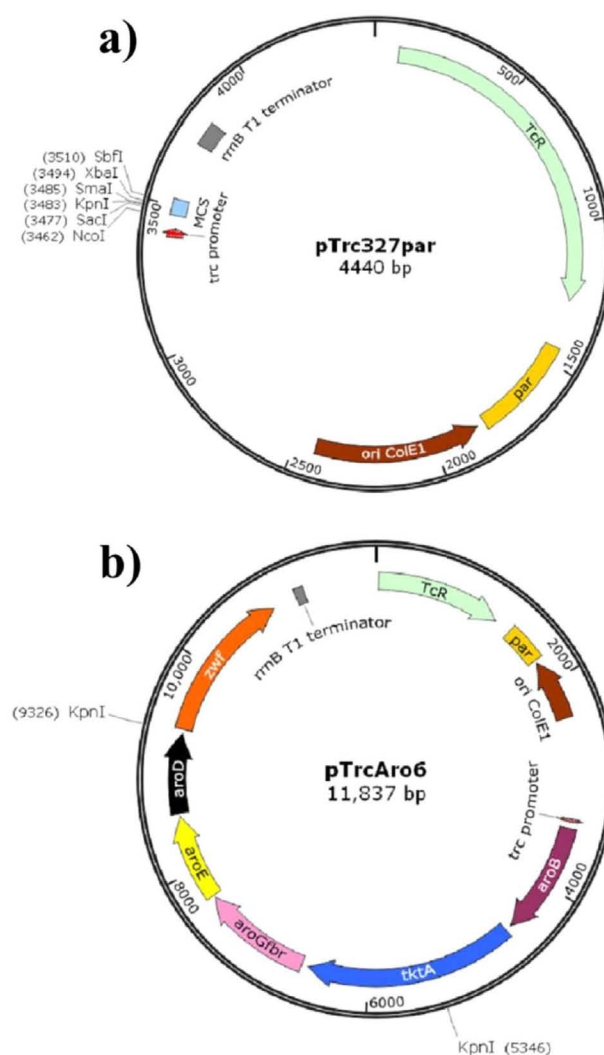
### Strains and Plasmids

All *E. coli* strains and plasmids used in this work have been reported before and their construction described in further detail (Rodríguez et al., 2013). Briefly, SA production strain AR36 was generated after introducing several genetic modifications to strain PB12, a JM101 derivative isolated from an adaptive evolution experiment to select mutants in the phosphotransferase system (PTS) with increased Glc consumption rates (Aguilar et al., 2012). The genetic engineering approaches that generated strain AR36 (besides inactivation of genes coding for PTS components in PB12) include the inactivation of pyruvate kinase I (*pykF*), shikimate kinases I and II (*aroK* and *aroL*) and the lactose operon repressor (*lacI*). AR36 also carries a pBR322-derived plasmid with higher copy number and segregational stability that contains a 6-gene synthetic operon controlled by a Trc promoter to increase the production of SA (Rodríguez et al., 2013). Since there is no functional *lacI* gene in the plasmid or chromosome of strain AR36, genes controlled by the Trc promoter are constitutively expressed without the need to add IPTG. The production plasmid is called pTrcAro6 and was assembled by cloning a synthetic operon into the backbone plasmid pTrc327par (Fig. 1). To simplify the nomenclature, a strain having the same genetic background as AR36 but carrying instead the empty

(pTrc327par) plasmid will be referred as AR3e and the derivative without any plasmid will be called AR3. Molecular verification of strains and plasmids is shown in Supplementary material S1.

### Cultivation Media

The composition (per liter) of the SA production medium is: K<sub>2</sub>HPO<sub>4</sub> (7.5 g), KH<sub>2</sub>PO<sub>4</sub> (7.5 g), citric acid monohydrate (2.1 g), ammonium iron (III) citrate (0.3 g), concentrated H<sub>2</sub>SO<sub>4</sub> (1.2 mL), MgSO<sub>4</sub> (640.0 mg), CaCl<sub>2</sub> (60.0 mg), (NH<sub>4</sub>)<sub>6</sub>(Mo<sub>7</sub>O<sub>24</sub>) (3.7 mg), ZnSO<sub>4</sub> (2.9 mg), H<sub>3</sub>BO<sub>3</sub> (24.7 mg), CuSO<sub>4</sub> (2.5 mg), MnCl<sub>2</sub> (15.8 mg), CoCl<sub>2</sub> (1.3 mg), thiamine (1.0 mg), and betaine (234.0 mg). The medium was adjusted to pH 7 with 10 N NaOH before autoclaving it. Tetracycline (30 µg/mL) was added whenever needed for plasmid maintenance. YE (added before autoclaving) and Glc (filter-sterilized) were supplied at varying concentrations: 25 g/L of Glc and 15 g/L of YE for inocula in



**Figure 1.** Maps of the plasmids carried by the strains used in this study. (a) Plasmid pTrc327par is a pBR327-derivative carrying a Trc promoter and expressing a gene conferring resistance to tetracycline; (b) the cloning of a synthetic operon into pTrc327par resulted in plasmid pTrcAro6. Details on the construction of these plasmids have been previously reported (Rodríguez et al., 2013).

shake flasks and 100 g/L of Glc and 30 g/L of YE for 1 L fermentors. Cells were grown following the same inoculation train, equipment, controlled conditions, and sampling scheme as previously reported (Rodriguez et al., 2013). Cell growth was measured by monitoring optical density at 600 nm ( $OD_{600}$ ) with a spectrophotometer (DU700, Beckman, Indianapolis, IN). A previously established correlation factor between cell dry weights and optical density of 0.36 (gCDW/L)/ $OD_{600}$  (Rodriguez et al., 2013) was used to calculate the biomass concentrations and the physiological parameters (see below).

For metabolomics experiments, a minimal medium with low phosphate and sulfate salts was used. This medium contains (per liter):  $Na_2HPO_4 \cdot 12H_2O$  (3.48 g),  $KH_2PO_4$  (0.61 g), NaCl (0.51 g),  $NH_4Cl$  (2.04 g),  $MgSO_4$  (98.0 mg),  $CaCl_2$  (4.4 mg),  $Na_2EDTA \cdot 2H_2O$  (15.0 mg),  $ZnSO_4 \cdot 7H_2O$  (4.5 mg),  $CoCl_2 \cdot 6H_2O$  (0.3 mg),  $MnCl_2 \cdot 4H_2O$  (1.0 mg),  $H_3BO_3$  (1.0 mg),  $Na_2MoO_4 \cdot 2H_2O$  (0.4 mg),  $FeSO_4 \cdot 7H_2O$  (3.0 mg),  $CuSO_4 \cdot 5H_2O$  (0.3 mg), and thiamine (1.0 mg). Glc (3 g/L) was used as the only carbon source.

### Preparation of Resting Cells and Extraction of Metabolites

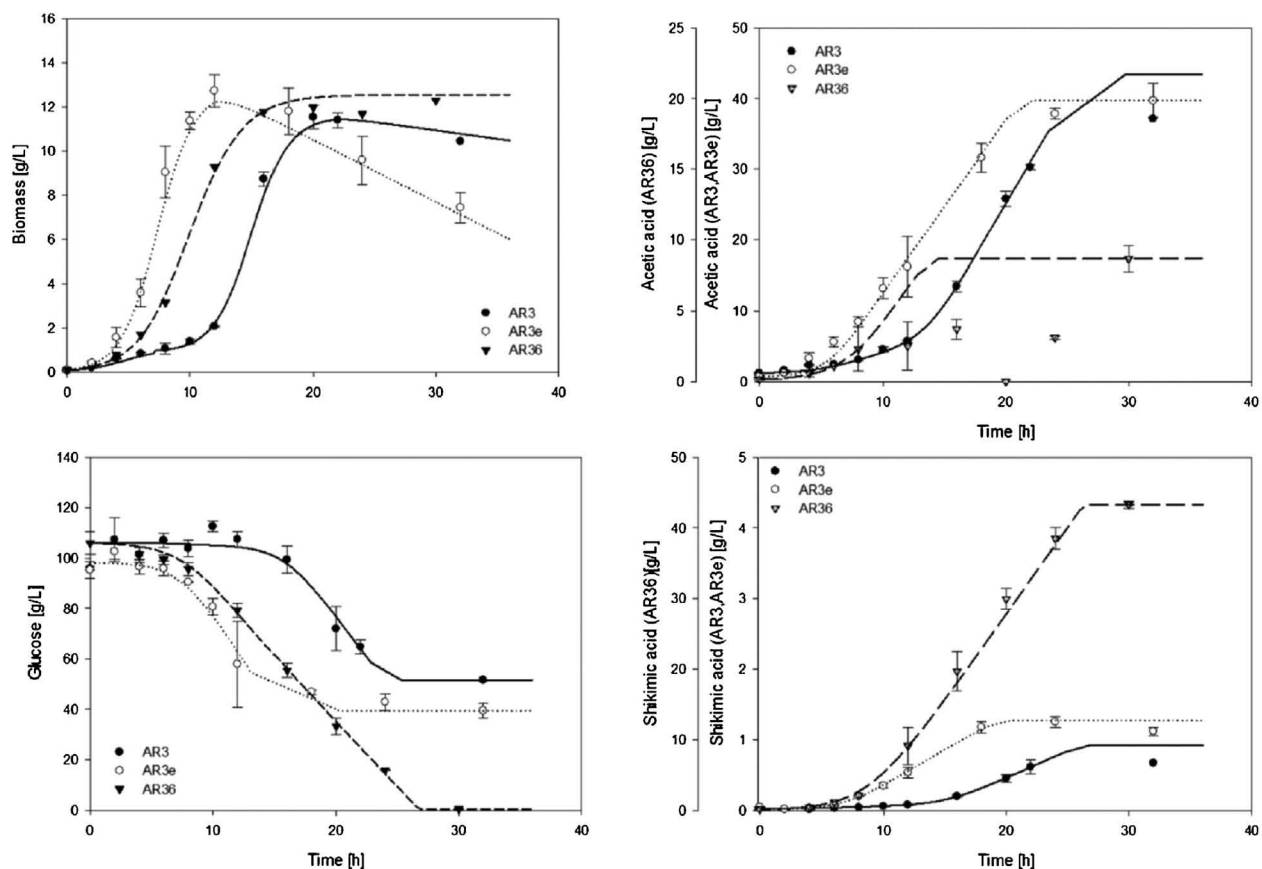
Seed cultures were started from frozen stocks in LB medium supplemented with 2 g/L of Glc and grown for 8 h at 37°C and 180 rpm. These cultures were used to inoculate 500 mL shake flasks with 50 mL of SA production medium, supplemented with 25 g/L of Glc and 15 g/L of YE and grown until mid-exponential phase. Cells from 50 mL of culture were then harvested by centrifuging (3000 rpm, 4 min, 22°C) and washed with 10 mL of low-salt minimal medium to remove the YE in the supernatant. To start the resting cells experiment, the biomass was centrifuged again and concentrated in a small volume, in order to reach approximately 5  $OD_{600}$  after transferring it to 250 mL shake flasks with 25 mL of low-salt medium containing 3 g/L of Glc. Samples were then collected at 1 h intervals to measure extracellular metabolites and to verify a constant Glc consumption and the absence of growth. To extract intracellular metabolites, whole-broth samples (120  $\mu$ L) were added to 1 mL of a cold solution of methanol:acetonitrile:water (40:40:20) and kept for 1 h at -20°C, to allow for complete cell lysis. Also, 120  $\mu$ L of a fully  $^{13}C$ -labeled metabolites cellular extract was added to serve as an internal standard for relative abundance comparisons. Metabolites in the quenching solution were separated from cell debris by brief centrifugation and the supernatant was then evaporated in a vacuum concentrator (SC110A SpeedVac Plus, ThermoSavant, Waltham, MA) for 12 h. The dried metabolites were stored at -80°C and resuspended in 120  $\mu$ L of water before analysis. Samples for ppGpp quantitation were collected from cells growing exponentially in the 500 mL shake flasks used as inocula for the resting cells (in SA production media). In this case, a 120  $\mu$ L aliquot of the whole broth was collected and the cells separated by fast-filtration using 0.22  $\mu$ m nylon filters. The cells were then quickly washed with a minimal medium salts solution, quenched by transferring the filter into 1 mL of a cold solution of methanol:acetonitrile:water (40:40:20) and processed as described above for the other metabolites. Glycogen was extracted from the resting cell cultures and measured using an adaptation of previously reported methods (Krisman, 1962; Shim et al., 2009). Cells from 1 mL of culture were resuspended in 250  $\mu$ L of 50 mM sodium acetate buffer,

sonicated, and centrifuged to separate cell debris. Glycogen was precipitated with two volumes of absolute ethanol, evaporated and quantified by measuring the absorbance at 460 nm after allowing it to react with an iodine reagent (0.15 M  $I_2$  + 1 M KI) in presence of high concentrations of  $CaCl_2$  (Shim et al., 2009). Samples for intracellular metabolites were taken in triplicate for each of three points in time from two biological replicates of each strain.

### Metabolite Quantitation, Data Analysis, and Calculations

Glc, SA, and acetate from the fermentor experiments were quantified with a Waters HPLC system (600E quaternary pump, 717 automatic injector, 2410 refraction index, and 996 photodiode array detectors) equipped with an Aminex HPX-87H column (300  $\times$  7.8 mm; 9  $\mu$ M). The mobile phase was 5 mM  $H_2SO_4$  with a flow rate of 0.5 mL/min, maintained at 50°C. For the resting cells experiments, a biochemical analyzer (YSI 2700 Select, Yellow Springs, OH) was used to track Glc consumption throughout the sampling time and all extracellular metabolites were detected and quantified by NMR using an Avance 500 MHz spectrometer with TOPSPIN 2.1 software (Bruker, Germany, Rheinstetten), using conditions and parameters described previously (Kiefer et al., 2007; Millard et al., 2014). Intracellular metabolites were detected by liquid anion exchange chromatography with a Dionex ICS 2000 system (ThermoFisher, Sunnyvale, CA) coupled to a triple-quadrupole mass spectrometer, 4000 QTrap (AB Sciex, Framingham, MA), and analyzed with a multiple reaction monitoring (MRM) mode using previously described conditions (Millard et al., 2014). Data was extracted with Maven software (Clasquin et al., 2012; Melamud et al., 2010) and values were adjusted to the relative concentration of compounds present in the internal standard using the IDMS method (Wu et al., 2005) and normalized to the average biomass obtained in each experiment. The presence of a metabolic pseudo-steady state was suggested by the low variations in the levels of most intracellular metabolites during the time course of the experiment and constant Glc consumption rates. This was further validated in a parallel setup after feeding the resting cells with  $^{13}C$ -labeled Glc and detecting low variations in the isotopic labeling patterns over the time (Supplementary material S2). The yield of total aromatic compounds from glucose ( $Y_{TAC/Glc}$ ) was calculated with the combined molar yields of 3-deoxy-arabinoheptulosonate-7-phosphate (DAHP), quinate (QA), 3-dehydroshikimate (DHS), and shikimate (SA), based on the amounts detected in supernatant samples. This yield was divided by the maximum theoretical yield of aromatic compounds from Glc, estimated as 0.86 mol/mol in strains lacking PTS (Chandran et al., 2003). Concentrations of ppGpp in the extracts were calculated with a calibration curve and adjusted to the biomass. Adenylate energy charge was calculated with the following equation:  $AEC = ([ATP] + \frac{1}{2}[ADP])/([ATP] + [ADP] + [AMP])$ , as reported elsewhere (Chapman et al., 1971; Schuhmacher et al., 2014). Graphs and tables show the average values for metabolite measurements and the error bars represent the standard deviation.

For fermentation data analysis, the growth rate was obtained by adjusting a logistic model to experimental data. Experiments showed a drop in  $OD_{600}$  values after exponential growth, therefore, a zero order parameter ( $-Kd$ ) for biomass decay was added to the logistic model. Yields were first approximated by a regression of linear segments of product versus substrate ( $Y_{p/s}$ ) and biomass



**Figure 2.** Kinetic profiles of the three studied *E. coli* strains grown in batch-mode fermentors with 100 g/L of glucose and 30 g/L of yeast extract. The concentrations of biomass, glucose, acetic acid, and shikimic acid were tracked during 32 h of cultivation; the symbols indicate experimental values and the error bars represent the standard deviation. The model that best fits the experimental values in each case is represented with a line. AR3: filled circles and solid lines; AR3e: open circles and dotted lines; AR36: triangles and dashed lines.

versus substrate ( $Y_{x/s}$ ) data. With these yields, specific exponential consumption and production rates ( $q_{\text{sexp}}$  and  $q_{\text{pexp}}$ , respectively) were calculated by standard equations taking into account the growth rate values. Stationary consumption and production rates ( $q_{\text{ssta}}$  and  $q_{\text{psta}}$ ) were approximated as  $\Delta\text{Glucose}/\Delta\text{Time} \cdot X_{\text{max}}$ , where  $X_{\text{max}}$  is the maximum biomass achieved in the fermentors. Since fermentations were performed on complex media, biomass calculation could render a biased or incorrect comparison between experiments. To provide more accurate comparisons, a consumption and production integrated model was constructed, represented with the following formulas:

$$G_{(t)} = G_{(t-1)} - \left[ (q_{\text{sexp}} X_{(t)} (\Delta t)) \left( 1 - \frac{X_{(t)}}{X_{\text{max}}} \right) \right] - \left[ (q_{\text{psta}} X_{(t)} (\Delta t)) \left( \frac{X_{(t)}}{X_{\text{max}}} \right) \right]$$

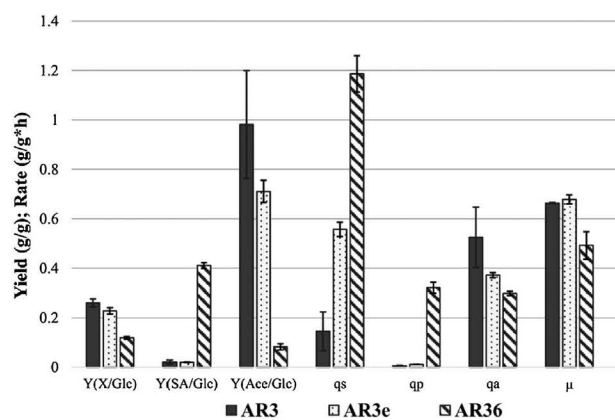
$$S_{(t)} = S_{(t-1)} + \left[ (q_{\text{pexp}} X_{(t)} (\Delta t)) \left( 1 - \frac{X_{(t)}}{X_{\text{max}}} \right) \right] + \left[ (q_{\text{psta}} X_{(t)} (\Delta t)) \left( \frac{X_{(t)}}{X_{\text{max}}} \right) \right]$$

where,  $G$  refers to glucose and  $S$  to any product calculated on our fermentations (shikimate or acetate).  $X_{(t)}$  is the biomass calculated at time  $t$  by the modified logistic model,  $X_{\text{max}}$  is the maximum biomass obtained at fermentations and  $q_{\text{sexp}}$  and  $q_{\text{pexp}}$  are the

specific exponential approximations for glucose consumption and product formation rates. Parameters  $q_{\text{ssta}}$  and  $q_{\text{psta}}$  are the specific stationary approximations for glucose consumption and product formation rates.  $\Delta t$  is the 1 min time step used for simulations. Mathematical model approximation was performed by 0.001 up and down modifications to consumption and production rate values and by optimization of the root mean square of the errors for data versus model. This approximation rendered better descriptors for strain behavior and allowed recalculation of yields and rates.

### <sup>13</sup>C-Flux Calculations

For <sup>13</sup>C-labeling experiments, resting cells were prepared as described above. The unlabeled glucose was replaced by a mixture of 80% [1-<sup>13</sup>C]- and 20% [U-<sup>13</sup>C]-glucose (99% isotopic purity; Eurisotop Saint-Aubin, France). Intracellular metabolites were collected, prepared, and analyzed by IC-MS as described above. The isotopic patterns of metabolites of the glycolytic and pentose phosphate pathways (PEP, G6P, F6P, S7P, F1,6BP, combined pools of 2 and 3 PG, and combined pools of Ru5P, R5P, and Xu5P) were corrected for the presence of naturally occurring isotopes of all elements (except carbon) using IsoCor (Millard et al., 2012). Fluxes were calculated using influx\_s software (Sokol et al., 2012), in which both mass balances and carbon atom transitions describing the biochemical



**Figure 3.** Calculated yields and rates from strains grown in fermentors containing 100 g/L of glucose and 30 g/L of yeast extract. Graphs show the average values of two fermentation runs for each strain with error bars representing the standard deviation.  $Y(X/Glc)$  = biomass to glucose yield;  $Y(SA/Glc)$  = shikimate to glucose yield;  $Y(Ace/Glc)$  = acetate to glucose yield;  $qs$  = specific glucose consumption rate;  $qp$  = specific shikimate production rate;  $qa$  = specific acetate production rate;  $\mu$  = specific growth rate. AR3: solid bars; AR3e: dotted bars; AR36: dashed bars.

reaction network were implemented. The metabolic network contained glycolysis and the pentose phosphate and Entner–Doudoroff pathways. To account for the non-linear relationship between fluxes and isotopic data, the confidence intervals of the fluxes were determined using a Monte-Carlo approach (with 100 iterations). The calculated fluxes were normalized to the rate of substrate uptake, which was arbitrarily set at 100. All experiments were performed in two biological replicates for each strain. The isotopic data, metabolic model, and calculated fluxes for each independent biological replicate are given in Supplementary material S3.

## Results and Discussion

### Plasmid Effects Observed Through Fermentation Profiles

Strain AR3 (containing no plasmid) and its derivatives, AR3e (containing the plasmid without the synthetic operon, pTrc327par) and AR36 (containing the shikimate production plasmid, pTrcAro6) were grown in batch-mode fermentors containing SA production medium supplemented with 100 g/L of glucose (Glc) and 30 g/L yeast extract (YE). These conditions allowed AR36 to completely consume the Glc and produce about 43 g/L of SA, 12 g/L of biomass, and 9 g/L of acetate in 30 h, as previously observed (Rodriguez et al., 2013) (Fig. 2). The same high-substrate conditions were applied to derivatives AR3 and AR3e to challenge them into a prolonged Glc consumption phase and exacerbate their metabolic responses. It is important to notice that even without the pTrcAro6 plasmid, these derivatives still express the SA pathway genes from the chromosome; however, they are controlled by native transcriptional and post-transcriptional regulation (Sprenger, 2006). In this setup, strains AR3 and AR3e consumed roughly half of the available Glc in the fermentors, producing only 1 g/L of SA and about 40 g/L of acetate. Although the low SA titers obtained from these strains is not surprising since they lack the synthetic operon, the large amount of acetate found in the media at the end of fermentations is striking

and evidences a clear metabolic trend towards the production of this compound. In fact, the yields of acetate from Glc ( $Y_{Ace/Glc}$ ) are 12 and 9 times higher in AR3 and AR3e than in AR36, respectively, although the specific acetate production rates ( $qa$ ) are only 1.8 and 1.2 times higher in AR3 and AR3e than in AR36 (Fig. 3). Considering that the enzymes encoded in the synthetic operon are expected to withdraw a significant fraction of glycolytic intermediates toward SA, we propose that this might be the most important factor in reducing acetate formation and allowing consumption of large amounts of Glc. Similar effects have been reported before (albeit with lower substrate concentrations), where the expression of enzymes that increase ethanol production caused less organic acid accumulation and improved growth, compared to strains with empty plasmids (Diaz-Ricci et al., 1992; Ingram and Conway, 1988). Moreover, it is known that acetate can be toxic from concentrations as low as 2.5 g/L (Luli and Strohl, 1990), so it seems likely that the measured concentrations of 40 g/L are causing the decline in biomass concentration observed at the end of AR3e and AR3 cultivations (Fig. 2). Given that SA is less toxic than acetate; the results presented here support previous discussions on how the expression of a biosynthetic pathway can provide a metabolic advantage by carbon redirection (Diaz Ricci and Hernández, 2000). The model correctly fit all the experimental data with the exception of the unusual acetate concentration profile observed with AR36 (Fig. 2). The latter profile indicates a particular acetate metabolism in this condition which may also contribute to the lower acetate titers obtained with this strain.

It is remarkable that the sole presence of the empty plasmid induced a 4-fold increase in the specific Glc consumption rate ( $qs$ ) and eliminated the lag phase observed during growth of strain AR3, allowing AR3e to reach maximum biomass before the other variants (Fig. 2). However, this effect was not reflected in an increased exponential growth rate ( $\mu$ ) or biomass to Glc yield ( $Y_{X/Glc}$ ) in AR3e, compared to AR3 (Fig. 3). In addition, the three strains produced roughly the same amount of biomass but at different times, evidencing different physiological states. The strain expressing the SA biosynthetic operon displayed a further increase in the value for  $qs$ , representing twice the value obtained with the empty plasmid or eight times the one obtained by the strain without plasmid, accompanied by a 30% drop in the value of  $\mu$  compared to the other strains. On this regard, the presence of high amounts of YE in the medium is mainly provided as a way to counteract the need for supplementation of aromatic amino acids in strains lacking both shikimate kinase enzymes, but is also used to increase biomass formation. By growing in the presence of YE, the cells should have more metabolic precursors, energy, and reducing power to compensate for the drain of resources that involves replicating a plasmid and expressing its encoded proteins (Carneiro et al., 2013). In particular, up to 1000-fold change (8–10 qPCR cycles) in mRNA abundance of the operon genes were detected in AR36 compared to AR3e, after these derivatives were grown in SA production media and sampled in mid-exponential phase (Supplementary material S1c). If this level of expression is in fact creating a higher demand of resources, it could also explain the observed increase in  $qs$  for strain AR36 to produce nucleotide and amino acid precursors derived from Glc. Nevertheless, since the YE is limiting biomass production by providing a finite amount of aromatic amino acids (that cannot



be synthesized from Glc in these derivatives) it seems reasonable that the large differences in  $q_s$  among the three strains are not translating into large differences in biomass yields.

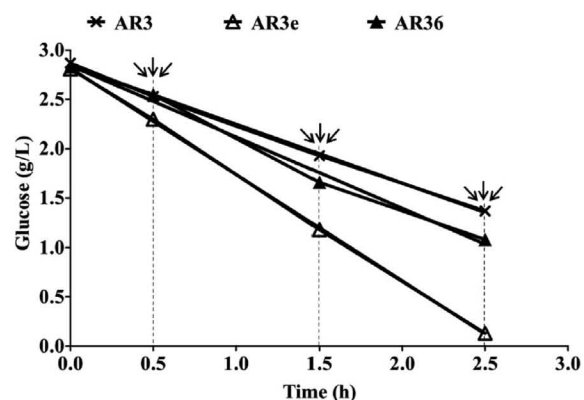
We acknowledge that both Glc and YE may contribute to the generation of precursors for biomass and SA production. Even when it is difficult to quantify the extent of these contributions, we have previously shown that an increase in YE concentration does not result in a significant change in SA titers or yields but correlates linearly with an increase in the biomass concentration. On the contrary, SA titers changed proportionally with the Glc concentration (Rodriguez et al., 2013). As a control, when AR36 was cultivated in fermentors containing 30 g/L of YE but no added Glc, only one-fourth of the biomass and less than 1.5 g/L of SA were produced (Supplementary material S4). This suggests that SA is mainly being produced from Glc in these conditions and there may be enough YE-derived components to maintain balanced growth. However, the physiological responses to plasmid presence observed here cannot be generalized until other plasmids and genetic backgrounds are also evaluated under the same conditions. For example, the inclusion of gene *zwf* in the synthetic operon might influence the growth capabilities of AR36, since it has been shown that this gene can recover the decrease in  $\mu$  caused by the plasmid-based expression of a recombinant protein (Flores et al., 2004). A direct comparison of the effects of expressing gene *zwf* under SA production conditions still remains to be tested.

### Establishment of a Metabolically Active Resting Cells State

Functional analyses of metabolic networks offer the possibility of obtaining a snapshot of the central metabolism in a defined physiological state and they have become an important tool to evaluate the impact of metabolic engineering strategies (Ellis and Goodacre, 2012). We decided to continue strain characterization by tracking and quantifying key intermediates from central carbon metabolism and the SA pathway and by calculating intracellular metabolic fluxes in the main metabolic pathways involved with SA production. However, the enriched culture medium used to produce SA in fermentors is not compatible with fine metabolite analyses where defined media are preferred. This became a particular issue because strain AR36 was not able to grow on minimal medium supplemented with aromatic amino acids and vitamins (even though strains AR3 and AR3e were able to), suggesting the presence of a constrained metabolism in AR36 that is alleviated by the presence of YE. To solve this situation and be able to make a fair comparison of the three strains, we decided to prepare a resting cell system from cells grown in rich media with 25 g/L of Glc and 15 g/L of YE that were then washed and transferred to minimal media with 3 g/L of Glc as only carbon source. With this procedure, it was possible to keep an active Glc consuming state and arrest cell growth without having to add chloramphenicol or limit nitrogen, as it is often done (Emmerling et al., 1999; Julsing et al., 2012). Even when this simulated state will not provide information about biomass generation and omits the contribution of YE to

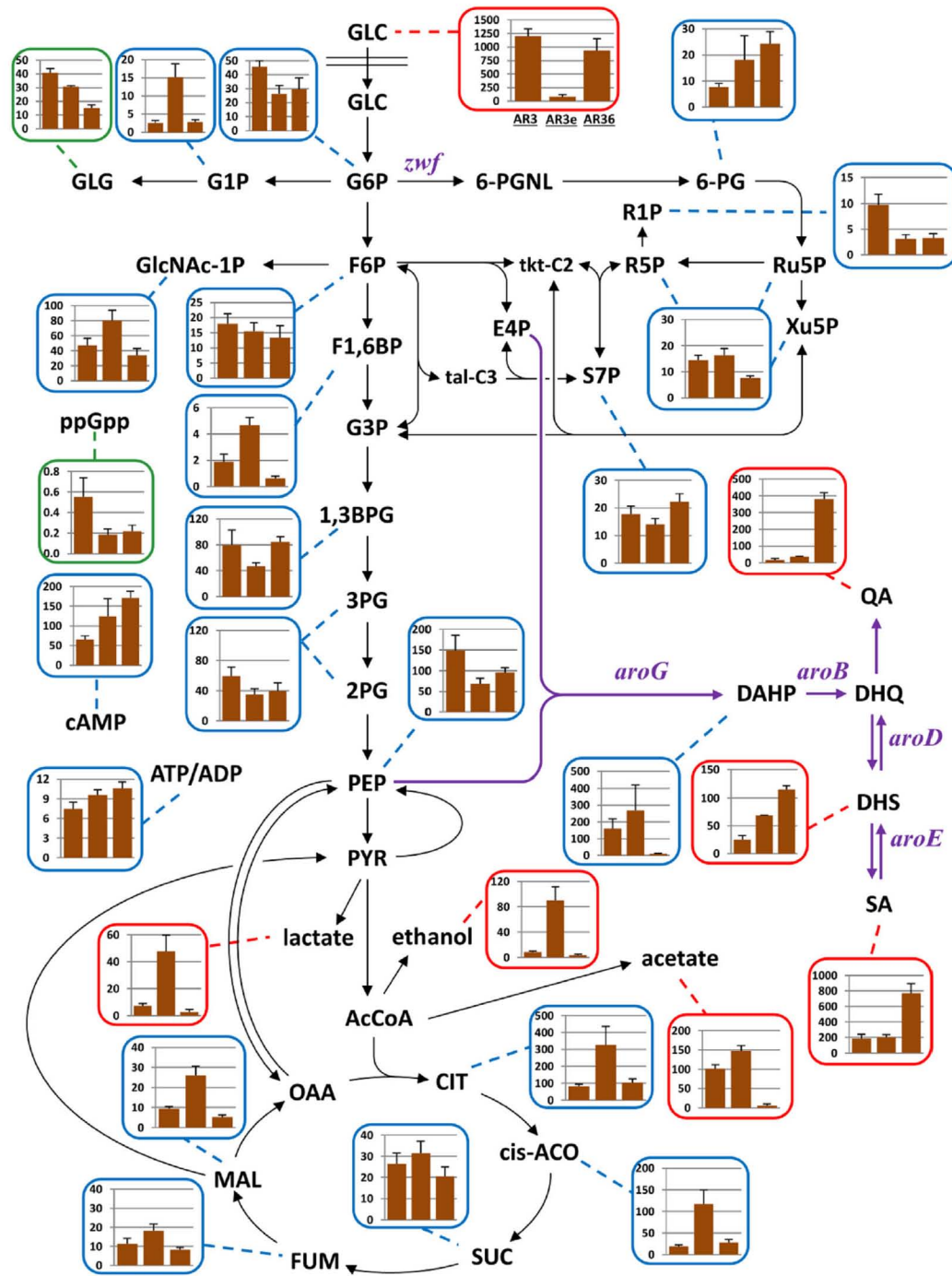
metabolism, it is still a useful setup to determine differences in the relative levels of intermediates derived from Glc, which partially fuel the central metabolism and can also get excreted as organic acids.

In these conditions, we observed differential  $q_s$  values but constant Glc consumption from the start of the experiments, suggesting the cells are likely in a metabolic pseudo-steady state (Fig. 4). Plasmid presence in AR3e considerably increased the  $q_s$  compared to the plasmid-less derivative AR3, keeping the same trend than the values obtained for these two strains when grown in rich media. Strain AR36, however, halved its  $q_s$  in the resting condition compared to fermentors with rich media, but retained similar  $q_p$  and  $Y_{SA/Glc}$  values (Figs. 3 and 4). The use of defined media made possible to detect all major compounds present in the supernatant after 2.5 h of starting the resting cell experiments and SA was the most abundant extracellular product observed in all cases. The SA pathway intermediates quinic acid and 3-dehydroshikimic acid were also detected in larger amounts in the supernatants obtained from strain AR36, as expected for a strain that redirects more precursors toward the SA pathway (Fig. 5). As a result, the total molar yield of aromatic compounds relative to the maximum theoretical yield ( $Y_{TAC/Glc}/Y_{max}$ ) obtained with strain AR36 was 81% (Fig. 4), also resembling previously reported values in fermentor experiments, where 67% was attained (Rodriguez et al., 2013). These results show that while YE presence is beneficial to AR36 by helping growth and promoting faster Glc



	AR3	AR3e	AR36
X (g/L)	1.63 ± 0.10	1.65 ± 0.04	1.59 ± 0.02
$q_s$ (g/g <sup>3</sup> h)	-0.36 ± 0.02	-0.66 ± 0.01	-0.48 ± 0.03
$q_p$ (g/g <sup>3</sup> h)	0.06 ± 0.05	0.07 ± 0.01	0.20 ± 0.02
$Y_{SA/Glc}$ (g/g)	0.11 ± 0.02	0.07 ± 0.01	0.39 ± 0.04
$Y_{TAC/Glc}/Y_{max}$ (%)	17	14	81
AEC	0.51	0.54	0.51

**Figure 4.** Glucose consumption profiles and metabolic parameters obtained in the resting cells system. The arrows show the sampling times that were used for collection and quantitation of extracellular and intracellular metabolites. Trend lines to the experimental values are graphed to show near-linear consumption rates during the sampling times. Data in the table represent the average values and standard deviations of the three time points analyzed for each strain. X = biomass;  $q_s$  = specific glucose consumption rate;  $q_p$  = specific shikimate production rate;  $Y_{SA/Glc}$  = shikimate to glucose yield;  $Y_{TAC/Glc}/Y_{max}$  = total molar yield of aromatic compounds relative to the maximum theoretical yield; AEC = adenylate energy charge.



**Figure 5.** Comparative metabolomics in the resting cells system. Graphs are assigned to metabolites by dashed lines and the bars represent the average values and standard deviations of metabolites collected over the time-course of the experiments. In all cases, the first bar from the left corresponds to strain AR3, the middle bar corresponds to strain AR3e, and the right-side bar represents strain AR36. Graphs enclosed in red boxes represent extracellular metabolites and their y-axis values are displayed in mg/L. Graphs enclosed in blue boxes represent intracellular metabolites and their y-axis values are displayed in intensity units relative to the amounts present in the mixture of  $^{13}\text{C}$ -labeled external standards. Graphs enclosed in green boxes represent intracellular concentrations and their y-axis values are displayed in milligrams per gram of dry cell weight (for glycogen) or micrograms per gram of dry cell weight (for guanosine tetraphosphate). The six plasmid-expressed genes in strain AR36 are indicated in purple. GLC = glucose; G6P = glucose 6-phosphate; G1P = glucose 1-phosphate; GLG = glycogen; 6-PGNL = 6-phosphogluconolactone; 6-PG = 6-phosphogluconate; Ru5P = ribulose 5-phosphate; Xu5P = xylulose 5-phosphate; R5P = ribose 5-phosphate; R1P = ribose 1-phosphate; G3P = glyceraldehyde 3-phosphate; S7P = sedoheptulose 7-phosphate; E4P = erythrose 4-phosphate; F6P = fructose 6-phosphate; GlcNAc-1P = N-acetylglucosamine 1-phosphate; F1,6BP = fructose 1,6-bisphosphate; 1,3BPG = 1,3-bisphosphoglycerate; 3PG = 3-phosphoglycerate; 2PG = 2-phosphoglycerate; PEP = phosphoenolpyruvate; PYR = pyruvate; AcCoA = acetyl coenzyme-A; CIT = citrate; cis-ACO = cis-aconitate; SUC = succinate; FUM = fumarate; MAL = malate; OAA = oxaloacetate; cAMP = cyclic adenosine monophosphate; ppGpp = guanosine tetraphosphate; ATP/ADP = ratio of adenosine triphosphate to adenosine diphosphate; DAHP = 3-dehydroarabinoheptulosonate 7-phosphate; DHQ = 3-dehydroquininate; QA = quinate; DHS = 3-dehydroshikimate; SA = shikimate; *zwf* = glucose 6-phosphate dehydrogenase; *tktA* = transketolase I; *aroG* = DAHP synthase; *aroB* = DHQ synthase; *aroD* = DHQ dehydratase; *aroE* = SA dehydrogenase.

consumption, the carbon partition and redirection of glycolytic intermediates to SA in the resting cells appears to be similar to the high-substrate production conditions. Altogether, these observations support the use of the resting cell system described here to simulate and assay the glycolytic metabolism (without YE-derived components) and the production of SA from Glc in fermentors. As discussed in the next section, this system is relevant because it allows for detection of pathway bottlenecks or depletion of precursors and permits faster small-scale comparisons of the consequences that genetic or environmental perturbations can have over the SA yield.

## Comparative Metabolomics

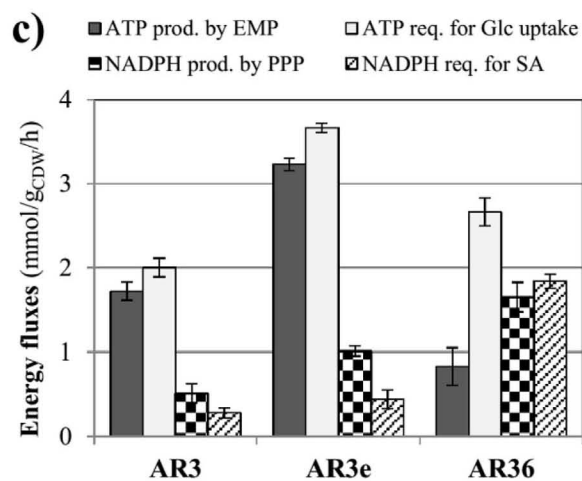
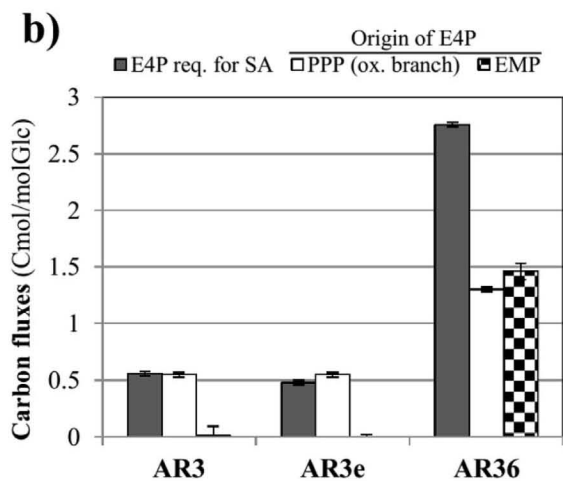
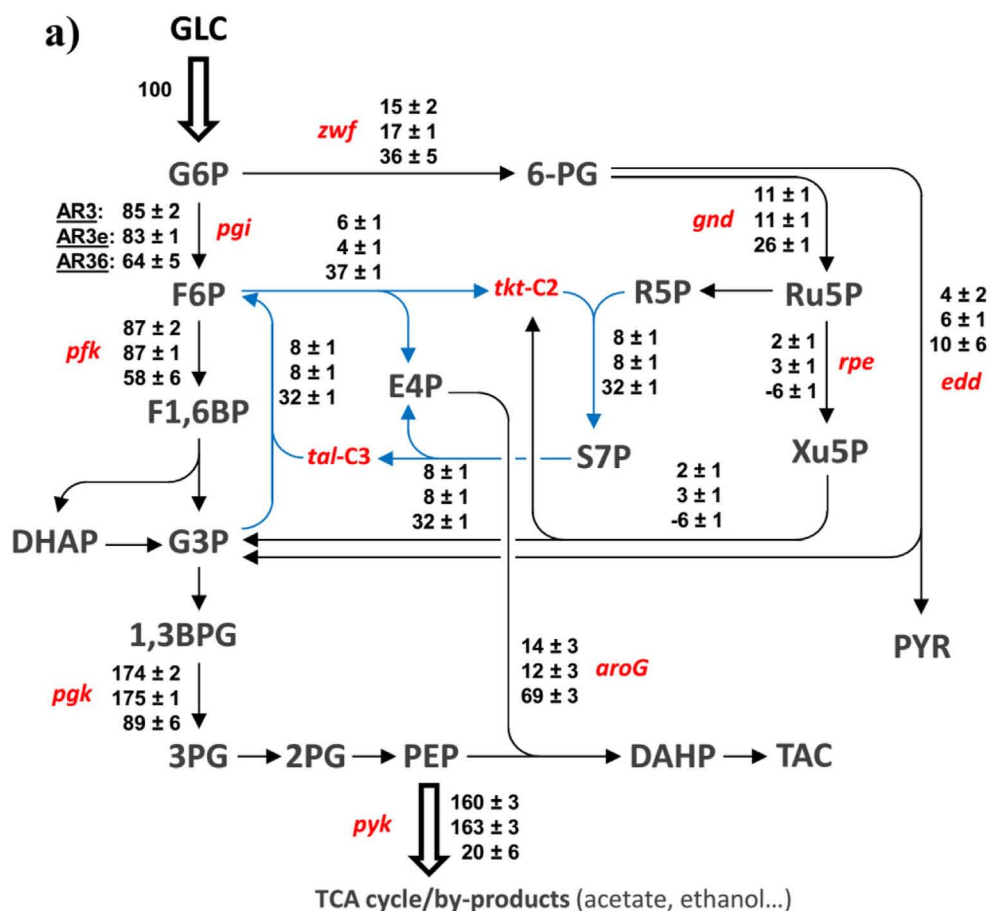
The resting cell cultures were used to track seven extracellular metabolites and 25 intracellular metabolites over time by measuring the absolute concentration of the first group in the supernatant and the relative concentration of the second group in cell extracts. The metabolite profiles are shown in Figure 5. Intermediates in red boxes refer to the absolute extracellular concentrations measured at the point of highest SA accumulation, intermediates in blue boxes indicate relative intracellular concentrations and metabolites in green boxes indicate absolute intracellular concentrations per unit of biomass. The columns in every graph, from left to right, refer to the average values obtained in AR3, AR3e, and AR36, respectively.

Strain AR3e, which consumes Glc faster than the other derivatives, accumulates higher concentrations of fermentation products, intermediates of the tricarboxylic acid (TCA) cycle, glucose-1-phosphate (G1P) as precursor of glycogen (GLG), and N-acetylglucosamine-1-phosphate (GlcNAc-1P), used in cell wall synthesis. In contrast, AR36 did not accumulate lactate, ethanol or acetate, probably because the overexpression of genes in the operon redirects glycolytic intermediates toward the SA pathway. As discussed previously, this may prevent toxic product formation, like acetate or ethanol, that accumulate almost 100 times more in AR3e than in AR36 (Fig. 5). On the other hand, the observed intracellular levels of fructose-1,6-bisphosphate (F1,6BP) are lower in AR36 than in AR3e. This could be partially explained by the overexpression of transketolase I enzyme (coded by *tktA*), which promotes conversion of F6P into erythrose-4-phosphate (E4P), presumably reducing the amount of substrate for the phosphofructokinase reaction and diminishing the flux to F1,6BP (see next section). However, phosphoenolpyruvate (PEP) concentration is not comparatively low in AR36, despite the overexpression of the 3-deoxy-arabinoheptulosonate-7-phosphate (DAHP) synthase (encoded by *aroG<sup>fbt</sup>*). The expression of this enzyme is also not resulting in a relatively higher intracellular concentration of DAHP in AR36, suggesting that AroB level is sufficient and does not likely significantly limit the flux (Fig. 5). The ATP/ADP ratio does not significantly differ between the three strains, being close to the value of 10 measured in the K-12 wild-type strain during exponential growth on Glc (Koebsmann et al., 2002). The adenylate energy charge (AEC), which is also very similar in the three strains

with a value of 0.5 (Fig. 4), corresponds to values typically found in metabolically active cells in stationary phase (Chapman et al., 1971). Therefore, it seems that there are active systems to prevent depletion of essential intermediates, such as G6P, F6P, 1,3BPG, PEP, and ATP, as previously suggested (Diaz Ricci and Hernández, 2000). This might be controlled to some extent by the low F1,6BP concentration in AR36, consistent with previous findings where it was demonstrated that a low F1,6BP level limits PEP depletion by reducing the PEP carboxylase activity (Xu et al., 2012). This response, which would save PEP for glucose transport via the PTS in a wild-type strain, is probably improving PEP availability and SA formation in the PTS<sup>-</sup> derivatives studied here.

Moreover, the lower Glc consumption displayed by strain AR3 may be a consequence of the relatively higher concentrations of intracellular intermediates like ribose-1-phosphate (R1P), G6P, F6P, and PEP. It is important to recall that all studied strains have a PTS<sup>-</sup> *pykF*<sup>-</sup> background, constructed precisely with the purpose of increasing PEP availability from Glc but knowing beforehand that a concomitant decrease in  $\mu$  and  $q_s$  values is commonly observed as a response to these inactivations (Meza et al., 2012; Ponce, 1999). It is after plasmid transformation that the metabolism changes to a more active state, presumably by redirecting carbon precursors toward amino acid and nucleotide biosynthetic pathways, as well as energy generation pathways like the TCA cycle (Fig. 5). A sound theory is that this switch is activated as a response to the need to replicate the plasmid and express the encoded proteins, mediated by a change in the transcriptional regulatory networks and leading to an increase in  $q_s$  (Diaz Ricci and Hernández, 2000). In AR3e, we detected strong build-up of most of the TCA cycle intermediates that may be in agreement with the higher flux in the TCA cycle. Furthermore, it is interesting to note that this accumulation correlates with the production of lactate and ethanol which are not expected in aerobic condition.

There is evidence that the central metabolic response to plasmid presence partially depends on a complex regulatory network mediated by cAMP-CRP and Cra transcriptional regulators (Ow et al., 2009). Moreover, it has been proposed that plasmids can increase the PTS-dependent sugar transport by means of a positive regulation of the glucose transporter gene *ptsG* by the complex cAMP-CRP, correlating with increased cAMP levels and the expression of catabolic operons (Botsford and Harman, 1992; Diaz Ricci and Hernández, 2000). The results presented here support the presence of such effect in a PTS<sup>-</sup> strain, since an increase in cAMP concentration was observed in plasmid-containing derivatives AR3e and AR36, compared to AR3 (Fig. 5). This means that other transporters could potentially be upregulated by cAMP-CRP in these conditions, for example GalP, which is responsible for the main Glc import capabilities in the parental strain PB12 (Flores et al., 2007). Additionally, F1,6BP is considered as a metabolic flux sensor in *E. coli*, where its interaction with Cra regulates the expression of central metabolic genes (Kochanowski et al., 2013). Based on this regulatory model, the high F1,6BP level observed in AR3e is expected to decrease Cra activity and to



**Figure 6.** Comparative fluxomics in the resting cells system. In vivo flux distributions in the central metabolism of strains AR3 (top values), AR3e (middle values), and AR36 (bottom values) are shown in panel (a). All fluxes are normalized to the specific Glc uptake rate, which is set to 100%. Values represent the mean  $\pm$  standard deviation of net fluxes estimated from two independent biological replicates for each strain. For reversible reactions, the direction of the net flux is indicated by an arrow. The carbon fluxes of E4P formation and consumption are given in panel (b) and the ATP and NADPH balances are shown in panel (c).

enhance the glycolytic flux. In contrast, in AR36 (or to a lesser extent in AR3), the low F1,6BP level is expected to increase Cra activity and reduce the glycolytic flux. Interestingly, we found that the plasmid also causes a drop in ppGpp concentrations in

exponential phase (Fig. 5). Values of 0.55, 0.18, and 0.22  $\mu\text{g/g}_{\text{CDW}}$  were detected for strains AR3, AR3e, and AR36, respectively, showing an inverse correlation to the  $q_s$  trends. This effect has been reported before (Hoffmann and

Rinas, 2004; Ow et al., 2009; Silva et al., 2012) and suggests that the plasmid may indirectly decrease the stress response of the cell in the tested conditions.

## Comparative Fluxomics

To get a comprehensive view of the metabolic state in the three derivative strains,  $^{13}\text{C}$  metabolic flux analysis was performed to estimate carbon flux distributions within the central metabolism and the SA pathway. Resting cell cultures were used in identical conditions as for metabolomics, substituting glucose by  $^{13}\text{C}$ -glucose, and metabolic fluxes were determined from the measured  $^{13}\text{C}$ -labeling patterns of seven intracellular metabolites and from extracellular fluxes (Fig. 6a). First of all, by sampling  $^{13}\text{C}$ -labeled metabolites at different time points, we detected low variations in the isotopic labeling patterns over the time (Supplementary material S2) validating that the resting cells experiment provides a metabolic and isotopic pseudo-steady state.

In AR3, 15% of the Glc is metabolized through the pentose phosphate pathway (PPP) (Zwf reaction), with a negligible contribution of the Entner–Doudoroff (ED) pathway. The flux through the oxidative branch of the PPP is thus significantly higher than previously reported for the parental strain PB12 (Flores et al., 2002). This indicates that the resting cell state and/or the additional genetic modifications of the PB12 strain contribute to rerouting carbon from glycolysis to the oxidative branch of the PPP. It is important to note that, in AR3, the oxidative branch of the PPP alone is sufficient to support the E4P demand for SA production (Fig. 6b). The flux distribution in the strain AR3e is similar to that of the strain AR3, indicating that the presence of the plasmid increases the global activity of the central metabolism (since the  $q_s$  is higher in AR3e) but does not impact the intracellular flux partition (Fig. 6a). In contrast, 36% of the Glc is metabolized through Zwf in the strain AR36, which is more than twice the flux observed in AR3 and AR3e, as intended by overexpressing the *zwf* gene. This flux primarily feeds the oxidative branch of the PPP in AR36, with a low contribution of the ED pathway. Despite this increased flux through Zwf and the 6-phosphogluconate dehydrogenase (Gnd) in AR36, the oxidative branch of the PPP alone is not sufficient to sustain the E4P demand for SA production, and about 50% of the E4P is supplied by the Embden–Meyerhof–Parnas (EMP) pathway (Fig. 6b) via the non-oxidative branch of the PPP. This metabolic remodeling is particularly striking at the level of the two-carbon block transfer reaction from xylulose-5-phosphate (Xu5P) to the transketolase, which reverses in AR36 leading to a slight formation of Xu5P from glyceraldehyde-3-phosphate (G3P).

Another important finding in AR36 is the significant increase of flux through both the transketolase (which is overexpressed) and transaldolase reactions. This leads to the establishment of a very active metabolic cycle which roughly produces two molecules of E4P from one molecule of ribose-5-phosphate (R5P) and one of G3P (displayed in blue in Fig. 6a). The increased contributions of both the oxidative and non-oxidative branches of the PPP in AR36 to sustain the E4P demand significantly decrease the carbon flux. This is consistent with the lower F1,6BP level observed in this strain, which increases the activity of Cra and thereby, presumably inhibits the expression of glycolytic genes while activating gluconeogenic

(e.g., *pck*) genes. The combined decrease of glycolytic flux and increase of demand in PEP for SA production, reduces the carbon available for the rest of metabolism (TCA cycle and byproducts formation), which is eight times lower in AR36 compared to AR3 and AR3e. This strengthens the hypothesis that the redirection of carbon towards the SA pathway in AR36 may be the primary cause of the reduced production of acetate, ethanol, and lactate.

Finally, we calculated the ATP and NADPH balances from the estimated intracellular carbon fluxes, considering only the EMP and PPP pathways, respectively (Fig. 6c). In this scenario, the EMP produces 86% and 88% of the ATP required for Glc uptake by gluco- and phosphofructo-kinases in strains AR3 and AR3e, respectively, but only 31% in AR36. This indicates that additional energy-producing processes (e.g., the TCA cycle or oxidative phosphorylation) are necessary in this strain to sustain continuous Glc uptake and ATP needs for maintenance. This condition might be one of the causes of the observed decrease in  $q_s$  values when AR36 was transferred to the resting cells system and would also explain the lack of growth in the absence of YE, while AR3 and AR3e were able to. In contrast to ATP, the NADPH requirements for SA production (i.e., one mole of NADPH per mole of SA) are balanced with the NADPH formation by Zwf and Gnd in the three strains (Fig. 6c).

## Conclusions

In this paper we describe the implementation and characterization of a resting cell state which was useful to emulate the SA production profiles of the strain AR36 in fermentation conditions, albeit only from Glc. It was observed that plasmid pTrc327par increases the  $q_s$  and the intracellular levels of metabolites involved in energy generation pathways, therefore preventing depletion of PEP and ATP, which might translate into a lower stress state. When the biosynthetic operon is expressed from this plasmid, the engineered strain was able to use the more active glycolytic metabolism to consume up to 100 g/L of Glc and produce SA, which worked as an escape valve for intermediates that would otherwise be converted into more toxic fermentation products. This is the first reported metabolomics and fluxomics characterization of an *E. coli* strain that overproduces SA with high titers and yields. Expressing the production plasmid in AR36 increases the demand in E4P and PEP for SA synthesis and results in a major, global remodeling of the carbon and energy metabolism. Consequently, the carbon available for oxidative and fermentative pathways is reduced, and both the oxidative and the non-oxidative branches of the PPP are strongly activated to supply E4P and balance NADPH requirements for SA production.

The severe combined modifications to Glc transport functions and the inactivation of *pykF*, *aroK*, and *aroL* performed during construction of strain AR36, cause a phenotype characterized by a constrained metabolism to increase precursor availability for the production of aromatic compounds. This is the phenotype displayed by strains AR3 and AR3e, which do not show optimal traits for a production strain. Here we provide an example of how this situation was turned into an advantage by redirecting intermediates towards SA with a constitutive expression system; nevertheless, supplementation of YE to the culture medium seems to be required and a large amount of acetate is still being accumulated at the end of the

fermentation. A further comparison of the effects caused by other combinations of plasmids and genetic backgrounds would provide valuable insights into the metabolism of engineered *E. coli* strains.

The authors are grateful to MetaToul (Metabolomics & Fluxomics facilities, Toulouse, France, [www.metatoul.fr](http://www.metatoul.fr)), which is part of the national infrastructure MetaboHUB-ANR-11-INBS-0010 (The French National Infrastructure for Metabolomics and Fluxomics, [www.metabohub.fr](http://www.metabohub.fr)), for access to NMR and MS facilities. Support from Lindsay Peyriga and Lara Gales (MetaToul) and from Noemi Flores and Ramon de Anda (IBT-UNAM) is gratefully acknowledged.

## References

- Aguilar C, Escalante A, Flores N, de Anda R, Riveros-McKay F, Gosset G, Morett E, Bolivar 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.
- Ahn J, Chung BKS, Lee D, Park M, Karimi IA, Jung J, Lee H. 2011. NADPH-dependent *pgi*-gene knockout *Escherichia coli* metabolism producing shikimate on different carbon sources. *FEMS Microbiol Lett* 324:10–16.
- Bogossian G, Frantz JP. 2013. Use of glyphosate to produce shikimic acid in microorganisms. US Patent 8, 435,769 B2.
- Botsford JL, Harman JG. 1992. Cyclic AMP in prokaryotes. *Microbiol Rev* 56:100–122.
- Carneiro S, Ferreira EC, Rocha I. 2013. Metabolic responses to recombinant bioprocesses in *Escherichia coli*. *J Biotechnol* 164:396–408.
- Chandran SS, Yi J, Draths KM, von Daeniken R, Weber W, Frost JW, Daeniken R Von, Weber W, Frost JW. 2003. Phosphoenolpyruvate availability and the biosynthesis of shikimic acid. *Biotechnol Prog* 19:808–814.
- Chapman AG, Fall L, Atkinson DE. 1971. Adenylate energy charge in *Escherichia coli* during growth and starvation. *J Bacteriol* 108:1072–1086.
- Chen K, Dou J, Tang S, Yang Y, Wang H, Fang H, Zhou C. 2012. Deletion of the *aroK* gene is essential for high shikimic acid accumulation through the shikimate pathway in *E. coli*. *Bioresour Technol* 119:141–147.
- Chen X, Li M, Zhou L, Shen W, Algasan G, Fan Y, Wang Z. 2014. Metabolic engineering of *Escherichia coli* for improving shikimate synthesis from glucose. *Bioresour Technol* 166:64–71.
- Clasquin MF, Melamud E, Rabinowitz JD. 2012. LC-MS data processing with MAVEN: a metabolomic analysis and visualization engine. *Curr Protoc Bioinforma* 37:14.11.1–14.11.23.
- Cui Y-Y, Ling C, Zhang Y-Y, Huang J, Liu J-Z. 2014. Production of shikimic acid from *Escherichia coli* through chemically inducible chromosomal evolution and cofactor metabolic engineering. *Microb Cell Fact* 13:21.
- Diaz-Ricci JC, Bode J, Jong Il Rhee, Schugerl K. 1995. Gene expression enhancement due to plasmid maintenance. *J Bacteriol* 177:6684–6687.
- Diaz-Ricci JC, Tsu M, Bailey JE. 1992. Influence of expression of the *pet* operon on intracellular metabolic fluxes of *Escherichia coli*. *Biotechnol Bioeng* 39:59–65.
- Díaz Quiroz DC, Carmona SB, Bolivar F, Escalante A. 2014. Current perspectives on applications of shikimic and aminoshikimic acids in pharmaceutical chemistry. *Res Reports Med Chem* 4:35–46.
- Diaz Ricci JC, Hernández ME. 2000. Plasmid effects on *Escherichia coli* metabolism. *Crit Rev Biotechnol* 20:79–108.
- Donovan RS, Robinson CW, Click BR. 1996. Review: Optimizing inducer and culture conditions for expression of foreign proteins under the control of the *lac* promoter. *J Ind Microbiol* 16:145–154.
- Ellis DI, Goodacre R. 2012. Metabolomics-assisted synthetic biology. *Curr Opin Biotechnol* 23:22–28.
- Emmerling M, Bailey JE, Sauer U. 1999. Glucose catabolism of *Escherichia coli* strains with increased activity and altered regulation of key glycolytic enzymes. *Metab Eng* 1:117–127.
- Escalante A, Calderón R, Valdivia A, de Anda R, Hernández G, Ramírez OT, Gosset G, Bolivar F. 2010. Metabolic engineering for the production of shikimic acid in an evolved *Escherichia coli* strain lacking the phosphoenolpyruvate:carbohydrate phosphotransferase system. *Microb Cell Fact* 9:21.
- Estévez AM, Estévez RJ. 2012. A short overview on the medicinal chemistry of (–)-shikimic acid. *Mini Rev Med Chem* 12:1443–1454.
- Flores S, Gosset G, Flores N, De Graaf AA, Bolivar F. 2002. Analysis of carbon metabolism in *Escherichia coli* strains with an inactive phosphotransferase system by (<sup>13</sup>C) labeling and NMR spectroscopy. *Metab Eng* 4:124–137.
- 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. *J Mol Microbiol Biotechnol* 13:105–116.
- Flores S, de Anda-Herrera R, Gosset G, Bolívar F. 2004. Growth-rate recovery of *Escherichia coli* cultures carrying a multicopy plasmid, by engineering of the pentose-phosphate pathway. *Biotechnol Bioeng* 87:485–494.
- Frost JW, Knop DR. 2002. Biocatalytic synthesis of shikimic acid. US patent 6, 472,169 B1.
- Ghosh S, Chisti Y, Banerjee UC. 2012. Production of shikimic acid. *Biotechnol Adv* 30:1425–1431.
- Glick BR. 1995. Metabolic load and heterologous gene expression. *Biotechnol Adv* 13:247–261.
- Hoffmann F, Rinas U. 2004. Stress induced by recombinant protein production in *Escherichia coli*. *Adv Biochem Eng Biotechnol* 89:73–92.
- Ingram LO, Conway T. 1988. Expression of different levels of ethanologenic enzymes from *Zymomonas mobilis* in recombinant strains of *Escherichia coli*. *Appl Environ Microbiol* 54:397–404.
- Johansson L, Lindskog A, Silfversparre G, Cimander C, Nielsen KF, Lidén G. 2005. Shikimic acid production by a modified strain of *E. coli* (W3110.shik1) under phosphate-limited and carbon-limited conditions. *Biotechnol Bioeng* 92:541–552.
- Jones KL, Kim SW, Keasling JD. 2000. Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. *Metab Eng* 2:328–338.
- Julsing MK, Kuhn D, Schmid A, Bühler B. 2012. Resting cells of recombinant *E. coli* show high oxidation yields on energy source and high sensitivity to product inhibition. *Biotechnol Bioeng* 109:1109–1119.
- Kiefer P, Nicolas C, Letisse F, Portais JC. 2007. Determination of carbon labeling distribution of intracellular metabolites from single fragment ions by ion chromatography tandem mass spectrometry. *Anal Biochem* 360:182–188.
- Kim HK, Park KJ. 2012. A new efficient synthesis of oseltamivir phosphate (Tamiflu) from (–)-shikimic acid. *Tetrahedron Lett* 53:1561–1563.
- Kochanowski K, Volkmer B, Gerosa L, Haverkorn van Rijsewijk BR, Schmidt A, Heinemann M. 2013. Functioning of a metabolic flux sensor in *Escherichia coli*. *Proc Natl Acad Sci USA* 110:1130–1135.
- Koebmann BJ, Westerhoff HV, Snoep JL, Nilsson D, Jensen PR. 2002. The glycolytic flux in *Escherichia coli* is controlled by the demand for ATP. *J Bacteriol* 184:3909–3916.
- Krisman C. 1962. A method for the colorimetric estimation of glycogen with iodine. *Anal Biochem* 23:17–23.
- Luli GW, Strohl WR. 1990. Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations. *Appl Environ Microbiol* 56:1004–1011.
- 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.
- Melamud E, Vastag L, Rabinowitz JD. 2010. Metabolomic analysis and visualization engine for LC-MS data. *Anal Chem* 82:9818–9826.
- Meza E, Becker J, Bolivar F, Gosset G, Wittmann C. 2012. Consequences of phosphoenolpyruvate: Sugar phosphotransferase system and pyruvate kinase isozymes inactivation in central carbon metabolism flux distribution in *Escherichia coli*. *Microb Cell Fact* 11:127.
- Millard P, Letisse F, Sokol S, Portais J-C. 2012. IsoCor: Correcting MS data in isotope labeling experiments. *Bioinformatics* 28:1294–1296.
- Millard P, Massou S, Wittmann C, Portais JC, Letisse F. 2014. Sampling of intracellular metabolites for stationary and non-stationary <sup>13</sup>C metabolic flux analysis in *Escherichia coli*. *Anal Biochem* 465:38–49.
- Nie LD, Ding W, Shi XX, Quan N, Lu X. 2012. A novel and high-yielding asymmetric synthesis of oseltamivir phosphate (Tamiflu) starting from (–)-shikimic acid. *Tetrahedron: Asymmetry* 23:742–747.
- Ow DSW, Lee D-Y, Tung H-H, Lin-Chao S. 2009. Plasmid regulation and systems-level effects on *Escherichia coli* metabolism. In: Lee SY, editor. *Systems biology and biotechnology of Escherichia coli*, Netherlands: Springer. p 237–256.
- Ponce E. 1999. Effect of growth rate reduction and genetic modifications on acetate accumulation and biomass yields in *Escherichia coli*. *J Biosci Bioeng* 87:775–780.

- Rawat G, Tripathi P, Saxena RK. 2013. Expanding horizons of shikimic acid. Recent progresses in production and its endless frontiers in application and market trends. *Appl Microbiol Biotechnol* 97:4277–4287.
- Rodríguez A, Martínez JA, Báez-Viveros JL, Flores N, Hernández-Chávez G, Ramírez OT, Gosset G, Bolívar F. 2013. 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 Fact* 12:86.
- Rodríguez A, Martínez JA, Flores N, Escalante A, Gosset G, Bolívar F. 2014. Engineering *Escherichia coli* to overproduce aromatic amino acids and derived compounds. *Microb Cell Fact* 13:126.
- Schuhmacher T, Löffler M, Hurler T, Takors R. 2014. Phosphate limited fed-batch processes: Impact on carbon usage and energy metabolism in *Escherichia coli*. *J Biotechnol* 190:96–104.
- Shim J-H, Park J-T, Hong J-S, Kim KW, Kim M-J, Auh J-H, Kim Y-W, Park C-S, Boos W, Kim J-W, Park K-H. 2009. Role of maltogenic amylase and pullulanase in maltodextrin and glycogen metabolism of *Bacillus subtilis* 168. *J Bacteriol* 191:4835–4844.
- Silva F, Queiroz JA, Domingues FC. 2012. Evaluating metabolic stress and plasmid stability in plasmid DNA production by *Escherichia coli*. *Biotechnol Adv* 30:691–708.
- Sokol S, Millard P, Portais J-C. 2012. Influx\_S: Increasing numerical stability and precision for metabolic flux analysis in isotope labelling experiments. *Bioinformatics* 28:687–693.
- Sprenger GA. 2006. Aromatic amino acids. *Microbiol Monogr* 5:93–127.
- Wang Z, Xiang L, Shao J, Wegrzyn A, Wegrzyn G. 2006. Effects of the presence of ColE1 plasmid DNA in *Escherichia coli* on the host cell metabolism. *Microb Cell Fact* 5:34.
- Ward P, Small I, Smith J, Suter P, Dutkowski R. 2005. Oseltamivir (Tamiflu®) and its potential for use in the event of an influenza pandemic. *J Antimicrob Chemother* 55:5–21.
- Wu G, Yan Q, Jones JA, Tang YJ, Fong SS, Koffas MAG. 2016. Metabolic burden: Cornerstones in synthetic biology and metabolic engineering applications. *Trends Biotechnol* 34:652–664.
- Wu L, Mashego MR, van Dam JC, Proell AM, Vinke JL, Ras C, van Winden WA, van Gulik WM, Heijnen JJ. 2005. Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly <sup>13</sup>C-labeled cell extracts as internal standards. *Anal Biochem* 336:164–171.
- Xu Y-F, Amador-Noguez D, Reaves ML, Feng X-J, Rabinowitz JD. 2012. Ultrasensitive regulation of anapleurosis via allosteric activation of PEP carboxylase. *Nat Chem Biol* 8:562–568.

## Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web-site.



# Shikimic acid production in *Escherichia coli*: from classical metabolic engineering strategies to omics applied to improve its production

## OPEN ACCESS

Juan Andrés Martínez, Francisco Bolívar and Adelfo Escalante\*

### Edited by:

Hilal Taymaz Nikerel,  
Bogazici University, Turkey

### Reviewed by:

Maria Suarez Diez,  
Helmholtz Zentrum für  
Infektionsforschung, Germany  
Hannes Link,  
Max Planck Institute for Terrestrial  
Microbiology, Germany

### \*Correspondence:

Adelfo Escalante,  
Departamento de Ingeniería Celular y  
Biotecnología, Universidad Nacional  
Autónoma de México,  
Avenida Universidad 2001,  
Colona Chamilpa, Cuernavaca,  
Morelos 62210, Mexico  
adelfo@ibt.unam.mx

### Specialty section:

This article was submitted to  
Systems Biology, a section of the  
journal *Frontiers in Bioengineering  
and Biotechnology*

**Received:** 08 July 2015

**Accepted:** 07 September 2015

**Published:** 23 September 2015

### Citation:

Martínez JA, Bolívar F and  
Escalante A (2015) Shikimic acid  
production in *Escherichia coli*: from  
classical metabolic engineering  
strategies to omics applied to  
improve its production.  
*Front. Bioeng. Biotechnol.* 3:145.  
doi: 10.3389/fbioe.2015.00145

Departamento de Ingeniería Celular y Biotecnología, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Mexico

Shikimic acid (SA) is an intermediate of the SA pathway that is present in bacteria and plants. SA has gained great interest because it is a precursor in the synthesis of the drug oseltamivir phosphate (OSF), an efficient inhibitor of the neuraminidase enzyme of diverse seasonal influenza viruses, the avian influenza virus H5N1, and the human influenza virus H1N1. For the purposes of OSF production, SA is extracted from the pods of Chinese star anise plants (*Illicium* spp.), yielding up to 17% of SA (dry basis content). The high demand for OSF necessary to manage a major influenza outbreak is not adequately met by industrial production using SA from plants sources. As the SA pathway is present in the model bacteria *Escherichia coli*, several “intuitive” metabolically engineered strains have been applied for its successful overproduction by biotechnological processes, resulting in strains producing up to 71 g/L of SA, with high conversion yields of up to 0.42 (mol SA/mol Glc), in both batch and fed-batch cultures using complex fermentation broths, including glucose as a carbon source and yeast extract. Global transcriptomic analyses have been performed in SA-producing strains, resulting in the identification of possible key target genes for the design of a rational strain improvement strategy. Because possible target genes are involved in the transport, catabolism, and interconversion of different carbon sources and metabolic intermediates outside the central carbon metabolism and SA pathways, as genes involved in diverse cellular stress responses, the development of rational cellular strain improvement strategies based on omics data constitutes a challenging task to improve SA production in currently over-producing engineered strains. In this review, we discuss the main metabolic engineering strategies that have been applied for the development of efficient SA-producing strains, as the perspective of omics analysis has focused on further strain improvement for the production of this valuable aromatic intermediate.

**Keywords:** *Escherichia coli*, metabolic engineering, shikimic acid, transcriptome, metabolome, antiviral drug, influenza



## Introduction

Compounds derived from the aromatic amino acid (AA) pathway play important roles in the pharmaceutical and food industries as raw materials, additives, or final products (Patnaik et al., 1995; Bongaerts, 2001; Báez et al., 2001; Yi et al., 2002; Chandran et al., 2003; Báez-Viveros et al., 2004; Gosset, 2009). This metabolic pathway is present in bacteria and plants, starting with condensation of the central carbon metabolism (CCM) intermediates phosphoenolpyruvate (PEP) and erythrose-4-phosphate (E4P) to form the first AA pathway intermediate *D-arabino*heptulosonate-7-phosphate (DAHP). From this compound to chorismic acid (CHA), the pathway is mostly linear and represents the first part of the AA pathway, known as the common AA pathway or the shikimic acid (SA) pathway (**Figure 1**). One of the specific intermediates on this pathway is SA, which is a highly functionalized six-carbon cyclic compound with three asymmetric centers. Therefore, SA is an enantiomeric precursor for the production of many high valuable biological active compounds for different industries. SA is the precursor for the synthesis of compounds with diverse pharmaceutical applications, including as an antipyretic, antioxidant, anticoagulant, antithrombotic, anti-inflammatory, or analgesic agent, for the synthesis of anticancer drugs, such as (+)-zeylonone (which has been shown to inhibit nucleoside transport in Ehrlich carcinoma cells and to be cytotoxic to cultured cancer cells), and for antibacterial or hormonal applications [reviewed in Estevez and Estevez (2012), Liu et al. (2012), and Diaz Quiroz et al. (2014)] (**Figure 2**).

Specifically, SA has great pharmaceutical relevance because it is the precursor for the chemical synthesis of oseltamivir phosphate (OSF), known as Tamiflu<sup>®</sup>, used as the antiviral inhibitor of the neuraminidase enzyme for the treatment of diverse seasonal influenza viruses, including influenza A and B, the avian influenza virus H5N1, and the human influenza virus H1N1 (Krämer et al., 2003; Estevez and Estevez, 2012; Ghosh et al., 2012; Diaz Quiroz et al., 2014). For this purpose, SA is obtained from the seed of the Chinese star anise plant *Illicium verum*, which contains between 2 and 7% of the intermediate. However, it can only be retrieved from plants after 6 years of crop growth and harvested in September and October (Li et al., 2007; Raghavendra et al., 2009; Wang et al., 2011). To recover SA from the seed, a 10-step process is required, taking ~30 kg of seed to produce 1 kg of SA. According to Li et al. (2007) on their 2007 patent, ~90% of the Chinese harvest is used by Roche (2009) for OSF production.

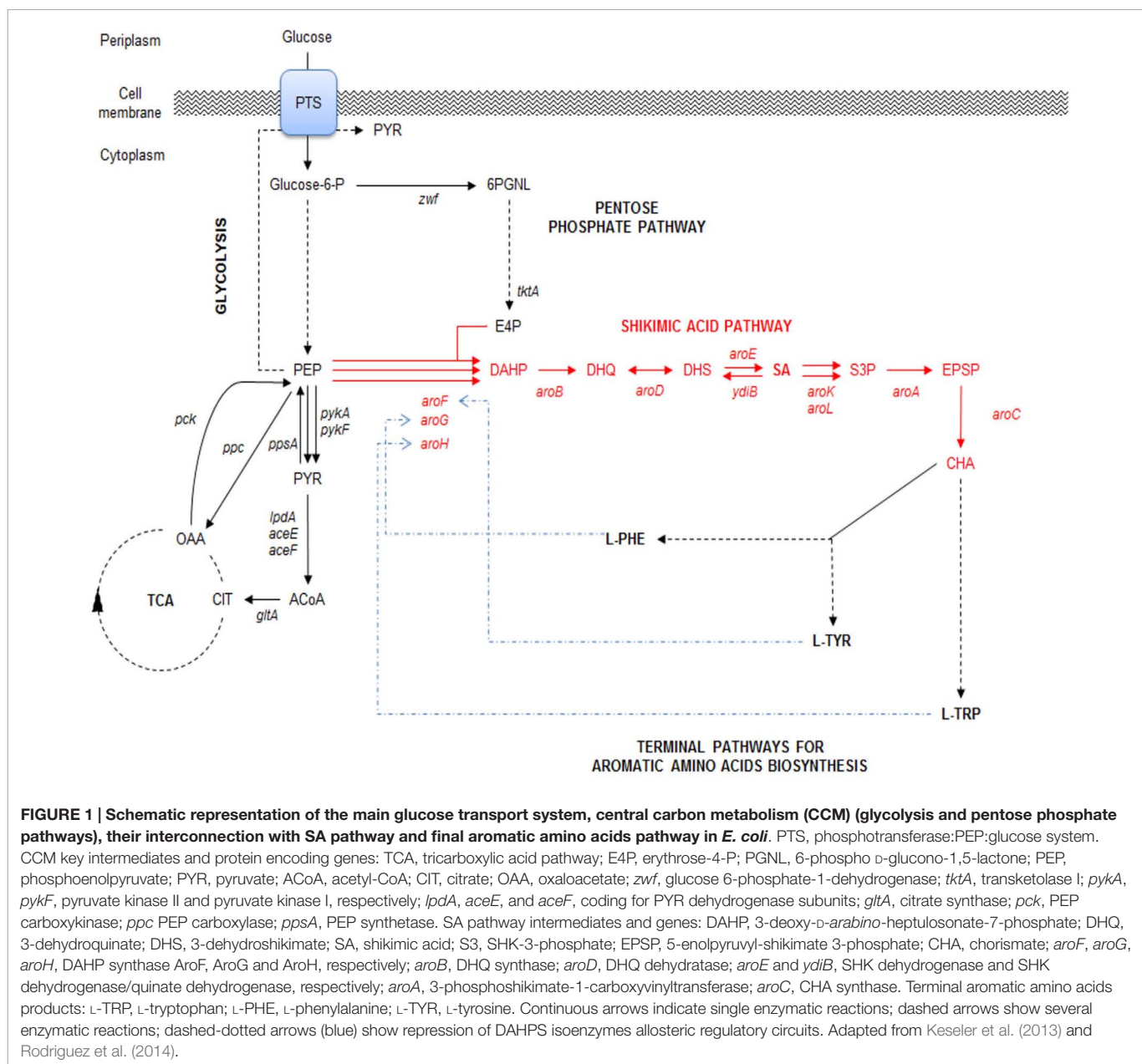
In 2009, Roche reported Tamiflu<sup>®</sup> sales to be 3.5 billion dollars, with a production capacity of up to 33 million treatments per month and 400 million packages per year (Scheiwiller and Hirschi, 2010). For the antiviral production, up to 1.3 g of SA are required to manufacture 10 doses to treat only one person, estimating a production requirement for this antiviral drug alone of ~520,000 kg/year (Rangachari et al., 2013). Even so, this reported production capacity could be insufficient in the case of an influenza pandemic, particularly with more pathogenic and infective strains. An estimated production of 30 billion doses, requiring 3.9 million kilograms of SA, would be necessary to cover a severe influenza outbreak (Rangachari et al., 2013).

According to the World Health Organization regarding influenza outbreak preparedness, only 66 million people in medium to low income countries are covered up with antiviral stocks, representing only 2.25% of the populations in these countries (World Health Organization, 2011). This situation results in a possibly low production capacity since in 2010, 100 million people were infected with common strains of influenza in Europe, Japan, and the United States alone. Moreover, before 2010, pandemic influenza has affected between 20 and 40% of the population, causing over 20 million deaths (Scheiwiller and Hirschi, 2010; World Health Organization, 2011).

For the reasons mentioned before and due to the relevance of SA in diverse industrial setups, many studies concerning SA production have been conducted within the past years, resulting in new and insightful strategies for its production, including recovery technologies, chemical synthesis methods, and biotechnological production methods using microorganisms. In fact, one of the most studied alternatives for SA production processes is biotechnological synthesis using recombinant microbial strains that are capable of producing high yields and that have high productivities, as there are key advantages over chemical synthesis, which include environmental friendliness, the availability and abundance of low-cost renewable feed stocks, and selectivity and diversity of the obtained products (Chen et al., 2013). These strains can be obtained by genetic modification, altering cellular properties to enhance their production capacity through the application of diverse metabolic engineering (ME) approaches (Krämer et al., 2003; Ghosh et al., 2012; Diaz Quiroz et al., 2014). However, despite the great achievements accomplished through this discipline, performance improvement has become limited after the first breakthroughs, mainly because of the traditional local pathway modification strategies. This is probably due to the limited understanding of the overall mechanism of metabolic regulation (Matsuoka and Shimizu, 2012). Therefore, given the importance of finding not only a particular pathway but also global information regarding cell physiology and metabolism to overcome production limitations, a systems biology approach supported by omics data may be the solution for improving SA production. The goal of this work is not only to review the literature on the great biotechnological achievements made for SA production, mainly in *Escherichia coli*, but also to outline future perspectives on research performed in the omics era, which could provide relevant tools for understanding cell behavior and production optimization via biotechnological processes.

## Classical Metabolic Engineering Approaches for SA Production

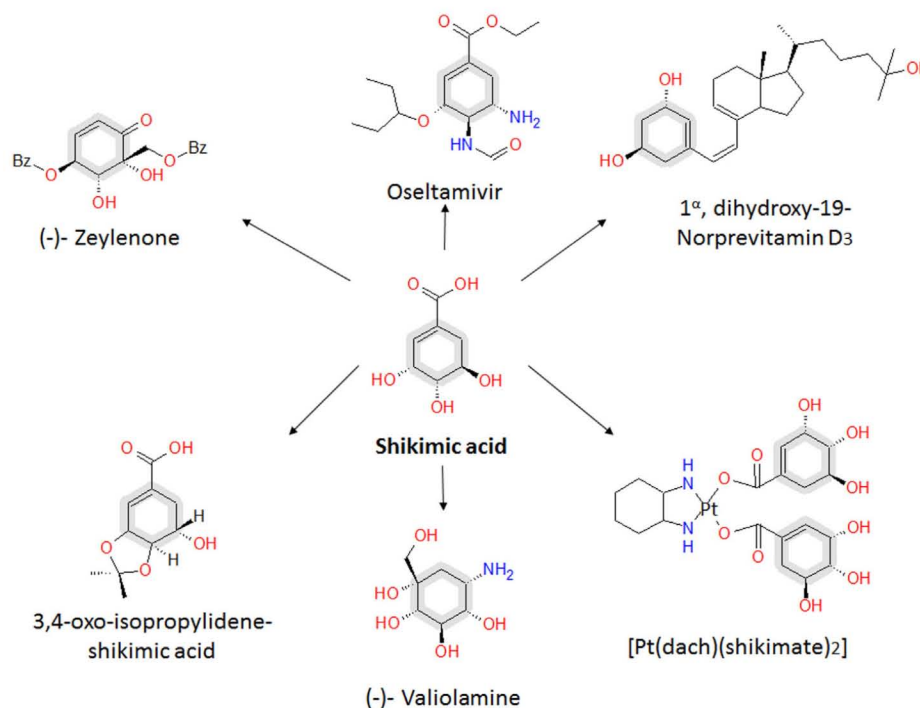
Metabolic engineering has been used since 1991 for strain modification by using recombinant DNA technology to enhance the production of specific metabolites (Matsuoka and Shimizu, 2012). The efforts to use ME have extended from the early years to optimize many cellular behaviors or parameters, such as substrate consumption, robustness, and tolerance toward toxic compounds and media conditions (Matsuoka and Shimizu, 2012). Classical ME strategies for strain development include various steps, such as the selection of a proper organism, elimination of competing pathways, deregulation of desired pathways at the enzyme activity



and transcriptional levels, and overexpression of enzymes at flux bottlenecks (Patnaik et al., 1995). Regarding the selection of an organism, *E. coli* has been preferred for industrial purposes and ME applications because of the knowledge available on *E. coli* physiology and the great numbers of tools developed to modify its genome (Chen et al., 2013). Therefore, many advances had been made regarding SA in *E. coli*, rendering strains capable of being used in industrial applications (Frost et al., 2002; Li et al., 2007).

In *E. coli*, the SA pathway starts by condensation of the CCM intermediates PEP and E4P by three DAHP synthase isoenzymes, AroG, AroF, and AroH (coded by *aroG*, *aroF*, and *aroH*, respectively), to produce DAHP. These three isoenzymes are responsible for the redirection from CCM intermediates toward the synthesis of aromatic compounds and are allosterically regulated specifically by the final products of AA biosynthesis. AroG catalyzes

~80% of DAHPS activity and is specifically feedback regulated by L-phenylalanine, AroF (~20% of DAHPS activity) is feedback regulated by L-tyrosine, and AroH (~1% of DAHPS activity) is regulated by L-tryptophan. Additionally, the transcription of *aroG* and *aroF* is controlled by the *tyrR* repressor, with the end products of the AA pathway (L-phenylalanine and L-tyrosine, respectively) acting as corepressors, whereas the transcription of *aroH* is controlled by the *trpP* repressor, with L-tryptophan acting as a corepressor (Keseler et al., 2013) (Figure 1). The ME solution for this first flux bottleneck is the expression of a DAHP AroG and AroF synthase that is not sensitive to feedback inhibition (*fbr*) (AroG<sup>fbr</sup> and AroF<sup>fbr</sup>). Mutations in the *aroG* and *aroF* genes lead to L-phenylalanine and L-tyrosine feedback-insensitive mutants with increased net carbon flux from CCM to the SA pathway (Keseler et al., 2013; Lin et al., 2014; Rodriguez et al., 2014); these



**FIGURE 2 | Relevant SA derivatives with high added value.** OSF, viral inhibitor of diverse influenza virus types, including seasonal types A and B, avian virus H5N1, and human virus H1N1. (–)-Zeylenone is a compound with antiviral, anticancer, and antibiotic activities. (–)-Valiolamine, a very strong  $\alpha$ -glucosidase with inhibitory activity against porcine intestinal enzymes sucrose, maltase, and isomaltase. [Pt(dach)(SA)<sub>2</sub>] is an active compound against L1210 leukemia. 3,4-Oxo-isopropylidene-SA, with antithrombotic activity and anti-inflammatory effects. Analogs of 1 $\alpha$ , dihydroxy-19-Nor previtamin D3 is a compound with promising applications in the treatment of osteoporosis and malignancies. Adapted from Estevez and Estevez (2012) and Diaz Quiroz et al. (2014).

mutants have been used in most SA production strains (Table 1) (Chandran et al., 2003; Escalante et al., 2010; Chen et al., 2012, 2014; Rodriguez et al., 2013). Forward reactions convert DAHP to dehydroquinic acid (DHQ), then to 3-dehydroquinic acid (DHS) and finally to SA by the enzymes 3-dehydroquinic acid synthase (*aroB*), 3-dehydroquinic acid dehydratase (*aroD*), and shikimate dehydrogenase (*aroE*), respectively (Figure 1). Although the pathway to SA conversion is small and linear, its regulation and the competition for precursor metabolites remain quite complicated because the SA pathway is dependent on the glycolytic and pentose phosphate pathways (PPPs) to provide the starting precursors PEP and E4P, respectively (Gosset, 2009; Escalante et al., 2012; Ghosh et al., 2012; Rodriguez et al., 2014).

SA production starts in CCM, further away from glucose consumption. In *E. coli*, the majority of glucose transport occurs via the PEP:glucose phosphotransferase system (PTS), which uses a phosphate group from one molecule of PEP to simultaneously import and phosphorylate periplasmic glucose, resulting into 6-phosphate glucose (G6P) and pyruvate (PYR) (Figure 1). For these reasons, the application of ME strategies only on the SA pathway would not render a significantly optimized strain for SA production (Ghosh et al., 2012).

To optimize productivity and yields from a given carbon source, modification of the CCM pathways supplying the needed precursors and energy sources for product synthesis is required (Patnaik and Liao, 1994). For the E4P supply, the PPP is the responsible for its production. The overexpression of transketolase I (TktA,

coded by *tktA*) and transaldolase (coded by *talA*), resulting in the preferential use of TktA to improve the E4P pool for the synthesis of DAHP (Flores et al., 1996; Draths et al., 1999; Frost et al., 2002; Chandran et al., 2003; Escalante et al., 2010; Rodriguez et al., 2013).

Regarding increasing the PEP pool, the first problem arises with the consumption of 50% of the PEP resulting from the catabolism of one molecule of glucose-6-P by PTS during the translocation and phosphorylation of one molecule of glucose. A rational approach is to reconvert PYR to PEP by overexpressing PEP synthase (coded by *ppe*); this solution, along with the expression of a DAHP<sup>fbt</sup> (AroG<sup>fbt</sup> or AroF<sup>fbt</sup>), leads to a 51% (mol/mol) yield of DHS and related SA pathway metabolites. This yield is in fact higher than the 43% (mol/mol) yield calculated from stoichiometric reactions, reflecting the effective redistribution of the PEP to PYR pool ratio and the ability of the strain to redirect this new imbalance into the SA pathway (Yi et al., 2002; Chandran et al., 2003; Krämer et al., 2003; Escalante et al., 2010; Rodriguez et al., 2013). Overexpression of the *ppe* gene has been studied; the maximum yield of SA is not obtained under the maximum concentration of the enzyme. In fact, it has been found that expression of this enzyme over the optimized level would only reduce the yields of SA intermediates, probably due to energetic imbalances (Yi et al., 2002).

The maximum theoretical yield limitation can be changed by restructuring the metabolic network, providing the system with a new stoichiometric matrix. Therefore, a natural solution for the

**TABLE 1 | Relevant *E. coli* engineered strains for SA production.**

Producing strain	Phenotypic traits	Special comments	Culture conditions	Title (g/L)	Yield (mol SA/mol glc)	Reference
SA112	BW25113 $\Delta$ aroKL, P <sub>pps</sub> ::P <sub>tac</sub> Q1, P <sub>csrB</sub> ::P <sub>tac</sub> Q1Pt5-pps, P <sub>T5</sub> -csrB, 5P <sub>tac</sub> -tktA	CiChE evolved to optimize SA production	Shake flasks cultures with 10 g/L glc, 1 g/L peptone	1.70	0.25	Cui et al. (2014)
DHPYAAS-T7	DH5 $\alpha$ $\Delta$ ptsHlcr, $\Delta$ aroKL, $\Delta$ ydiB pAOC-TGEFB:aroE, aroB, glk, tktA, aroF <sup>tr</sup>	Plasmid overexpression of SA related genes	Shake flasks cultures (50 mL), M9 broth supplemented with 25 g/L glycerol, 10 g/L peptone, 15 g/L YE	1.066	0.23	Chen et al. (2012)
PB12.SA22	JM101 $\Delta$ ptsH, ptsI, crr::Km <sup>r</sup> $\Delta$ aroKL::cm <sup>r</sup> pJLBaroG <sup>tr</sup> tktA, pTOPO-aroBaroE	Laboratory evolved PTS <sup>-</sup> glucose <sup>+</sup> into glucose <sup>+</sup> derivative phenotype	1 L batch bioreactor 25 g/L glc and 15 g/L YE	7.05	0.22	Escalante et al. (2010)
SA5	B0013 $\Delta$ aroKL::dif $\Delta$ ptsG::dif $\Delta$ ydiB::dif $\Delta$ ackA-pta::dif pTH-aroG <sup>tr</sup> -ppsA-tktA	Plasmid over expression of SA related genes	7 L fed-batch bioreactor, initial 15 g/L glc supplemented with AA and vitamins	14.6	0.3	Chen et al. (2014)
SA114	BW25113 $\Delta$ aroKL, P <sub>pps</sub> ::P <sub>tac</sub> Q1, P <sub>csrB</sub> ::P <sub>tac</sub> Q1 P <sub>T5</sub> -pps, P <sub>T5</sub> -csrB, 5P <sub>tac</sub> -tktA, 5P <sub>tac</sub> -pntAB	CiChE evolved to optimize SA production	Shake flasks, 10 g/L glc, 1 g/L peptone	2.99	0.31	Cui et al. (2014)
SP1.1pts-/pSC6.090B	DH5 $\alpha$ $\Delta$ ptsH-ptsI-crr $\Delta$ serA::aroB $\Delta$ aroL::Tn10 $\Delta$ aroK::Cm <sup>r</sup> P <sub>tac</sub> Glf glk, aroF <sup>tr</sup> tktA, P <sup>tac</sup> aroE, serA	Heterologous glk and glf from <i>Z. mobilis</i> to restore glc transport and phosphorylation in PTS <sup>-</sup> glc <sup>-</sup> phenotype	10 L fed-batch bioreactor 55–170 mM Glc + 15 g/L YE	84	0.33	Chandran et al. (2003)
SA116	BW25113 $\Delta$ aroKL, P <sub>pps</sub> ::P <sub>tac</sub> Q1, P <sub>csrB</sub> ::P <sub>tac</sub> Q1 P <sub>T5</sub> -pps, P <sub>T5</sub> -csrB, 5P <sub>tac</sub> -tktA, 5P <sub>tac</sub> -nadK	CiChE evolved to optimize SA production	Shake flasks 10 g/L glc, 1 g/L peptone	3.12	0.33	Cui et al. (2014)
AR36	JM101 $\Delta$ ptsH, ptsI, crr::Km <sup>r</sup> $\Delta$ aroKL::cm <sup>r</sup> $\Delta$ pykF $\Delta$ lacI pTrcAro6-aroB, tktA, aroG <sup>tr</sup> , aroE, aroD zwf	Constitutive strong over expression by synthetic operon on plasmid	1 L batch bioreactor 100 g/L glc + 15 g/L YE	41.8	0.42	Rodriguez et al. (2013)

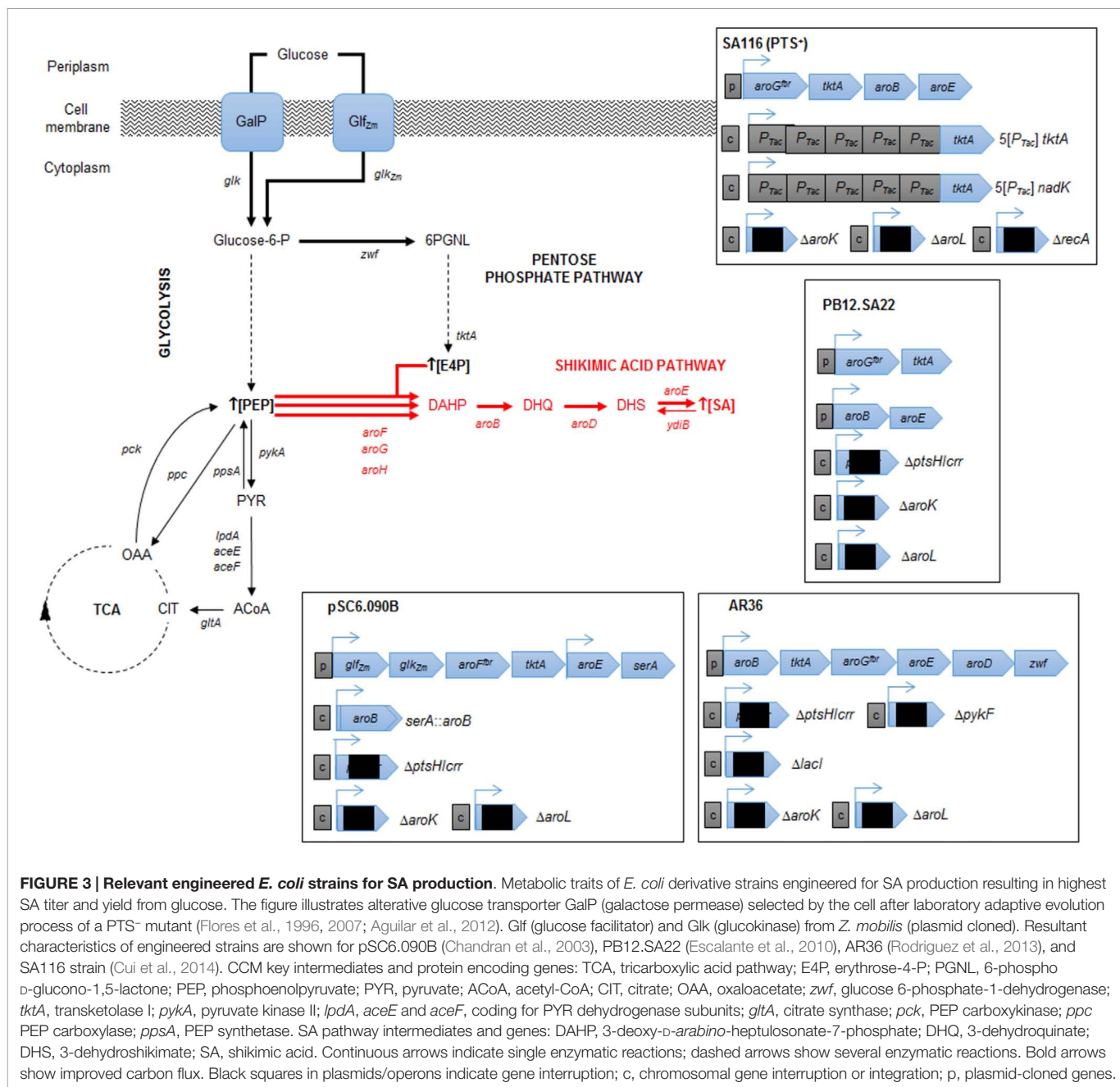
glc, glucose; YE, yeast extract.

PEP pool was to eliminate the PTS system, which would not only modify the amount of PEP but also redistribute the stoichiometric matrix to raise the maximum theoretical yield to 86% (mol/mol) (Chandran et al., 2003; Krämer et al., 2003). The main problem with this solution is the resultant low level of glucose transport, which results in a strain with hampered growth (PTS<sup>-</sup> phenotype) (Flores et al., 1996, 2007; Aguilar et al., 2012). Nevertheless, various strategies have been developed to revert this low glucose consumption and low growth phenotype. Using rational ME strategies, substitution of the PTS for another glucose transport system has been performed. Chandran et al. used heterologous expression of the *Zymomonas mobilis* (Glf) glucose transporter, a glf-encoded glucose facilitator and a glk-encoded glucose kinase (Glc), thereby allowing cells to consume glucose more efficiently without consuming PEP (Frost et al., 2002; Chandran et al., 2003; Krämer et al., 2003). Another strategy is to apply laboratory adaptive evolution onto a PTS<sup>-</sup> strain. Flores et al. used a continuous culture with glucose as a single carbon source to select high glucose consumption-evolved derivative strains (PTS<sup>-</sup> glc<sup>+</sup> phenotype). Characterization of these mutants revealed overexpression of the galP and glk genes encoding galactose permease and glucokinase, respectively, allowing and improving glucose transport and phosphorylation capabilities and resulting in an increased specific growth rate and PEP availability (Flores et al., 1996, 2007; Aguilar et al., 2012).

Finally, with precursors known to induce redirection, deregulation, and overexpression of the SA pathway genes aroB, aroD, and aroE have been achieved, resulting in an efficient carbon flux

from CCM to the SA pathway. The highest production to date corresponds to the SP1.1pts-/pSC6.090B strain, a PTS<sup>-</sup> derivative strain with a plasmid containing two tac promoters, the first of which controls expression of the glf, glk, aroF<sup>tr</sup>, and tktA genes and the second of which controls expression of the aroE and serA genes (Chandran et al., 2003). The reasoning behind this construction was to increase the PEP pool by deleting PTS, to recuperate glucose consumption by overexpressing glf and glk, to assure E4P pool enhancement by overexpressing tktA and to induce a deregulated pull toward the AA pathway via an aroF<sup>tr</sup>, as discussed before. The second promoter in the plasmid was designed to overexpress aroE, allowing continuous flux of the SA pathway; additionally, a second copy of aroB was introduced into the chromosome instead of reintroducing the serine production-related gene serA to the cell via the plasmid and was used as a selection marker for plasmid retention. This approximation, along with the deletion of genes related to SA consumption (aroK and aroL), allowed SA accumulation, achieving a production capacity of 87 g/L SA, with a yield of 36% (mol/mol) and a productivity of ~5.3 g/L h when a 10 L glucose-fed batch was cultured. This strain has the highest titer accumulation recorded in the literature to date (Table 1; Figure 3).

Even with these rational strategies, the yield of the SP1.1pts-/pSC6.090B strain is far from the theoretical maximum yields of PTS<sup>-</sup> derivative strains. In 2010, Escalante et al. presented a JM101 PTS<sup>-</sup> derivative strain with high glucose consumption capacity that was capable of overexpressing the galP and glk genes and that was produced from an adaptive evolution process. This



strain (PB12), along with a two-plasmid expression system for *aroG<sup>br</sup>-tktA* and *aroB-aroE*, respectively, under *lacUV5* promoters inducible by IPTG (PB12.SA22), allowed a yield of 29% (mol/mol) (Escalante et al., 2010). Further modifications allowed them to find that a *pykF* deletion could result in higher yields of total aromatic compounds, up to 50% (mol/mol), even when presenting an SA yield diminution (0.21%). In this case, the amount of flux reduced from PEP to PYR was redirected throughout the SA pathway, and without the correct amounts of enzymes, new bottlenecks appeared, causing other metabolites and intermediates to accumulate (Escalante et al., 2010). Therefore, it was clear that regulating gene expression and dosage remained a problem for more efficiently redirecting flux not only toward but also within

the SA pathway. Regarding that topic, Rodriguez et al. utilized the PB12 *pykF-aroKL<sup>-</sup>* strain and developed a plasmid with a constitutively strong promoter onto a synthetic operon containing the *aroB*, *tktA*, *aroG<sup>br</sup>*, *aroE*, *aroD*, and *zwf* genes (AR36) for synchronous expression of the relevant genes found in previous research. With this expression design, the AR36 derivative strain is able to redirect the carbon flow to SA even in high glucose conditions (above 100 g/L of the initial substrate concentration) without producing high acetate titers. This strain produced up to 43 g/L of SA via simple batch processes, with SA yields of 42% (mol/mol) and total SA pathway intermediate yield up to 67% of the theoretical maximum, representing the highest yield managed to be produced to date (Rodriguez et al., 2013) (Table 1; Figure 3).

Regarding the expression and regulation of key SA production genes, most of the research has been performed using plasmid expression; however, there are multiple drawbacks, ranging from structural and segregational instability to metabolic burden, of plasmid replication. Cui et al. (2014) resolved this problem by constructing a strain with an *aroG<sup>br</sup>*, *aroB*, *aroE*, and *tktA* gene cluster integrated into the chromosome and by tuning the copy number and expression by using chemically induced chromosomal evolution (CICHe) with triclosan. They also overexpressed the *ppsA* and *csrB* genes to enhance the PEP pyruvate pool. This strain rendered a 1.70 g/L SA titer, with a yield up to 0.25 (mol/mol). Finally, they studied and improved cofactor availability for SA production optimization; in this case, NADPH availability was increased because *aroE*-encoded enzymes require this specific cofactor for the DHS to SA conversion reaction. By plasmid-based or chromosomal overexpression of the NADPH availability-related genes *pntAB* or *nadK*, this cofactor pool was enhanced, which was directly correlated to the SA production capabilities of the strain. As they changed the promoters and the expression of all the chromosomally inserted genes related to SA production mentioned above, they managed to construct a strain capable of producing a yield of 0.33 (mol/mol) SA from glucose (Figure 3).

Many other examples of SA production platforms in *E. coli* have been studied in the literature, the most relevant of which are referred to in Table 1, rendering many industrially competitive strains and processes. Nevertheless, the main efforts throughout the past two decades were directed toward a particular pathway approach. As shown in Table 1, few SA production processes have been designed utilizing an overview of global regulation and manipulation, which can be obtained from omics data. Transforming this global information into global knowledge on the complexity of cell regulation would reveal the existing regulatory bottlenecks, allowing us to metabolically engineer potential strains using a systems biology approach, finally ensuring a truly rational strain design with optimized production capabilities.

## Omics Approaches for the Study of the SA Pathway in *Escherichia coli*

Classical ME approaches applied to diverse *E. coli* strains to obtain SA-overproducing derivatives have targeted key genes in the CCM and SA pathways, allowing successful reconfiguration of the biochemical network of engineered strains and resulting in the efficient redirection of carbon flow from CCM to SA production. However, the inactivation of key genes coding for enzymes involved in global regulatory processes, such as the PTS system or coding for key node enzymes, such as the PykF enzyme results in global metabolic reconfiguration, which frequently introduces significant flux imbalances. This often produces undesirable outcomes, including the accumulation of intermediates, feedback inhibition of upstream enzymes, the formation of unwanted byproducts, and the diminution of cellular fitness via the rerouting of resources toward the unnecessary or non-essential production of pathway enzymes. By understanding these newly created flux imbalances in SA-overproducing derivative strains,

it is possible to boost the overall cellular physiology, product titer, productivity, and yield, taking into account a global view of cellular metabolism (Biggs et al., 2014). Combinatorial approaches allow researchers to work with this scenario by conducting global cellular searches, but the necessity for high-throughput screening is often a drawback for pathway engineering. The other approach is to augment knowledge and computational tools to properly predict designs to achieve a desired metabolic outcome (Fong, 2014). Several high-throughput approximations, such as genomic, transcriptomic, and proteomic predictions, have been applied to aromatic AAs and engineered SA-overproducing strains for the identification of non-intuitive targets other than those genes/enzymes involved in the CCM and SA pathways that might be suitable for further modification by ME.

## The Identification of YdiB (*ydiB*) as a Key Enzyme in Byproduct Formation During SA Synthesis

The analysis of available genome sequences using Hidden Markov Model profiles to identify all known enzymes of the SA pathway has shown that some genes have been lost in diverse microbial groups, particularly in host-associated bacteria (Zucko et al., 2010). This condition has been proposed to result in the development of undesirable metabolic traits, such as the hydroaromatic equilibration observed in *E. coli*, resulting in the synthesis of so-called missing metabolites, such as quinic acid (QA) and DHQ, by a reversion of the SA biosynthetic pathway (Knop et al., 2001; Zucko et al., 2010). The coproduction of high quantities of the byproducts DHS and QA is not a desirable trait; they significantly reduce the SA yield because QA is co-purified during the downstream process of SA purification from the culture supernatant (Knop et al., 2001; Krämer et al., 2003; Diaz Quiroz et al., 2014).

The strain W3110.shik1 ( $\Delta$ *aroL*, *aroG<sup>br</sup>*, *trpE<sup>br</sup>*, and *tnaA*) engineered for SA production growing in low glucose (high phosphate) or glucose-rich (low phosphate) conditions resulted in the production of SA in cultures with mineral broth, as the single inactivation of shikimate kinase II (*aroL*) allows carbon flux to CHA through shikimate kinase I (*aroK*), resulting in the synthesis of aromatic AAs. However, under carbon-limited conditions, SA production decreased by 59%, and the byproducts DHS, DHQ, gallic acid (GA), and QA were detected in the culture supernatant with respect to phosphate limiting culture conditions (Johansson et al., 2005). Global transcriptomic analysis (GTA) of the strain W3110.shik1 in chemostatic culture conditions, comparing between glucose and phosphate limiting conditions, allowed identification of the significantly upregulated genes *ydiB* (coding for shikimate dehydrogenase/quininate dehydrogenase), *aroD*, and *ydiN*, which encodes a putative transporter, in carbon limiting conditions. The upregulation of these genes, particularly *ydiB* (10× with respect to its paralogs, *aroE*), was proposed to increase the YdiB level, which uses DHQ and SA as substrates, as this enzyme has a lower  $K_m$  for SA in the presence of NAD<sup>+</sup> (Keseler et al., 2013). Additionally, the intracellular concentration of NAD<sup>+</sup> is reported to be 40-fold higher than that of NADH<sup>+</sup>, suggesting that the dehydrogenase activity on SA to produce DHS is favored by YdiB *in vivo* (Johansson and

Lidén, 2006). These results suggest that byproduct formation during SA production was associated with the reversal of the biosynthetic pathway from (1) SA + NAD(P)<sup>+</sup> ↔ DHS + NAD(P)H + H<sup>+</sup> and (2) DHS + NAD(P)H + H<sup>+</sup> ↔ QA + NAD(P)<sup>+</sup> by YdiB or (3) DHS + H<sub>2</sub>O ↔ DHQ by AroD (Figure 4). The presence of a large amount of intracellular SA was proposed to drive the reversal of the pathway, whereas YdiN was proposed to be the exporter of the aromatic byproducts (Johansson and Lidén, 2006).

As these results suggest an important role of YdiB in byproduct synthesis during SA production and its intracellular accumulation under glucose limiting conditions, a rational strategy to avoid byproduct synthesis was the inactivation of *ydiB* and/or the upregulation of its paralogs, *aroE*, coupled to efficient SA secretion from the cell. The upregulation of *aroE* expression (simultaneously with other key genes of the CCM and SA pathways) in PTS<sup>-</sup> *gluc*<sup>+</sup>*aroK*<sup>-</sup>*aroL*<sup>-</sup> engineered strains resulted in the highest SA titer and yield reported with low byproduct formation (Chandran et al., 2003; Rodriguez et al., 2013) (Table 1).

The replacement of *ydiB* by its paralogs, *aroE*, in a modular biosynthetic pathway design for L-tyrosine production in *E. coli* MG1655 resulted in the elimination of a bottleneck caused by the high affinity of YdiB protein for the accumulation of QA and DHS. This replacement in the modular plasmid construction P<sub>lac-UV5</sub>*aroE*, *aroD*, *aroB*<sup>op</sup>, *aroG*<sup>ibr</sup>, *ppsA*, *tktA* (*op* = optimize codon usage) resulted in the accumulation of 700 mg/L of SA, which was in turn successfully channeled to L-tyrosine (Juminaga et al., 2012). However, combinational plasmid overexpression of the *aroB*, *aroD*, *aroE*, *ydiB*, *aroK*, *aroL*, *aroA*, *aroC*, and *tyrB* genes with *ydiB* resulted in high L-tyrosine production. This result suggested that *ydiB* but not its paralog, *aroE*, is an attractive target for the overproduction of this aromatic AA because *aroE* in *E. coli* codes for a feedback-inhibited shikimate dehydrogenase, resulting

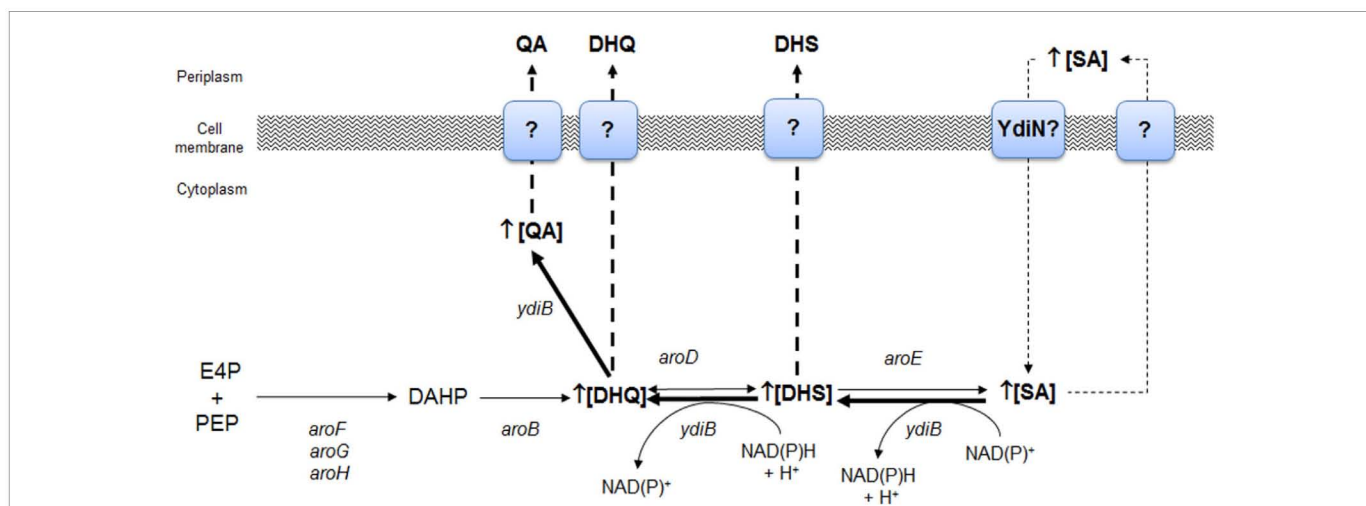
in a bottleneck for L-tyrosine production (Lütke-Eversloh and Stephanopoulos, 2008).

### The Impact of *pykF* Inactivation on the Protein Levels of SA Pathway Enzymes

The pyruvate kinase isoenzymes Pyk I and Pyk II (coded by *pykF* and *pykA*, respectively) play key roles in CCM via Pyk activity, together with 6-phospho-fructokinase I (coded by *pkfA*) and glucokinase (*glk*), controlling carbon flux through the glycolytic pathway (Keseler et al., 2013). Pyk I and Pyk II are key allosteric enzymes that catalyze one of the two substrate-level phosphorylation steps yielding ATP and the irreversible trans-phosphorylation of PEP and ADP into PYR and ATP, maintaining a permanent flux of PYR to acetyl-CoA (Keseler et al., 2013).

Inactivation of the *pykF* gene in *E. coli* PTS<sup>-</sup> derivatives (PB12 strain) engineered for SA production has resulted in the increased flux of carbon into the SA pathway (Escalante et al., 2010), increasing the DAHP concentration above 370% (and the total SA pathway aromatic yield) with respect to the *pykF*<sup>+</sup> parental strain. Further applications of ME strategies in the PB12 strain *pykF*<sup>-</sup> resulted in the derivative strain AR36, which produces up to 40 g/L SA with a yield of 0.42 mol SA/mol glc (Table 1) (Rodriguez et al., 2013), demonstrating that the inactivation of *pykF* in a PTS<sup>-</sup> derivative strain significantly improves PEP flux toward SA synthesis.

Global proteomic analysis in a *pykF*<sup>-</sup> derivative of *E. coli* (BW25113) compared with its *pykF*<sup>+</sup> parental strain revealed the differential overexpression of 24 proteins, including enzymes from the SA pathway and aromatic AAs. The upregulation of key SA pathway enzymes, including the DAHPS AroG isoenzyme (2.66 times more abundant with respect to the *pykF*<sup>+</sup> strain), which is involved in the synthesis of DAHP, the first intermediate of the SA pathway, and the AroB enzyme (DHQ synthase, 4.72



**FIGURE 4 | Identification of key genes of the SA pathway involved in the biosynthesis of aromatic byproducts QA and DHS from SA as determined by global transcriptomic analysis in *E. coli* W3110.shik1.** Overexpression of *ydiB*, *aroD*, and *ydiN* genes allowed proposing that under carbon limiting growth conditions, SA is intracellularly accumulated as consequence of an inefficient export to periplasmic space or as consequence of its back transport to the cytoplasm as consequence of extracellular accumulation. YdiN, a putative transporter coded by *ydiN* was proposed to be involved in SA back import. Backflow of SA to DHS was possibly catalyzed by YdiB, whereas synthesis of DHQ from DHS was performed by AroD enzyme and finally, YdiB performed synthesis of QA from DHQ. Adapted from Johansson and Lidén (2006).

times more abundant with respect to the *pykF*<sup>+</sup> strain) (Kedar et al., 2007). These results support the positive impact of *pykF*<sup>-</sup> inactivation not only on increased PEP availability but also on increased carbon flux toward the SA pathway.

### The Identification of Other Possible Key Catabolic and Biosynthetic Genes Involved in SA Production

Batch fermentation cultures of the *E. coli* PB12.SA22-derivative strain for SA production (PTS<sup>-</sup> Glc<sup>+</sup> *aroK*<sup>-</sup>, *aroL*-*aroG*<sup>tr</sup>, *tktA*, *aroB*, and *aroD*; **Table 1**) using complex production media containing 25 g/L glucose and 15 g/L yeast extract (YE) showed two characteristic growth stages: a fast growth phase associated with low glucose consumption during the first 8–10 h of cultivation and low SA production, and a second slow growth stage with high glucose consumption until this carbon source was completely consumed (25 h of cultivation). Interestingly, SA production continues during the STA phase after glucose, used as a carbon source, was completely consumed, until the end of fermentation (50 h) (Escalante et al., 2010). This behavior suggested that during the EXP growth phase, this strain preferentially consumed some YE components to support growth, whereas glucose was used to produce SA and other pathway intermediates, suggesting the existence of regulatory and physiological differences between EXP and STA phases (Cortés-Tolalpa et al., 2014).

GTA was performed to corroborate this hypothesis during SA production in batch fermentation cultures using complex fermentation broth (Chandran et al., 2003; Escalante et al., 2010; Rodriguez et al., 2013) by comparing global expression profiling between the mid-exponential growth phase (EXP, 5 h of cultivation), the early stationary phase (STA1, 9 h) and the late STA phase (44 h); EXP/STA1, EXP/STA2, and STA1/STA2 comparisons were conducted (Cortés-Tolalpa et al., 2014) (**Figure 5**).

The relevant results showed EXP growth in the derivative strain PB12.SA22 during the first 9 h of cultivation. When the L-tryptophan provided by YE available in the supernatant was completely consumed (6 h), the strain entered the low-growth phase (even in the presence of glucose) until 26 h of cultivation, when glucose was completely consumed; this was associated with low SA production. Interestingly, during the stationary stage, SA production continued until the end of fermentation (50 h), achieving the highest accumulation (7.63 g/L of SA) in the absence of glucose (**Figures 5A,B**, upper panel).

GTA comparisons among EXP/STA1, EXP/STA2, and STA1/STA2 showed no significant differences in the regulation of genes from the CCM and SA pathways, but for the EXP/STA1 comparison, the upregulation of genes coding for sugar transport, AA catabolism and biosynthesis, and nucleotide/nucleoside salvage was observed (**Figure 5A**). Interestingly, in the STA2 phase, the highest SA production was observed in the absence of glucose in supernatant, associated with the upregulation of genes encoding transporters for the AAs L-lysine, L-arginine, L-histidine, L-ornithine, and L-glutamic acid and enzymes involved in the synthesis, interconversion, and catabolism of L-arginine. As all of these AAs are provided by YE, this result suggests that this AA could play a key role in fueling carbon to SA synthesis, and likely also in L-arginine conversion to the

TCA intermediate succinate through the super-pathway of L-arginine and L-ornithine degradation (Keseler et al., 2013) (**Figure 5B**). These results indicate the origin of carbon required for the highest SA production during the STA phase after glucose was completely consumed. Additionally, the upregulation of genes involved in the pH stress response and inner and outer membrane modifications suggests a cellular response to environmental conditions imposed on the cell at the end of fermentation (44 h) (Cortés-Tolalpa et al., 2014).

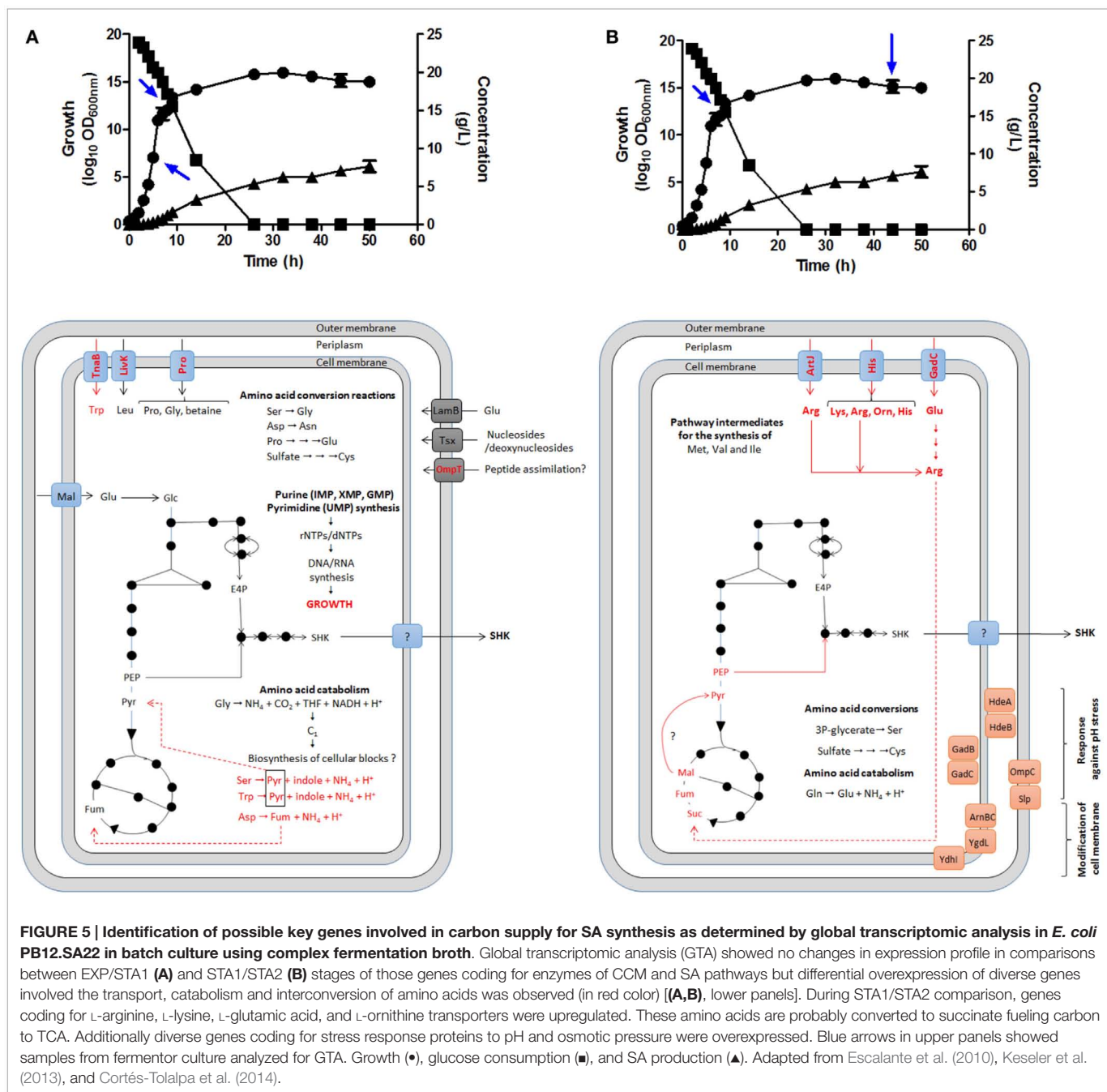
The upregulation of genes coding for the biosynthesis and interconversion pathways of almost all AAs was also observed by GTA in cultures under C-limiting condition of the derivative strain W3110.shik1 grown in minimal broth. These changes were postulated to correlate to aromatic AA starvation with these culture conditions, although this strain maintained functional shikimate kinase I (*aroK*), allowing the accumulation of SA but maintaining carbon flux toward CHA and aromatic AAs (Johansson and Lidén, 2006).

As demonstrated by GTA in the SA-producing strain PB12.SA22 during batch culture fermentations in complex media containing YE, several metabolic constraints limit the growth capabilities of this strain, stopping growing even in the presence of glucose. The highest SA production observed in the late stationary stage in the absence of glucose was probably supported by the non-aromatic AA content of YE. This evidence supports valuable information to further optimize culture strategies, as YE feeding increased the SA titer and yield in engineered strains.

### Omics Data Integration into Metabolic Modeling: Moving Toward Data Integration for Rational Strain Improvement

Although ME is capable of reconfiguring a biochemical network to redirect the substrate conversion into valuable compounds by manipulating the microorganism genetic code, its classical rational approach often introduces significant new flux imbalances. This has often caused undesirable outcomes due to the accumulation of intermediates, feedback inhibition of upstream enzymes, and the formation of unwanted byproducts of cellular fitness diminution via the rerouting of resources toward the unnecessary or non-essential production of pathway enzymes (Biggs et al., 2014). By understanding these newly created flux imbalances on mutant strains, it is possible to boost overall cellular health and the product titer, productivity, and yield, taking into account a holistic view of cellular metabolism (Biggs et al., 2014). Since the development of the omics, there has been an increased interest to understand the behavior of complete biological systems. Omics renders biological data from all levels of metabolism going all the way from genome to metabolome, these data combined give us the possibility to study the whole organism instead of single components. To achieve this, mathematical models play the important role of converting omics data into organismal information and knowledge (Åkesson et al., 2004; Fong, 2014). There are several frameworks and approaches for the mathematical modeling of metabolism developed to collect high-throughput data to understand as well as to predict phenotypic





function. Computational applications have been developed using models as quantitative mathematical representations of biological systems and or their components to a suitable level of simplification (Jouhten, 2012). These computational tools can be used to identify new biological pathways in the host microorganism for the selection and improvement of important genotypic characteristics to improve the production of the desired compound (Long et al., 2015). In this section, we discuss some mathematical models and computational tools that can be used in ME to utilize all high-throughput omics data and render new insights into flux distributions, regulation constraints, and modification targets to optimize the production of desired metabolites.

To understand the challenges and virtues of mathematical modeling, we must observe that biological systems are complex in nature, involving the transport of information through many layers, including the genome, transcriptome, proteome, and metabolome; therefore, regulatory steps between the interactions of these layers finally render the complex outcome of the phenotypic behavior (Cloots and Marchal, 2011; Fong, 2014). Therefore, mathematical models have been used to evolve and clarify the complex network interactions and system characteristics to reveal the underlying mechanisms. Despite this high degree of complexity, with all the recent advances and data sets available, mathematical modeling promises to generate experimentally

testable hypotheses, predictions, and new insights into systems biology to better understand cell behavior (Stelling, 2004).

The first step in mathematical modeling is reconstructing the metabolic network. With the advent of the genomic era since approximately 1999, reconstruction can be achieved on a genome-wide scale for many organisms and has been used to expand the knowledge on metabolic networks and to identify new or non-intuitive metabolic reactions to be engineered for further strain improvements (Åkesson et al., 2004; Kim et al., 2012). Genome-scale models are assembled and manually curated from the annotated genome, and biochemical information is used to render a representation of the metabolic network on which mathematical representations will set a matrix of equations to model its behavior. The reconstruction of a genomic metabolic network starts through the examination and identification of the coding regions or open reading frames on the sequence. After analysis with established algorithms and biochemical and physiological databases (EcoCyc, MPW, and KEGG WIT), sequences can be converted into feasible reactions, and a metabolic network can be reconstructed from genomic information (Covert et al., 2001). This reconstructed network, based on genomic data, is now the backbone of an *in silico* organism. Many organisms have been completely sequenced and have simultaneously been extensively biochemically studied, which in turn can make the reconstructed metabolic network more complete (Covert et al., 2001). In recent years, ~40% of all eukaryotic models and 30% of the total prokaryotic models have been published, advancing from highly characterized organisms (*E. coli* and *Saccharomyces cerevisiae*) to less characterized species with more complex biological systems that have special characteristics for specific applications (Kim et al., 2012). When a network is described with sufficient detail, some qualitative predictions can be made, and with the inclusion of stoichiometric, thermodynamic, and kinetic data, the reconstructed metabolic map of an organism can be used to generate quantitative predictions regarding phenotype via the construction of mathematical models (Covert et al., 2001). For example, individual genes have been deleted from *in silico* models, and correlations between the model and experimental data for the consequences of each deletion have been found to be 60% accurate for *Helicobacter pylori* and 86% accurate for *E. coli* (Price et al., 2003). Nevertheless, the challenges for the construction of these *in silico* models include obtaining high-throughput data to reconstruct more complete models, which can be sorted out by using omics data and combinatorial experimentation, and constructing mathematical approaches to model and render specific solutions for the highly complex systems of biological networks. Because genome-scale metabolic networks comprise hundreds to thousands of reactions, a large number of parameters are required to mathematically describe networks, which, therefore, requires the development of informatic intensive modeling approaches to describe its complexity and to make useful predictions regarding phenotypic behavior for strain design (Price et al., 2003).

The most used approaches are those arising from stoichiometric modeling, which uses mass balances over the metabolic network and assumes a pseudo-steady-state condition to determine intracellular metabolic fluxes, along with additional experimental data to solve the underdetermined linear equation

system (Åkesson et al., 2004). Stoichiometric modeling creates a matrix ( $S$ ) for the metabolites and metabolic reactions, in which each element indicates a stoichiometric coefficient, along with a vector that contains all of the unknown reaction rates ( $v$ ); under the steady state assumption, flux distribution will be represented by  $Sv = 0$  (Jouhten, 2012; Kim et al., 2012). As expected, this equation system will have many solutions, or more precisely, it will render a convex solution space, and because genome-scale metabolic models include all possible metabolic reactions whether or not they are expressed, meaningful solutions must be narrowed down to render a viable solution (Kim et al., 2012). The main problem is that due to the high number of equations and parameters, these systems are always underdetermined; thus, the use of thermodynamic, metabolic, kinetic, and all other experimental data available is required to impose constraints, to reveal a plausible solution, and therefore to conduct quantitative analysis and make predictions regarding cell behavior (Fong, 2014).

To accomplish this desirable outcome, mathematical modeling researchers have developed many approaches to render the complexity, including the use of interaction-based, constraint-based, and mechanism-based methodologies for calculations. Interaction-based approaches isolate autonomous units performing distinct functions in cellular systems, accounting for modularity, which simplifies networks and systems to perform a topological analysis to reveal the principles of cellular organization. Constraint-based approaches account for the physicochemical invariance of networks in addition to network topology. This approach along with stoichiometric modeling, is capable of confining the numerous steady-state flux distributions the metabolic reconstruction network can have (convex space of solutions) into a smaller group, which complies with the constraints indicated by the knowledge regarding the system (a set of feasible states). Even so, this approach accounts only for the steady state, and therefore produces static models; thus, the final phenotypic behavior in changing intracellular or extracellular environments is difficult to address (Stelling, 2004). Mechanism-based approaches use kinetic parameters along with stoichiometric parameters to render the dynamic behavior of cells; thus, such approaches can formulate precise flux distributions and explore the regulation over time. The main problem with this approach is that the knowledge on mechanisms and associated parameters (kinetic reaction parameters) has, thus, far been limited, as so much effort and so many resources must be used to accomplish this type of models (Stelling, 2004; Jouhten, 2012; Long et al., 2015).

Constraint-based approaches are the most used ones to date because of their capability to render flux distribution modeling even with a relatively small amount of information. These approaches state the constraints under which the reconstructed network operates based on stoichiometry and thermodynamics, including directionality and biochemical loops (Price et al., 2003). Such constraints can be imposed by linear optimization; for example, standard flux-base analysis (FBA) uses growth optimization, selecting only the flux solutions, that in turn, produce the maximum growth rate for network topology (Åkesson et al., 2004). Newer flux solution reduction methods have been developed to study the solution space, accounting for the

optimization of not only growth but also many other linear and non-linear objective functions, such as the maximum biomass, maximum ATP, minimum overall intracellular flux, maximum ATP yield per flux unit, maximum biomass yield per flux unit, maximum substrate consumption, minimum number of reaction steps, minimum redox potential, and minimum flux production between others (Price et al., 2003; Schuetz et al., 2007). These optimization principles, along with other constraints arising from specific conditions being either biotic (e.g., the inactivation, subexpression, or overexpression of specific target genes) or abiotic (e.g., aerobic culture, anaerobic culture, nitrogen limitation, carbon limitation, available substrates), will help not only to render the most feasible flux distribution solution but also to study the consequences of changing the genetic cellular output or fermentation parameters for a specific objective. This information is of great use for ME because it renders the ability through different modeling frameworks to study and predict the effects of knocking out genes, tuning the expression of target genes involved in specific reactions, network robustness, the endpoint of adaptive evolution, the identification and characterization of regulation, and heterologous reactions and *de novo* reactions on strain design. There are many reviews that discuss and compare multiple modeling frameworks, such as OptGene, OptStrain, CosMos, OptForce, FaceCon, and FOCAL, for the constraint-based analysis of genome-wide metabolic networks (Price et al., 2003; Schuetz et al., 2007; Krull and Wittmann, 2010; Cloots and Marchal, 2011; Jouhten, 2012; Fong, 2014; King et al., 2015; Long et al., 2015). In this review, we will focus only on one or two framework examples given the scope of this work.

The first strain design method involving knockouts is OptKnock, a bi-level optimization framework used to identify optimal reaction deletion strategies, coupling cellular growth, and target metabolite production. OptKnock identifies deletions with the highest chemical production within the solution space obtained by the maximum growth rate constraint (Long et al., 2015). Pharkya et al. (2003) used this framework to explore the overproduction of amino acids; specifically for AA, they addressed the channeling of flux from PEP to AA by removing the *ppc* gene, which could lead to the redirection of carbon flux to the formation of CHA via the accompanying deletions of pyruvate oxidase, pyruvate dehydrogenase, and pyruvate lyase reactions. The deletion of *ppc* by itself fails to redirect PEP to AA; the ability to detect its contribution through the co-inactivation of other reactions is a very useful tool of ME because the classical experimental deletion of this gene would have produced negative results for pathway optimization. In other words, *in silico* modeling enables researchers to avoid designs toward a local maxima or minima when trying to identify the modifications required to achieve a global maxima for their specific purposes.

FBA with grouping reaction constraints (FBAwGR) was developed to improve the accuracy of metabolic simulation by incorporating the grouping of reaction constraints of functionally and physically related reactions in the model. This framework allows the consideration of genomic context and flux-converging analyses. Genomic context accounts for conserved neighborhoods, gene fusion, and co-occurrences of genes to organize fluxes that are likely to be on or off together. Flux-converging

analyses then restrict the carbon flux solution space to the number of metabolites participating in reactions and converging patterns from a specific carbon source. This framework has been used to predict changes in flux patterns caused by several genetic modifications, such as *pykF*, *zwf*, *ppc*, and *sucA* deletions in *E. coli*, showing good agreements with experimentally obtained fluxes (Kim et al., 2012).

Regarding the scope of this review for SA production in *E. coli*, we have found few studies in the literature that account for metabolic modeling. Nevertheless, the notable work by Chen et al. (2011), described FBA constraint analysis by stoichiometry and mass balance, assuming no growth and optimizing SA as the objective function to design modifications for the production of intermediate metabolites of the aromatic pathway. The model identified several key reaction steps for overexpression, similarly to those previously reported for AA optimization (overexpression of the *aroF*, *tktA*, *ppsA*, and *glf* genes, as well as deletions of the *ldhA* and *ackA* genes) by avoiding carbon waste through lactate and acetate fluxes. Finally, with all of the modifications made, their model identified the *zwf* gene as the critical node for redirection of the carbon flux into the AA pathway; its deletion led to an optimized accumulation of QA, GA, and SA, accounting for a 47% molar conversion of glucose (Chen et al., 2014).

Regarding other SA related work, Rizk and Liao (2009), managed to use EM to model, study, and predict DAHP production in *E. coli* toward aromatic production. Ensemble modeling (EM) is a mechanism-based modeling approach that decomposes metabolic reactions into elementary reaction steps, incorporating all available phenotypic observations for the wild type and mutant strains, integrating this information into the mathematical approach to identify the kinetic variables of each elementary reaction step (Rizk and Liao, 2009; Khodayari et al., 2014). Rizk and Liao (2009), using different flux bounds on the pathway split ratio between glycolysis and the PPP. Then, by using data from literature for overexpression of the *tktA*, *talA*, and *pps* genes, they were able to screen the solution space models compared with the phenotypic behavior, selecting the ones that properly described the experimental data (from a 1500 solution space to 7, 171, and 195 solution spaces, according to glycolysis:PPP ratios of 25:75, 75:25, and 95:5, respectively). This subset of flux solutions revealed that TktA is the first controlling rate step and that PPS, only with simultaneous overexpression of TktA can augment DAHP production; these findings are in accordance with the phenotypic observations in the literature. Based on these results, they conclude that the flux distributions found could be reverse engineered to enhance aromatic production in *E. coli* (Rizk and Liao, 2009).

Notably, despite the existence of many genome-scale metabolic models and various mathematical approaches, many of the fluxes remain undetermined, as many solutions remain plausible. Thus, more information is needed to ensure the modeling quality by the validation and incorporation of *in vivo* experimental data. These experimental data can be acquired from transcriptomic, proteomic, or fluxomic data. Strategies incorporating these extensive experimental data have been developed to enhance the quality and the accuracy of metabolic models (Kim et al., 2012). Fluxomic data in its core provide us with the most important

information as fluxes are the modeling outcome, but experimental procedures can only be used for relatively smaller networks and in specific conditions. Nevertheless, these data are of utmost importance and are commonly used to validate model solutions to flux distributions. ME models have been used to integrate protein expression data to reconstruct and add constraints to genome-level metabolic models, relating kinetic equations into catalytic constraints to approximate stoichiometric relationships between enzyme abundance and catalyzed fluxes (O'Brien and Palsson, 2015). This integration of proteomic data adds thermodynamic and allocation constraints that help in the identification of a consistent flux state, allowing an explanation of aspects of cell behavior and relationships that have remained elusive, such as the interaction of ribosomes with metabolism, carbon limited to carbon excess metabolic shifts, substrate uptake regulation, membrane protein relationships, and other protein spatial constraints that can utterly dominate and/or change metabolic responses (O'Brien and Palsson, 2015). Transcriptomic data have been used to exploit the regulatory information in the expression data to provide additional constraints for the metabolic fluxes in the model by analyzing if or when gene expression correlates with a given metabolic flux (Åkesson et al., 2004; Kern et al., 2006). Computational protocols have been developed for this type of data integration, such as mixed integer linear programming (MILP), which seeks to maximize the agreement between experimental data and computational fluxes by limiting the presentation of entities with the capability to carry flux; meanwhile, the flux of absent entities would be 0 (Fong, 2014). Åkesson et al. (2004) used gene expression microarray data from chemostat and batch cultures of *S. cerevisiae* to create Boolean variables for all of the reactions encompassed on a genome-scale metabolic model to ascertain the absent/present fluxes using analysis software. These new constraints allowed the computation of metabolic flux distributions to enhance the metabolic behavior in batch cultures, along with the quantitative prediction of exchange fluxes as well as the qualitative estimation of changes in intracellular fluxes compared with the model without transcription constraints, as verified by experimental measurements of flux (Åkesson et al., 2004).

Many methods have been developed to introduce transcriptomic regulation into modeling predictions, such as probabilistic regulation of metabolism (PROM), which calculates the probability that a metabolic target gene will be expressed relative to the activity of its regulating transcription factor, metabolic adjustment by differential expression (MADE), which creates a sequence of binary expression states so that when the gene expression changes from one condition to another, the flux reaction will change in accordance with its value, and gene inactivity moderated by metabolism and expression (GIMME), which is a context metabolic model that predicts the subsets of reactions used under a particular condition using gene expression data and which identifies a flux distribution to optimize a given biological objective, such as growth and/or ATP production, along with FBA (Kim et al., 2012). Finally, a method called E-Flux can map continuous gene expression into flux bound constraints according to gene-protein-reaction (GPR) associations, limiting the upper and lower bounds on fluxes so that genes expressed at higher

levels will result in higher flux values (Kim et al., 2012). This and other methods have been reviewed and compared by Machado and Herrgård (2014), who concluded that the prediction of flux levels from gene expression remains far from solved because the predictions obtained by simple FBA with growth maximization and parsimony criteria were as good or even better than those obtained using the incorporation of transcriptomic data. Nevertheless, they acknowledge that some methods evaluated give reasonable predictions under certain conditions that there is no universal method that performs well under all scenarios and that the transcriptome should provide some guidelines for the correct phenotype determination within the space of solutions resulting from the large number of degrees of freedom in metabolic networks, recommending that users should perform a careful evaluation of the meaningfulness of the results for their particular applications (Machado and Herrgård, 2014).

There are many successful mathematical modeling approaches to produce good and accurate predictions of phenotypic behavior in the literature; all of these methods help us to understand and simplify metabolic regulation and systems to comprehend and find new or non-intuitive targets for ME. Even so, there is still much work to be conducted to understand and construct better models of metabolic networks. There are many challenges because cell behavior is a complex system that, therefore, has complex outcomes and regulation. These challenges range from network reconstruction, mathematical treatments, and true flux distribution determination to the integration of all systems data (omics) to achieve regulation and phenotypic predictions. Nevertheless, the effort put into understanding this matter has produced and will continue to produce new insights for strain design and ME. Explaining all the considerations, challenges and achievements in this field is not within the scope of this review as many reviews have been published on these matters (Liu et al., 2004; Patil et al., 2004; Stelling, 2004; Schuetz et al., 2007; Kim et al., 2012; Machado and Herrgård, 2014; Saha et al., 2014; Long et al., 2015; O'Brien and Palsson, 2015). Rather, this review is aimed to provide the reader with interesting findings and perspective on how the mathematical modeling of biological systems can be and is useful for ME, especially regarding SA and AA production, for which these methods can be of relevance to exploit the maximum production capability of *E. coli* that remains unachieved.

## Summary and Perspectives

SA is a key intermediate of the common aromatic pathway with diverse applications in the synthesis of valuable pharmaceutical compounds, but major interest relies on SA as the precursor for the chemical synthesis of OSF, the neuraminidase inhibitor of diverse influenza viruses, including pandemic strains. Diverse efforts have been made to produce high titers and yields of SA in metabolically engineered strains of *E. coli* with successful genetic modifications, including the following: (1) interruption of the SA pathway by the inactivation of shikimate kinase coding genes (*aroK* and *aroL*), which results in the high accumulation of SA; (2) increasing the intracellular availability of the CCM intermediate PEP by inactivation of the PTS system and replacing this glucose translocation system by other housekeeping or

heterologous glucose transporters and by inactivation of the *pykF* gene; and (3) the overexpression of diverse key genes of the CCM and SA pathways, such as *zwf*, *tktA*, *aroB*, *aroD*, and *aroE*, under the control of constitutively expressed or inducible promoters in plasmid-cloned operons or chromosome-integrated copies. These engineered strains have been cultured in batch or fed-batch culture conditions using a complex fermentation media including glucose and YE, resulting in the highest titer and yield of SA reported (Chandran et al., 2003; Rodriguez et al., 2013).

The above-described genetic changes impose global nutritional, regulatory, and metabolic constraints on the resultant engineered strains, which must be explored to determine their relevance on SA production. GTA of the SA-producing strain W3110.shik1 provided evidence supporting the roles of *ydiB*-, *aroD*-, and *ydiN*-encoded proteins in byproduct formation during SA production under glucose limiting conditions (Johansson and Lidén, 2006). Recent ME strategies applied for L-tyrosine (Juminaga et al., 2012) and SA production (Rodriguez et al., 2013) demonstrated the relevance of *ydiB* inactivation and *aroD* overexpression to avoid byproduct formation and to improve carbon flux toward the desired aromatic products.

Interruption of the SA pathway by inactivation of the *aroK* and *aroL* genes imposes an auxotrophic requirement for aromatic AAs and probably other metabolites derived from CHA on the cell; these effects were successfully reversed by the addition of YE to the fermentation media.

As the chemical complexity of YE or peptone significantly interferes in the study of carbon flux through the CCM and SA pathway metabolic networks, no studies to date have been reported on the application of metabolic models to identify possible targets for the application of further ME strategies focused on the improvement of SA production in fermentation culture using complex production media (Chandran et al., 2003; Escalante et al., 2010; Chen et al., 2012; Rodriguez et al., 2013; Cui et al., 2014). The application of omics, such as GTA, in SA-producing conditions, including YE, as reported for the strain P12.SA22, provides valuable information on the role

of diverse transporter systems and other pathways involved in carbon supply from YE to SA synthesis (Cortés-Tolalpa et al., 2014). These results highlight the relevance of information retrieved from the application of omics, such as GTA, or proteomic approaches in successful aromatic compound-producing strains to obtain data for mathematical modeling of metabolism.

Further application of synthetic biology strategies based on modular combinational design including key genes from the CCM and SA pathways in operons and optimized codon usage, and the construction of continuous genetic modules regulated by the same promoter but coupled to an efficient translational level by the selection of efficient ribosome binding sites (RBS) from tailored-made RBS libraries are promising strategies for the subsequent optimization of SA-producing strains. These synthetic strategies have been applied for the efficient production of L-tyrosine in *E. coli* (Juminaga et al., 2012) and for the successful production of SA in *Corynebacterium glutamicum* (Zhang et al., 2015), respectively. Great advances in SA production in *E. coli* have been made over the past decades. However, more and new developments must be made, taking into account the vast, recently acquired data from omics technology. These data, along with their integration with ME technology and experience, can lead to more global insight into cell physiology, allowing new engineering techniques from a systems ME perspective to be identified and developed.

## Author Contributions

All authors participated equally in the preparation of this contribution. All authors have read and approved the final manuscript.

## Acknowledgments

This work was supported by CONACYT Ciencia Básica project 240519.

## References

- Aguilar, C., Escalante, A., Flores, N., de Anda, R., Riveros-McKay, F., Gosset, G., et al. (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. doi:10.1186/1471-2164-13-385
- Åkesson, M., Förster, J., and Nielsen, J. (2004). Integration of gene expression data into genome-scale metabolic models. *Metab. Eng.* 6, 285–293. doi:10.1016/j.ymben.2003.12.002
- Báez, J. L., Bolívar, F., and Gosset, G. (2001). Determination of 3-deoxy-D-arabino-heptulosonate 7-phosphate productivity and yield from glucose in *Escherichia coli* devoid of the glucose phosphotransferase transport system. *Biotechnol. Bioeng.* 73, 530–535. doi:10.1002/bit.1088
- Báez-Viveros, J. L., Osuna, J., Hernández-Chávez, G., Soberón, X., Bolívar, F., and Gosset, G. (2004). Metabolic engineering and protein directed evolution increase the yield of L-phenylalanine synthesized from glucose in *Escherichia coli*. *Biotechnol. Bioeng.* 87, 516–524. doi:10.1002/bit.20159
- Biggs, B. W., De Paepe, B., Santos, C. N. S., De Mey, M., and Kumaran Ajikumar, P. (2014). Multivariate modular metabolic engineering for pathway and strain optimization. *Curr. Opin. Biotechnol.* 29, 156–162. doi:10.1016/j.copbio.2014.05.005
- Bongaerts, J. (2001). Metabolic engineering for microbial production of aromatic amino acids and derived compounds. *Metab. Eng.* 3, 289–300. doi:10.1006/mben.2001.0196
- Chandran, S. S., Yi, J., Draths, K. M., von Daeniken, R., Weber, W., and Frost, J. W. (2003). Phosphoenolpyruvate availability and the biosynthesis of shikimic acid. *Biotechnol. Prog.* 19, 808–814. doi:10.1021/bp025769p
- Chen, K., Dou, J., Tang, S., Yang, Y., Wang, H., Fang, H., et al. (2012). Deletion of the *aroK* gene is essential for high shikimic acid accumulation through the shikimate pathway in *E. coli*. *Bioresour. Technol.* 119, 141–147. doi:10.1016/j.biortech.2012.05.100
- Chen, P. T., Chiang, C.-J., Wang, J.-Y., Lee, M.-Z., and Chao, Y.-P. (2011). Genomic engineering of *Escherichia coli* for production of intermediate metabolites in the aromatic pathway. *J. Taiwan Inst. Chem. Eng.* 42, 34–40. doi:10.1016/j.jtice.2010.03.010
- Chen, X., Li, M., Zhou, L., Shen, W., Algasan, G., Fan, Y., et al. (2014). Metabolic engineering of *Escherichia coli* for improving shikimate synthesis from glucose. *Bioresour. Technol.* 166, 64–71. doi:10.1016/j.biortech.2014.05.035
- Chen, X., Zhou, L., Tian, K., Kumar, A., Singh, S., Prior, B. A., et al. (2013). Metabolic engineering of *Escherichia coli*: a sustainable industrial platform for bio-based chemical production. *Biotechnol. Adv.* 31, 1200–1223. doi:10.1016/j.biotechadv.2013.02.009
- Cloots, L., and Marchal, K. (2011). Network-based functional modeling of genomics, transcriptomics and metabolism in bacteria. *Curr. Opin. Microbiol.* 14, 599–607. doi:10.1016/j.mib.2011.09.003

- Cortés-Tolalpa, L., Gutiérrez-Ríos, R. M., Martínez, L. M., De Anda, R., Gosset, G., Bolívar, F., et al. (2014). Global transcriptomic analysis of an engineered *Escherichia coli* strain lacking the phosphoenolpyruvate: carbohydrate phosphotransferase system during shikimic acid production in rich culture medium. *Microb. Cell Fact.* 13, 28. doi:10.1186/1475-2859-13-28
- Covert, M. W., Schilling, C. H., Famili, I., Edwards, J. S., Goryanin, I. I., Selkov, E., et al. (2001). Metabolic modeling of microbial strains *in silico*. *Trends Biochem. Sci.* 26, 179–186. doi:10.1016/S0968-0004(00)01754-0
- Cui, Y.-Y., Ling, C., Zhang, Y.-Y., Huang, J., and Liu, J.-Z. (2014). Production of shikimic acid from *Escherichia coli* through chemically inducible chromosomal evolution and cofactor metabolic engineering. *Microb. Cell Fact.* 13, 21. doi:10.1186/1475-2859-13-21
- Díaz Quiroz, D. C., Carmona, S. B., Bolívar, F., and Escalante, A. (2014). Current perspectives on applications of shikimic acid aminoshikimic acids in pharmaceutical chemistry. *Res. Rep. Med. Chem.* 4, 35–46. doi:10.2147/RRMC.S46560
- Draths, K. M., Knop, D. R., and Frost, J. W. (1999). Shikimic acid and quinic acid: replacing isolation from plant sources with recombinant microbial biocatalysis. *J. Am. Chem. Soc.* 121, 1603–1604. doi:10.1021/ja9830243
- Escalante, A., Calderón, R., Valdivia, A., de Anda, R., Hernández, G., Ramírez, O. T., et al. (2010). Metabolic engineering for the production of shikimic acid in an evolved *Escherichia coli* strain lacking the phosphoenolpyruvate: carbohydrate phosphotransferase system. *Microb. Cell Fact.* 9, 21. doi:10.1186/1475-2859-9-21
- Escalante, A., Salinas Cervantes, A., Gosset, G., and 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. doi:10.1007/s00253-012-4101-5
- Estevez, A., and Estevez, R. (2012). A short overview on the medicinal chemistry of (-)-shikimic acid. *Mini Rev. Med. Chem.* 12, 1443–1454. doi:10.2174/138955712803832735
- F. Hoffmann-La Roche Ltd. (2009). *Roche Annual Report 09*. Basel, Switzerland: F. Hoffmann-La Roche Ltd. 127 p. Available at: <http://www.roche.com/gb09e.pdf>
- Flores, N., Leal, L., Sigala, J. C., de Anda, R., Escalante, A., Martínez, A., et al. (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. doi:10.1159/000103602
- Flores, N., Xiao, J., Berry, A., Bolívar, F., and Valle, F. (1996). Pathway engineering for the production of aromatic compounds in *Escherichia coli*. *Nat. Biotechnol.* 14, 620–623. doi:10.1038/nbt0596-620
- Fong, S. S. (2014). Computational approaches to metabolic engineering utilizing systems biology and synthetic biology. *Comput. Struct. Biotechnol. J.* 11, 28–34. doi:10.1016/j.csbj.2014.08.005
- Frost, J. W., Frost, K. M., and Knop, D. R. (2002). *Biocatalytic Synthesis of Shikimic Acid*. Available at: <http://www.google.com/patents/US6472169> (accessed July 5, 2015).
- Ghosh, S., Chisti, Y., and Banerjee, U. C. (2012). Production of shikimic acid. *Biotechnol. Adv.* 30, 1425–1431. doi:10.1016/j.biotechadv.2012.03.001
- Gosset, G. (2009). Production of aromatic compounds in bacteria. *Curr. Opin. Biotechnol.* 20, 651–658. doi:10.1016/j.copbio.2009.09.012
- Johansson, L., and Lidén, G. (2006). Transcriptome analysis of a shikimic acid producing strain of *Escherichia coli* W3110 grown under carbon- and phosphate-limited conditions. *J. Biotechnol.* 126, 528–545. doi:10.1016/j.jbiotec.2006.05.007
- Johansson, L., Lindskog, A., Silfversparre, G., Cimander, C., Nielsen, K. F., and Lidén, G. (2005). Shikimic acid production by a modified strain of *E. coli* (W3110.shik1) under phosphate-limited and carbon-limited conditions. *Biotechnol. Bioeng.* 92, 541–552. doi:10.1002/bit.20546
- Jouhten, P. (2012). Metabolic modelling in the development of cell factories by synthetic biology. *Comput. Struct. Biotechnol. J.* 3, 1–9. doi:10.5936/csbj.201210009
- Juminaga, D., Baidoo, E. E. K., Redding-Johanson, A. M., Batth, T. S., Burd, H., Mukhopadhyay, A., et al. (2012). Modular engineering of L-tyrosine production in *Escherichia coli*. *Appl. Environ. Microbiol.* 78, 89–98. doi:10.1128/AEM.06017-11
- Kedar, P., Colah, R., and Shimizu, K. (2007). Proteomic investigation on the *pyk-F* gene knockout *Escherichia coli* for aromatic amino acid production. *Enzyme Micro. Technol.* 41, 455–465. doi:10.1016/j.enzmictec.2007.03.018
- Kern, R., Malki, A., Abdallah, J., Tagourti, J., and Richarme, G. (2006). *Escherichia coli* HdeB is an acid stress chaperone. *J. Bacteriol.* 189, 603–610. doi:10.1128/JB.01522-06
- Keseler, I. M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martínez, C., et al. (2013). EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res.* 41, D605–D612. doi:10.1093/nar/gks1027
- Khodayari, A., Zomorodi, A. R., Liao, J. C., and Maranas, C. D. (2014). A kinetic model of *Escherichia coli* core metabolism satisfying multiple sets of mutant flux data. *Metab. Eng.* 25, 50–62. doi:10.1016/j.ymben.2014.05.014
- Kim, T. Y., Sohn, S. B., Kim, Y. B., Kim, W. J., and Lee, S. Y. (2012). Recent advances in reconstruction and applications of genome-scale metabolic models. *Curr. Opin. Biotechnol.* 23, 617–623. doi:10.1016/j.copbio.2011.10.007
- King, Z. A., Lloyd, C. J., Feist, A. M., and Palsson, B. O. (2015). Next-generation genome-scale models for metabolic engineering. *Curr. Opin. Biotechnol.* 35, 23–29. doi:10.1016/j.copbio.2014.12.016
- Knop, D. R., Draths, K. M., Chandran, S. S., Barker, J. L., von Daeniken, R., Weber, W., et al. (2001). Hydroaromatic equilibration during biosynthesis of shikimic acid. *J. Am. Chem. Soc.* 123, 10173–10182. doi:10.1021/ja0109444
- Krämer, M., Bongaerts, J., Bovenberg, R., Kremer, S., Müller, U., Orf, S., et al. (2003). Metabolic engineering for microbial production of shikimic acid. *Metab. Eng.* 5, 277–283. doi:10.1016/j.ymben.2003.09.001
- Krull, R., and Wittmann, C. (2010). *Linking Cellular Networks and Bioprocesses*. Heidelberg: Springer.
- Li, S., Yuan, W., Wang, P., Zhang, Z., Zhang, W., and Ownby, S. (2007). *Method for the Extraction and Purification of Shikimic Acid*. Available at: <http://www.google.com/patents/US20070149805>
- Lin, S., Liang, R., Meng, X., OuYang, H., Yan, H., Wang, Y., et al. (2014). Construction and expression of mutagenesis strain of *aroG* gene from *Escherichia coli* K-12. *Int. J. Biol. Macromol.* 68, 173–177. doi:10.1016/j.ijbiomac.2014.04.034
- Liu, A., Liu, Z. Z., Zou, Z. M., Chen, S. Z., Xu, L. Z., and Yang, S. L. (2004). Synthesis of (+)-zeylonene from shikimic acid. *Tetrahedron* 60, 3689–3694. doi:10.1016/j.tet.2004.02.066
- Liu, Q., Cheng, Y., Xie, X., Xu, Q., and Chen, N. (2012). Modification of tryptophan transport system and its impact on production of L-tryptophan in *Escherichia coli*. *Bioresour. Technol.* 114, 549–554. doi:10.1016/j.biortech.2012.02.088
- Long, M. R., Ong, W. K., and Reed, J. L. (2015). Computational methods in metabolic engineering for strain design. *Curr. Opin. Biotechnol.* 34, 135–141. doi:10.1016/j.copbio.2014.12.019
- Lütke-Eversloh, T., and Stephanopoulos, G. (2008). Combinatorial pathway analysis for improved L-tyrosine production in *Escherichia coli*: identification of enzymatic bottlenecks by systematic gene overexpression. *Metab. Eng.* 10, 69–77. doi:10.1016/j.ymben.2007.12.001
- Machado, D., and Herrgård, M. (2014). Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLoS Comput. Biol.* 10:e1003580. doi:10.1371/journal.pcbi.1003580
- Matsuoka, Y., and Shimizu, K. (2012). Importance of understanding the main metabolic regulation in response to the specific pathway mutation for metabolic engineering of *Escherichia coli*. *Comput. Struct. Biotechnol. J.* 3, 1–10. doi:10.5936/cs bj.201210018
- O'Brien, E. J., and Palsson, B. O. (2015). Computing the functional proteome: recent progress and future prospects for genome-scale models. *Curr. Opin. Biotechnol.* 34, 125–134. doi:10.1016/j.copbio.2014.12.017
- Patil, K. R., Åkesson, M., and Nielsen, J. (2004). Use of genome-scale microbial models for metabolic engineering. *Curr. Opin. Biotechnol.* 15, 64–69. doi:10.1016/j.copbio.2003.11.003
- Patnaik, R., and Liao, J. C. (1994). Engineering of *Escherichia coli* central metabolism for aromatic metabolite production with near theoretical yield. *Appl. Environ. Microbiol.* 60, 3903–3908.
- Patnaik, R., Spitzer, R. G., and Liao, J. C. (1995). Pathway engineering for production of aromatics in *Escherichia coli*: confirmation of stoichiometric analysis by independent modulation of AroG, TktA, and Pps activities. *Biotechnol. Bioeng.* 46, 361–370. doi:10.1002/bit.260460409
- Pharkya, P., Burgard, A. P., and Maranas, C. D. (2003). Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock. *Biotechnol. Bioeng.* 84, 887–899. doi:10.1002/bit.10857
- Price, N. D., Papin, J. A., Schilling, C. H., and Palsson, B. O. (2003). Genome-scale microbial *in silico* models: the constraints-based approach. *Trends Biotechnol.* 21, 162–169. doi:10.1016/S0167-7799(03)00030-1
- Raghavendra, T. R., Vaidyanathan, P., Swathi, H. K., Ramesha, B. T., Ravikanth, G., Ganeshiah, K. N., et al. (2009). Prospecting for alternate sources of shikimic acid, a precursor of Tamiflu, a bird-flu drug. *Curr. Sci.* 96, 771–772.

- Rangachari, S., Friedamna, T. C., Hartmann, R., and Weisenfeld, R. B. (2013). *Processes for Producing and Recovering Shikimic Acid*. Available at: <http://www.google.com/patents/US8344178> (accessed July 5, 2015).
- Rizk, M. L., and Liao, J. C. (2009). Ensemble modeling for aromatic production in *Escherichia coli*. *PLoS ONE* 4:e6903. doi:10.1371/journal.pone.0006903
- Rodríguez, A., Martínez, J. A., Báez-Viveros, J. L., Flores, N., Hernández-Chávez, G., Ramírez, O. T., et al. (2013). 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 Fact.* 12, 86. doi:10.1186/1475-2859-12-86
- Rodríguez, A., Martínez, J. A., Flores, N., Escalante, A., Gosset, G., and Bolívar, F. (2014). Engineering *Escherichia coli* to overproduce aromatic amino acids and derived compounds. *Microb. Cell Fact.* 13, 126. doi:10.1186/s12934-014-0126-z
- Saha, R., Chowdhury, A., and Maranas, C. D. (2014). Recent advances in the reconstruction of metabolic models and integration of omics data. *Curr. Opin. Biotechnol.* 29, 39–45. doi:10.1016/j.copbio.2014.02.011
- Scheiwiller, T., and Hirschi, S. (2010). *09 Roche Annual Report*. Finance Report. Basel: F. Hoffmann-LaRoche Ltd.
- Schuetz, R., Kuepfer, L., and Sauer, U. (2007). Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Mol. Syst. Biol.* 3, 119. doi:10.1038/msb4100162
- Stelling, J. (2004). Mathematical models in microbial systems biology. *Curr. Opin. Microbiol.* 7, 513–518. doi:10.1016/j.mib.2004.08.004
- Wang, G.-W., Hu, W.-T., Huang, B.-K., and Qin, L.-P. (2011). *Illicium verum*: a review on its botany, traditional use, chemistry and pharmacology. *J. Ethnopharmacol.* 136, 10–20. doi:10.1016/j.jep.2011.04.051
- World Health Organization. (2011). *Pandemic Influenza Preparedness. Framework for the Sharing of Influenza Viruses and Access to Vaccines and Other Benefits*. France: World Health Organization. Available at: [http://www.who.int/influenza/resources/pip\\_framework/en/](http://www.who.int/influenza/resources/pip_framework/en/)
- Yi, J., Li, K., Draths, K. M., and Frost, J. W. (2002). Modulation of phosphoenolpyruvate synthase expression increases shikimate pathway product yields in *E. coli*. *Biotechnol. Prog.* 18, 1141–1148. doi:10.1021/bp020101w
- Zhang, B., Zhou, N., Liu, Y.-M., Liu, C., Lou, C.-B., Jiang, C.-Y., et al. (2015). Ribosome binding site libraries and pathway modules for shikimic acid synthesis with *Corynebacterium glutamicum*. *Microb. Cell Fact.* 14, 71. doi:10.1186/s12934-015-0254-0
- Zucko, J., Dunlap, W. C., Shick, J. M., Cullum, J., Cercelet, F., Amin, B., et al. (2010). Global genome analysis of the shikimic acid pathway reveals greater gene loss in host-associated than in free-living bacteria. *BMC Genomics* 11:628. doi:10.1186/1471-2164-11-628

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Martínez, Bolívar and Escalante. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

REVIEW

Open Access

# Engineering *Escherichia coli* to overproduce aromatic amino acids and derived compounds

Alberto Rodriguez, Juan A Martínez, Noemí Flores, Adelfo Escalante, Guillermo Gosset and Francisco Bolivar\*

## Abstract

The production of aromatic amino acids using fermentation processes with recombinant microorganisms can be an advantageous approach to reach their global demands. In addition, a large array of compounds with alimentary and pharmaceutical applications can potentially be synthesized from intermediates of this metabolic pathway. However, contrary to other amino acids and primary metabolites, the artificial channelling of building blocks from central metabolism towards the aromatic amino acid pathway is complicated to achieve in an efficient manner. The length and complex regulation of this pathway have progressively called for the employment of more integral approaches, promoting the merge of complementary tools and techniques in order to surpass metabolic and regulatory bottlenecks. As a result, relevant insights on the subject have been obtained during the last years, especially with genetically modified strains of *Escherichia coli*. By combining metabolic engineering strategies with developments in synthetic biology, systems biology and bioprocess engineering, notable advances were achieved regarding the generation, characterization and optimization of *E. coli* strains for the overproduction of aromatic amino acids, some of their precursors and related compounds. In this paper we review and compare recent successful reports dealing with the modification of metabolic traits to attain these objectives.

**Keywords:** Aromatic compounds, *Escherichia coli*, Metabolic engineering, Systems biotechnology, Synthetic biology, Shikimate pathway, Phenylalanine, Tyrosine, Tryptophan

## Introduction

The aromatic amino acids (AAA), L-tryptophan (L-TRP), L-phenylalanine (L-PHE) and L-tyrosine (L-TYR), are the final products of the aromatic biosynthetic pathway comprising the shikimate (SHK) pathway, which connects central carbon metabolism (CCM) with the biosynthesis of chorismate (CHA), the last common precursor in the terminal branches for AAA biosynthesis (Figure 1) [1,2]. These pathways are present in bacteria and in several eukaryotic organisms such as ascomycetes fungi, apicomplexans, and plants [3,4]. The AAA are essential components in the diet of higher animals and humans, hence they are used as dietary supplements (e.g. diet of swine and poultry consisting of grains of corn and soybean is low in L-TRP) and key precursors of industrial and pharmaceutical compounds (e.g. L-PHE is the key ingredient in the synthesis of the artificial sweetener

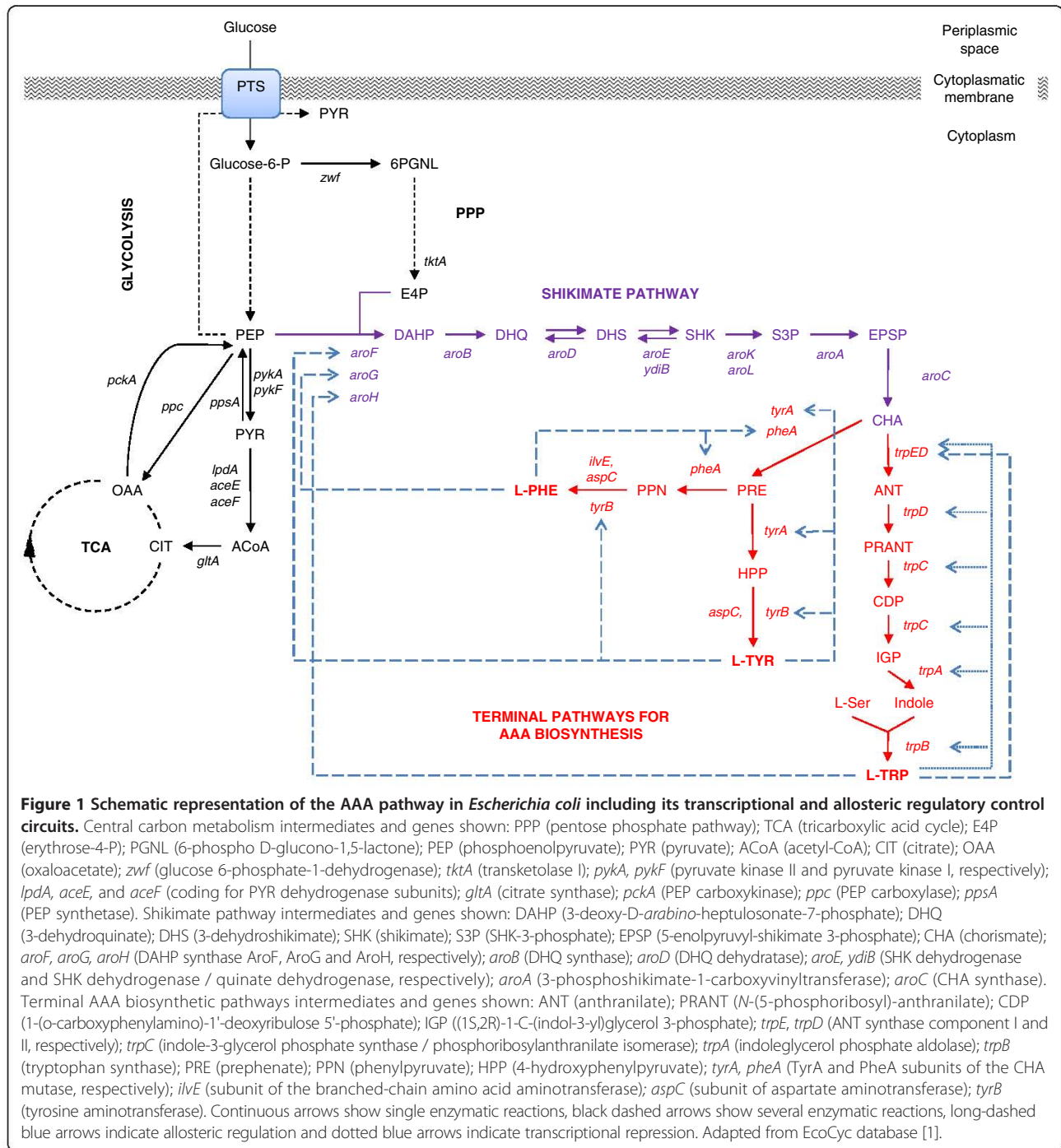
aspartame, whereas L-TYR is an essential dietary component for phenylketonuria patients as the starter material for L-DOPA or melanin production) [5]. The annual worldwide production of amino acids is estimated to be above 4.5 million tons/year, with a market growth for most amino acids of ~10% and higher [6,7]. Among the aromatic amino acids, L-TRP has a market size of more than 14,000 tons/year [8] and the production of L-PHE exceeds 30,000 tons/year [9].

It is well established that the production of high-valued commodities can be performed cost-efficiently by the rational design, modification and cultivation of a recombinant microorganism. In particular, the development of efficient microbial processes for accumulation of compounds derived from the AAA biosynthetic pathway has not been an easy task for metabolic and bioprocess engineers. For more than 20 years, considerable efforts have been directed towards characterizing and purposely overriding the naturally tight metabolic regulation of this pathway. These continued efforts have relied on knowledge obtained from pioneer works on the biosynthesis of aromatic compounds by the

\* Correspondence: bolivar@ibt.unam.mx

Departamento de Ingeniería Celular y Biotecnología, Instituto de Biotecnología, Universidad Nacional Autónoma de México (UNAM), Avenida Universidad 2001, Col. Chamilpa, Cuernavaca, Morelos 62210, México





groups of B.D. Davis, F. Gibson, C. Yanofsky, A.J. Pittard, K.M. Herrmann and J.W. Frost, among others, whose contributions have been comprehensively reviewed in the past [2,10-12].

Recently, the availability of omics-scale data has allowed significant advances in metabolic reconstruction and modeling, resulting in better strain development [13]. Likewise, the increased use of combinatorial and evolutionary approaches, fueled by a rapid expansion of synthetic

molecular tools, opened the possibility for testing novel and large combinations of gene expression systems and genetic backgrounds [14,15]. Additionally, efforts concerning the optimization of fermentation conditions have succeeded in scaling-up many AAA production processes, while simultaneously providing important feedback on the physiological behavior of engineered strains [16,17]. However, the availability of operational tools and techniques, as well as the amount of physiological and molecular information,

are unevenly distributed among the microorganisms currently used for the production of AAA. These circumstances have contributed to positioning *E. coli* as the organism with most reported success cases and has resulted in a wide array of well-characterized production strains [18,19].

In this paper we review some notable advances in the generation, characterization and optimization of *E. coli* strains for the overproduction of AAA, some of their important precursors and related compounds. Although these studies were classified in accordance to the main schemes employed for each case, the constant expansion and complementarity of such approaches has encouraged scientists to apply a systems-based perspective [20,21]. Therefore, recent and representative works on the subject using different strategies were selected and discussed.

#### **Engineering of the CCM: glucose transport, glycolytic, gluconeogenic, and pentose phosphate pathways**

Successful metabolic engineering efforts for the generation of *E. coli* strains that can overproduce AAA include: (i) increasing the availability of the direct precursors phosphoenolpyruvate (PEP) and erythrose-4-phosphate (E4P); (ii) enhancement of the first enzymatic reaction in the SHK pathway to yield 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP); (iii) improving the carbon flow through the biosynthetic pathway of interest by removal of transcriptional and allosteric regulation; (iv) identifying and relieving rate-limiting enzymatic reactions; (v) preventing loss of carbon flow towards competing pathways; (vi) enhancement of product export; and (vii) prevention of product degradation or re-internalization.

Regarding PEP metabolism, *E. coli* uses the phosphotransferase system (PTS) as the main system for the translocation and phosphorylation of glucose from the periplasmic space to the cytoplasmic environment, consuming one PEP molecule which is converted to pyruvate (PYR) [22,23]. This reaction yields one molecule of glucose-6-phosphate which is catabolized by the glycolytic pathway, resulting in two PEP molecules (Figure 1). PEP is a precursor feeding several biosynthetic pathways and also participates in ATP generation, either by substrate-level phosphorylation of ADP or indirectly as an acetyl coenzyme-A (ACoA) precursor. When *E. coli* grows in mineral broth containing glucose as the sole carbon source the PTS consumes 50% of the available PEP, whereas the reactions catalyzed by other enzymes such as PEP carboxylase, PYR kinases, UDP-N-acetylglucosamine enolpyruvyl transferase, and DAHP synthases (DAHPS), consume approximately 16%, 15%, 16%, and 3% of remaining PEP, respectively [23,24]. Therefore, PEP can be converted to PYR by PTS and PYR kinases I and II (coded by *pykF* and *pykA* respectively), and PYR is converted to ACoA by the PYR dehydrogenase

multienzyme complex (coded by *aceE*, *aceF* and *lpd*), a reaction connecting the glycolytic pathway with the tricarboxylic acid cycle (TCA) [1]. Moreover, PEP and PYR are key intermediates of the CCM as they are substrate of at least six enzymes which determine the metabolic fate of these intermediates (biosynthetic/catabolic pathways and glycolytic/gluconeogenic capabilities of the cell): DAHPS isoenzymes (AroF, AroG, and AroH coded by *aroF*, *aroG* and *aroH*, respectively) [3,11]; PYR kinases I and II; PEP synthetase (PpsA coded by *ppsA*); PEP carboxylase (Ppc, coded by *ppc*); and PEP carboxykinase (PckA coded by *pckA*) [25] (Figure 1).

Detailed knowledge of these nodes permitted the development of strategies that allowed higher PEP availability for the biosynthesis of aromatic compounds, including the replacement of glucose transport and phosphorylation capabilities of the PTS by alternative enzymes such as the glucose facilitator and glucokinase from *Zymomonas mobilis* (coded by *glf* and *glk*, respectively) [26-28], the galactose permease and glucokinase from *E. coli* (coded by *galP* and *glk*, respectively) [29,30], or the use of an adaptive evolution process to select PTS<sup>-</sup> derivatives growing at high specific growth rates ( $\mu$ ) on glucose [31,32]. Additionally, high PEP availability has been achieved by modulation of the carbon flux from PEP to the TCA caused by the inactivation of one or both of the PYR kinases [33,34], as well as improving the recycling of PYR to PEP by a plasmid-encoded copy of PEP synthetase [35-37]. The overexpression of *pckA*, in combination with an enhanced carbon flow through the glyoxylate shunt, has also been proposed as a strategy to increase the yield of aromatic compounds [38,39]. An alternative approach to increase PEP is the attenuation of CsrA, a regulatory protein of carbohydrate metabolism, either by direct gene knockout or by increasing the expression of its negative regulatory RNA, coded by *csrB* [40,41].

On the other hand, E4P is a metabolite that participates in reversible reactions present in the non-oxidative branch of the pentose phosphate pathway (PPP), as well as a substrate in irreversible reactions that lead to the production of aromatic amino acids or vitamin B<sub>6</sub> [42]. E4P can also be directly produced from sedoheptulose-1,7-bisphosphate in a reaction that is probably favored when the intracellular levels of sedoheptulose-7-phosphate are high [43]. Metabolic engineering reports have shown that a considerable increase in availability of E4P (inferred by the increased production of aromatic compounds and pathway intermediates, such as DAHP) can be achieved by overexpression of genes coding for a transketolase (*tktA*) [35,44-46] or a transaldolase (*talB*) [26,47]. Additional attempts to increase the carbon flow towards the PPP for enhanced production of aromatic compounds include the use of mutants lacking the enzyme phosphoglucose isomerase [48,49], the overexpression of enzyme

glucose-6-phosphate dehydrogenase [41,50], or the use of multiple carbon sources, mainly hexoses, pentoses and glycerol [51-54]. After an adequate supply of precursors has been established, it is essential to commit this carbon towards the SHK pathway and to remove control points and limiting steps to increase the production of target compounds.

#### Deregulation of the AAA pathway: identifying and relieving rate-limiting steps

In *E. coli*, the DAHPS isoenzymes AroG, AroF and AroH contribute to the total DAHPS activity and are subjected to allosteric control by L-PHE, L-TYR and L-TRP, respectively (Figure 1). AroG contributes about 80% of the overall DAHPS activity, AroF about 15%, and the remaining activity corresponds to AroH DAHPS [3,11]. Both the AroG and AroF isoenzymes are completely inhibited by about 0.1 mM of the corresponding amino acids, but AroH is only partially inhibited by L-TRP. Apparent inability of L-TRP to totally inhibit this isoenzyme is proposed to be a mechanism to ensure a sufficient supply of CHA for the biosynthesis of other aromatic compounds when AAA are present in excess in the growth medium [3]. Specific amino acid residues involved in the allosteric sites have been identified by structural analysis of feedback-insensitive mutant enzymes, resulting in the targeted generation of the feedback resistant (fbr) variants AroG<sup>fbr</sup> and AroF<sup>fbr</sup> [28,31,55]. Additionally to allosteric control of DAHPS isoenzymes, their transcriptional expression can be controlled by the *tyr*- and *trp*- repressors complexed with the AAA [3,11].

Consequently, amplification and deregulation of DAHPS activity is an essential strategy to overproduce aromatic compounds and its precursor SHK. Introduction of plasmid-encoded copies of *aroF*<sup>fbr</sup> and *aroG*<sup>fbr</sup> combined with additional plasmid-cloned gene *tktA*, or their chromosomal integrations in gene clusters, have resulted in increased carbon flow from the CCM to the SHK pathway for the production of L-PHE [11,55,56], L-TYR [5,57,58] and L-TRP [59-61]. Positive results were also obtained with the insertion of an *aroG*<sup>fbr</sup> gene into the chromosome of an L-PHE producing strain while being controlled by a promoter that is active during late cultivation stages, in order to counteract the fall of DAHPS activity in stationary phase [62].

Further increases in carbon flux through the SHK pathway have been attained by the removal of transcriptional and allosteric control points and by relieving limiting enzymatic reactions [2,11,19,23]. The reactions catalyzed by DHQ synthase (encoded by *aroB*) and SHK kinase isoenzymes I and II (encoded by *aroK* and *aroL*, respectively) are considered as rate-limiting [63-65]. In addition, the reaction catalyzed by the enzyme quinate/shikimate dehydrogenase (coded by *ydiB*) was also reported as

limiting in the development of L-TYR production strains [58]. Either the overexpression of some of these genes by plasmid-cloned copies [28,66], their co-expression in a modular operon under control of diverse promoters [50,58,67], or their expression by chromosomal integration of additional gene copies and promoter engineering by chromosomal evolution [68], have relieved to a great extent these rate-limiting steps typically encountered during the development of SHK and AAA overproducing strains (Table 1). To date, genetically modified *E. coli* strains can overproduce SHK from glucose with yields in the range of 0.08 to 0.42 mol SHK / mol glucose under diverse culture conditions [28,50,68-70]. SHK is a key intermediate of the common biosynthetic aromatic pathway (Figure 1) gaining relevance as the substrate for the chemical synthesis of the drug oseltamivir phosphate, known commercially as Tamiflu®, an efficient inhibitor of the surface protein neuraminidase of seasonal influenza, avian influenza H5N1, and human influenza H1N1 viruses [71-74].

In addition to modifications in the SHK pathway, metabolic engineering approaches to overproduce L-TYR typically include alterations in TyrR and/or *trp* regulons. The TyrR regulon comprises diverse essential genes implicated in AAA biosynthesis and transport [1,75]. TyrR acts as a dual transcriptional activator and repressor; however, the repression mechanism requires the ATP-dependent binding of AAA to the central protein domain. L-TYR is the major effector of TyrR-mediated repression, although some repression occurs with L-PHE as co-repressor for *aroE*, *aroL*, *tyrP* (coding for a L-TYR specific permease), *aroP* (coding for an aromatic amino acid permease) and *aroG* genes, whereas activation does not apparently involve an ATP-dependent binding of aromatic amino acids [1,2,5,11]. Inactivation of TyrR-mediated regulation by deletion of *tyrR* and overexpression of *aroG*<sup>fbr</sup> and *tyrA*<sup>fbr</sup>, combined with the overexpression of CCM genes (e.g. *pptsA* and *tktA*) and genes of the L-TYR biosynthetic pathway (e.g. *tyrB*, *aroC*, *aroA*) have improved the production of L-TYR in diverse *E. coli* strains [5,11,58].

Similar results were obtained for L-PHE in resting cells by overexpression of a feedback-resistant or an evolved (ev) CHA mutase/prephenate dehydratase enzymes (coded by *pheA*<sup>fbr</sup> and *pheA*<sup>ev</sup>, respectively) [55,76]. The bifunctional enzyme chorismate mutase/prephenate dehydrogenase TyrA, catalyzes the shared first step in L-PHE and L-TYR final biosynthetic pathways (the conversion of CHA to prephenate), as well as the second step in L-TYR biosynthesis (the subsequent NAD<sup>+</sup>-dependent oxidative decarboxylation of prephenate to 4-hydroxyphenylpyruvate) (Figure 1). TyrA catalyzes both reactions in separate domains of the protein and the CHA mutase/prephenate dehydrogenase is feedback-inhibited by L-TYR (up to 95% inhibition of the prephenate dehydrogenase and 45% of the CHA mutase activity) [1,2]. The bifunctional enzyme CHA

**Table 1 Relevant *E. coli* strains engineered for the overproduction of compounds derived from the aromatic biosynthetic pathway**

Strain	Relevant characteristics	Main compound produced (titer <sup>a</sup> , and/or yield <sup>b, c, d</sup> ). Relevant culture conditions	References
SP1.1pts/pSC6.090B (RB791 derivative)	$\Delta$ ptsHlcr $\Delta$ aroK $\Delta$ aroL serA::aroB / (plasmid) <i>aroF</i> <sup>fbt</sup> <i>tktA</i> , P <sub>tac</sub> <i>aroE serA</i> , P <sub>tac</sub> <i>glf glk</i> <sup>c</sup>	SHK (84, 0.33 <sup>b</sup> ). 10 L fed-batch reactors with glucose, AAA and 15 g/L of yeast extract	[28]
AR36 (JM101 derivative)	$\Delta$ ptsHlcr $\Delta$ aroK $\Delta$ aroL $\Delta$ lacI $\Delta$ pykF / (plasmid) P <sub>trc</sub> <i>aroB tktA aroG</i> <sup>fbt</sup> <i>aroE aroD zwf</i>	SHK (43, 0.42 <sup>b</sup> ). 1 L batch reactors with 100 g/L of glucose and 30 g/L of yeast extract	[50]
SA116 (BW25113 derivative)	$\Delta$ aroK $\Delta$ aroL P <sub>ppp</sub> -P <sub>lacQ1</sub> , P <sub>csrB</sub> -P <sub>lacQ1</sub> / (chromosome) <i>aroG</i> <sup>fbt</sup> <i>tktA aroB aroE</i> , P <sub>T5</sub> <i>ppsA csrB</i> , 5P <sub>tac</sub> <i>tktA nadK</i>	SHK (3, 0.33 <sup>b</sup> ). Medium supplemented with 10 g/L of glucose, 1 g/L of peptone and 1 g/L of proline	[68]
W14/pR15BABKG (W3110 derivative)	$\Delta$ cr $\Delta$ tyrA / (plasmid) P <sub>R</sub> <i>aroG15 tyrB</i> , P <sub>L</sub> <i>pheA</i> <sup>fbt</sup> <i>ydiB aroK yddG</i>	L-PHE (47, 0.25 <sup>d</sup> ). 15 L fed-batch reactors with glucose and 1 g/L of tyrosine	[132]
FUS4.11/pF81 <sub>kan</sub> (W3110 derivative)	$\Delta$ pheA $\Delta$ tyrA $\Delta$ aroF $\Delta$ lacIZYA $\Delta$ pykA $\Delta$ pykF / (chromosome) P <sub>tac</sub> <i>aroF aroB aroL</i> , (plasmid) P <sub>tac</sub> <i>pheA</i> <sup>fbt</sup> <i>aroF aroB aroL</i>	L-PHE (13, 0.15 <sup>d</sup> ). 15 L multi-phase fed-batch reactors with glycerol and lactic acid	[131]
BL21 (DE3)	(plasmid) containing the phenylalanine dehydrogenase gene of <i>Acinetobacter lwoffii</i>	L-PHE (5, 0.58 <sup>d</sup> ) 2 L batch reactors with 10 g/L of glycerol	[130]
MG1655 derivative	(plasmid) P <sub>lac-UV5</sub> <i>aroE aroD aroB</i> <sup>pp</sup> , P <sub>L-tetO1</sub> <i>aroG</i> <sup>fbt</sup> <i>ppsA tktA</i> , (plasmid) P <sub>lac-UV5</sub> <i>tyrB tyrA</i> <sup>fbt</sup> <i>aroC</i> , P <sub>trc</sub> <i>aroA aroL</i>	L-TYR (2, 0.44 <sup>d</sup> ). Shake flask cultures with 5 g/L of glucose	[58]
rpoA14 <sup>R</sup> (K-12 derivative)	$\Delta$ pheA $\Delta$ tyrR / (chromosome) P <sub>L</sub> <i>tyrA</i> <sup>fbt</sup> <i>aroG</i> <sup>fbt</sup> , point mutations in <i>hisH</i> and <i>purF</i> , (plasmid) <i>rpoA</i>	L-TYR (14, 0.12 <sup>d</sup> ) 2 L fed-batch reactors with glucose	[101]
MG1655 derivative	$\Delta$ pheA $\Delta$ pheL / (chromosome) P <sub>trc</sub> <i>tyrA</i>	L-TYR (55, 0.30 <sup>d</sup> ). 200 L fed-batch reactors with glucose	[17]
FB-04/pSV03 (W3110 derivative)	$\Delta$ trpR $\Delta$ tnaA $\Delta$ pheA $\Delta$ tyrA / (plasmid) <i>aroF</i> <sup>fbt</sup> <i>trpE</i> <sup>fbt</sup> <i>D</i>	L-TRP (13, 0.10 <sup>d</sup> ). 3 L fed-batch reactors with glucose, 2 g/L of L-PHE and 3 g/L of L-TYR	[59]
GPT1017 (W3110 derivative)	$\Delta$ trpR $\Delta$ tnaA $\Delta$ ptsG $\Delta$ aroP $\Delta$ tnaB $\Delta$ mtr / (chromosome) swapping of tryptophan attenuator and <i>trp</i> promoter by 5CP <sub>tac</sub> , (plasmid) <i>aroG</i> <sup>fbt</sup> <i>trpE</i> <sup>fbt</sup> <i>tktA</i>	L-TRP (16). 5 L fed-batch reactors with glucose and 1 g/L of yeast extract	[80]
TRTH0709/pMEL03 (MG1655 derivative)	$\Delta$ trpR $\Delta$ tnaA $\Delta$ pta $\Delta$ mtr / (plasmid) <i>aroG</i> <sup>fbt</sup> <i>trpE</i> <sup>fbt</sup> <i>DCBA serA</i> , (plasmid) <i>tktA ppsA yddG</i>	L-TRP (49). 30 L fed-batch reactors with glucose and 1 g/L of yeast extract	[61]
Vio-4 (MG1655 derivative)	$\Delta$ trpR $\Delta$ tnaA $\Delta$ sdaA $\Delta$ lac $\Delta$ trpL $\Delta$ gal $\Delta$ xyl $\Delta$ fuc / (chromosome) P <sub>tac</sub> <i>aroF aroB aroL tktA serA</i> <sup>fbt</sup> <i>vioD</i> <sup>1</sup> , <i>trpE</i> <sup>fbt</sup> , (plasmid) <i>vioABC</i> <sup>E9</sup>	Violacein (0.7). 0.7 L fed-batch reactors with arabinose, 12 g/L of tryptone and 24 g/L of yeast extract	[82]
BKD5 (BW25113 derivative)	$\Delta$ ptsG $\Delta$ tyrR $\Delta$ pykA $\Delta$ pykF $\Delta$ pheA / (plasmid) P <sub>lac-UV5</sub> <i>aroG</i> <sup>fbt</sup> <i>tyrA</i> <sup>fbt</sup> <i>aroE</i> , P <sub>trc</sub> <i>ppsA tktA glk</i> , (plasmid) P <sub>lac-UV5</sub> T7 RNA polymerase, (plasmid) <i>hpaBC d-ldh</i> <sup>H</sup>	Salvianic acid A (7, 0.47 <sup>b</sup> ). 0.5 L fed-batch flasks with glucose and 1 g/L of yeast extract	[94]
QH23 (ATCC 31884 derivative)	$\Delta$ pheLA $\Delta$ tyrA / (plasmid) P <sub>L-lacO1</sub> <i>tyrA</i> <sup>fbt</sup> <i>ppsA tktA aroG</i> <sup>fbt</sup> , (plasmid) P <sub>L-lacO1</sub> <i>tal</i> <sup>cod</sup> <i>hpaBC</i>	Caffeic acid (0.8). Shake flask cultures with 2.5 g/L of glucose, 10 g/L of glycerol and phenylalanine	[105]
pAD-AG/ $\Delta$ tyrR (BL21 (DE3) derivative)	$\Delta$ tyrR / (plasmid) <i>aroG</i> <sup>fbt</sup> <i>tyrA</i> <sup>fbt</sup> , (plasmid) <i>tal</i>	4-coumaric acid (1). Shake flask cultures with 15 g/L of glucose	[103]
VH33 $\Delta$ tyrR_DOPA (W3110 derivative)	$\Delta$ ptsHlcr / (chromosome) P <sub>trc</sub> <i>galP</i> , (plasmid) <i>tktA</i> P <sub>lac-UV5</sub> <i>aroG</i> <sup>fbt</sup> , (plasmid) P <sub>trc</sub> <i>tyrC</i> <sup>c</sup> <i>pheA</i>	L-DOPA (1.5, 0.05 <sup>d</sup> ). 1 L batch reactors with LB and 50 g/L glucose	[84]
W3110 <i>trpD9923</i> /pS0 + pY + pAvnD	(plasmid) P <sub>L-tetO1</sub> <i>ydiB aroD aroB aroG</i> <sup>fbt</sup> <i>ppsA tktA</i> , (plasmid) P <sub>lac-UV5</sub> <i>tyrB tyrA</i> <sup>fbt</sup> <i>aroC aroA aroL</i> , (plasmid) P <sub>lac-UV5</sub> <i>HCBT</i> <sup>8</sup> <i>4CL1</i> <sup>1</sup> <i>tal</i>	Avenanthramide D (27 <sup>a</sup> ). Shake flask cultures with 10 g/L of glucose	[87]

<sup>a</sup>g/L; <sup>b</sup>mol substrate/mol product; <sup>c</sup>gene from *Z. mobilis*; <sup>d</sup>g substrate/g product; <sup>e</sup>μM; <sup>f</sup>gene from *J. lividum*; <sup>g</sup>genes from *C. violaceum*; <sup>h</sup>gene from *L. pentosus*; <sup>i</sup>gene from *R. glutinis*; <sup>j</sup>gene from *S. espanaensis*; <sup>k</sup>gene from *D. caryophyllus*; <sup>l</sup>gene from *N. tabacum*; <sup>cod</sup>codon-optimized variant.

mutase/prephenate dehydratase (PheA) also catalyzes the first step in the parallel biosynthetic pathways for L-TYR and L-PHE as well as the second step in L-PHE biosynthetic pathway (prephenate to phenylpyruvate) (Figure 1). The native enzyme is a dimer and each monomer contains a dehydratase active site, a mutase active site and an L-PHE binding site. PheA enzyme is inhibited by L-PHE (up to 90% of the prephenate dehydratase and 55% of the mutase activity) [1,2]. Feedback-resistant mutants of TyrA and PheA *E. coli* enzymes have been used for the efficient overproduction of L-TYR [11,17,57,67] and L-PHE [55,76] in combination with some of the previously described alterations in CCM and the SHK pathway (Table 1). An alternative approach to take advantage of the natural feedback-resistant diversity in the TyrA enzyme family was the expression of the TyrC<sup>fbr</sup> enzyme (cyclohexadienyl dehydrogenase) from *Z. mobilis* and the CHA mutase domain of native PheA from *E. coli*, relieving rate-limiting steps and increasing the carbon flux towards L-TYR [57].

A strategy to minimize carbon loss to competing pathways was exemplified in the CHA node with the construction of L-PHE production strains expressing TyrA enzymes containing tags for increased proteolytic degradation, instead of completely removing the enzyme. The resultant strains have the advantage of not being auxotrophic to L-TYR while displaying a higher L-PHE/L-TYR production ratio than the strain containing the wild-type TyrA [77].

Additional modifications applied in L-TRP overproducers include the overexpression of exporter protein YddG [61,78,79], the inactivation of permeases AroP, Mtr and TnaB to avoid re-internalization [61,79,80], the deletion of gene *tnaA* coding for a tryptophanase to avoid product degradation [59,81,82] and expression of genes included in the tryptophan biosynthetic branch, including a feedback-resistant version of anthranilate synthase, TrpE<sup>fbr</sup> [60,81].

Variations on the strategies described in this section have also been applied to the production of other valuable compounds derived from the AAA pathway such as phenyllactate, phenylacetate and phenylethanol [83], L-DOPA [84], mandelic acid [85], deoxyviolacein and violacein [82,86], avenanthramides [87], and resveratrol [88] (Table 2).

#### **Increasing the genetic engineering repertoire: development and application of synthetic biology strategies and techniques**

The field of synthetic biology has been continuously evolving and it is now acknowledged that this discipline is primarily concerned with the design and characterization of biological parts [89,90]. Indeed, modular and predictable parts find many applications in the modification of cellular metabolism, whether these alterations are direct

(modulation of the expression and function of enzymes comprised in metabolic pathways) or indirect (rewiring and repositioning of sensing components and cellular effectors). In this sense, the powerful recent advances in synthesis and assembly of macromolecules have changed the way to approach challenges in metabolic engineering. This has helped to generate a degree of biological diversity and reprogramming not previously reached with traditional biological controllers, promoting the merging of rational and combinatorial approaches to direct cellular design [91,92].

The aromatic biosynthetic pathway in *E. coli* was no exception to this paradigm shift, resulting in notable accomplishments over the last years. It is worth noting that even when the upregulation of a few genes can increase the carbon flux from CCM towards the aromatic biosynthetic pathway, the outcome is importantly influenced by a variety of factors, such as the combination of expression modules, genetic background and cultivation conditions. It is therefore ideal to design experiments to obtain a characterization of the contribution of each factor to the phenotype. Illustrative examples on this subject include the assessment of differences in the production of L-TYR by overexpressing various sets of genes in a stepwise approach (Table 1) [58,93].

The generation of synthetic parts in a faster, cheaper and more targeted way has also enabled metabolic engineers to reach unprecedented biochemical diversity, exemplified by the production of plant compounds using precursors present in the aromatic biosynthetic pathway in *E. coli*. In this way, combinations of simultaneous transcriptional modules and genetic platforms have resulted in strains with the ability to produce attractive compounds such as salvianic acid A [94],  $\delta$ -tocotrienol and its intermediate 2-methyl-6-geranylgeranyl-benzoquinol [95,96] (Figure 2) and (*S*)-reticuline [97] (Figure 3).

Aside from the product titers reached so far, these approaches are appealing because the systematic evaluation of conditions permits a more precise identification of targets for future improvement. In this respect, more structurally complex compounds can be produced by the optimization of expression parameters, for example when approaching problems with the heterologous insertion of genes and pathways into *E. coli*. One successful case concerning a systematic analysis of heterologous expression is the production of (2*S*)-pinocembrin from glucose as the only carbon source [98] (Figure 2). In this report, the authors assembled gene expression modules, including genes from the SHK pathway as well as heterologous sources. With this arrangement it was possible to accumulate up to 40 mg/L of (2*S*)-pinocembrin, even when using four plasmids and enzymes with naturally low catalytic efficiencies. The same system was used to evaluate the capabilities for resveratrol production after

**Table 2 Proposed applications of high-valued compounds derived from the aromatic pathway and synthesized by engineered *E. coli* strains**

Compound	Summary of pharmaceutical and industrial applications	References
Shikimate ((3R,4S,5R)-3,4,5-trihydroxycyclohexene-1-carboxylic acid)	Antipyretic, antioxidant, anticoagulant, antithrombotic, anti-inflammatory, and analgesic agent. Has a key role in the synthesis of important pharmacological compounds such as anti-cancer and antibacterial agents, as well as hormones. Substrate in the chemical synthesis of the antiviral Tamiflu®.	[72,74]
Salvianic acid or danshensu (3,4-dihydroxyphenyllactic acid)	A naturally occurring plant polyphenolic acid, considered as a superior antioxidant. Its scavenging activities against free hydroxyl radicals and superoxide anion radicals are higher than vitamin C. Has a variety of other pharmacological effects, including improving cerebral blood flow, inhibiting platelet activation and arterial thrombosis, as well as anti-cancer and anti-inflammatory effects.	[94]
(2S)-pinocembrin (5,7-dihydroxyflavanone)	Flavonoid with demonstrated activity decreasing the neurological scores, alleviating brain edema, reducing the permeability of blood-brain barrier and alleviating cerebral ischemic injury in the middle cerebral artery occlusion in rats. Has been proposed as a novel therapeutic agent to reduce cerebral ischemia/reperfusion and blood-brain injury, useful for its antioxidant and anti-apoptotic effects.	[98]
Caffeic acid (3,4-dihydroxycinnamic acid)	Possesses various pharmacological activities including antioxidant, antitumoral, antiviral, antidepressive and antidiabetic functions.	[104]
Resveratrol (3,4',5-trihydroxystilbene)	Potential therapeutic effects in humans as antioxidant, anti-inflammatory, anticancer, and chemopreventive agent.	[99]
Violacein ((3E)-3-[5-(5-hydroxy-1H-indol-3-yl)-2-oxo-1H-pyrrol-3-ylidene]-1H-indol-2-one) and deoxyviolacein	Activity against herpes simplex virus and pathogenic bacteria such as <i>Staphylococcus aureus</i> and <i>Pseudomonas aeruginosa</i> . Violacein has shown successful activity against leukemia, lung cancer, human uveal melanoma and lymphoma cells, where it mediates apoptosis. It is also an interesting bio-dye showing attractive color tone and stability.	[82]
PDC (2-pyrone-4,6-dicarboxylic acid)	Proposed as a novel starting material for several useful synthetic polymers such as polyesters and polyamides.	[100]
(S)-reticuline ((1S)-1-[(3-hydroxy-4-methoxyphenyl)methyl]-6-methoxy-2-methyl-3,4-dihydro-1H-isoquinolin-7-ol)	Building block for benzyloisoquinoline alkaloids, including the analgesic compounds morphine and codeine, as well as the antibacterial agents berberine and palmatine. Useful in the development of novel antimalarial and anticancer drugs.	[97]
Hydroxytyrosol (3,4-dihydroxyphenylethanol)	Powerful antioxidant activity. Potential antitumoral, antiatherogenic, anti-inflammatory and antiplatelet aggregation agent.	[106]
Avenanthramides	Natural hydroxycinnamoyl anthranilates with antioxidant, anti-inflammatory, and antiproliferative effects, considered to contribute to the health benefits of oatmeal consumption. Potential antitumor activities.	[87]
δ-tocotrienol	Vitamin E component naturally produced by photosynthetic organisms. It has shown to induce apoptosis and inhibit proliferation of cancer cells. Possess to some extent neuroprotective, anticancer, and cholesterol lowering properties.	[95,96]

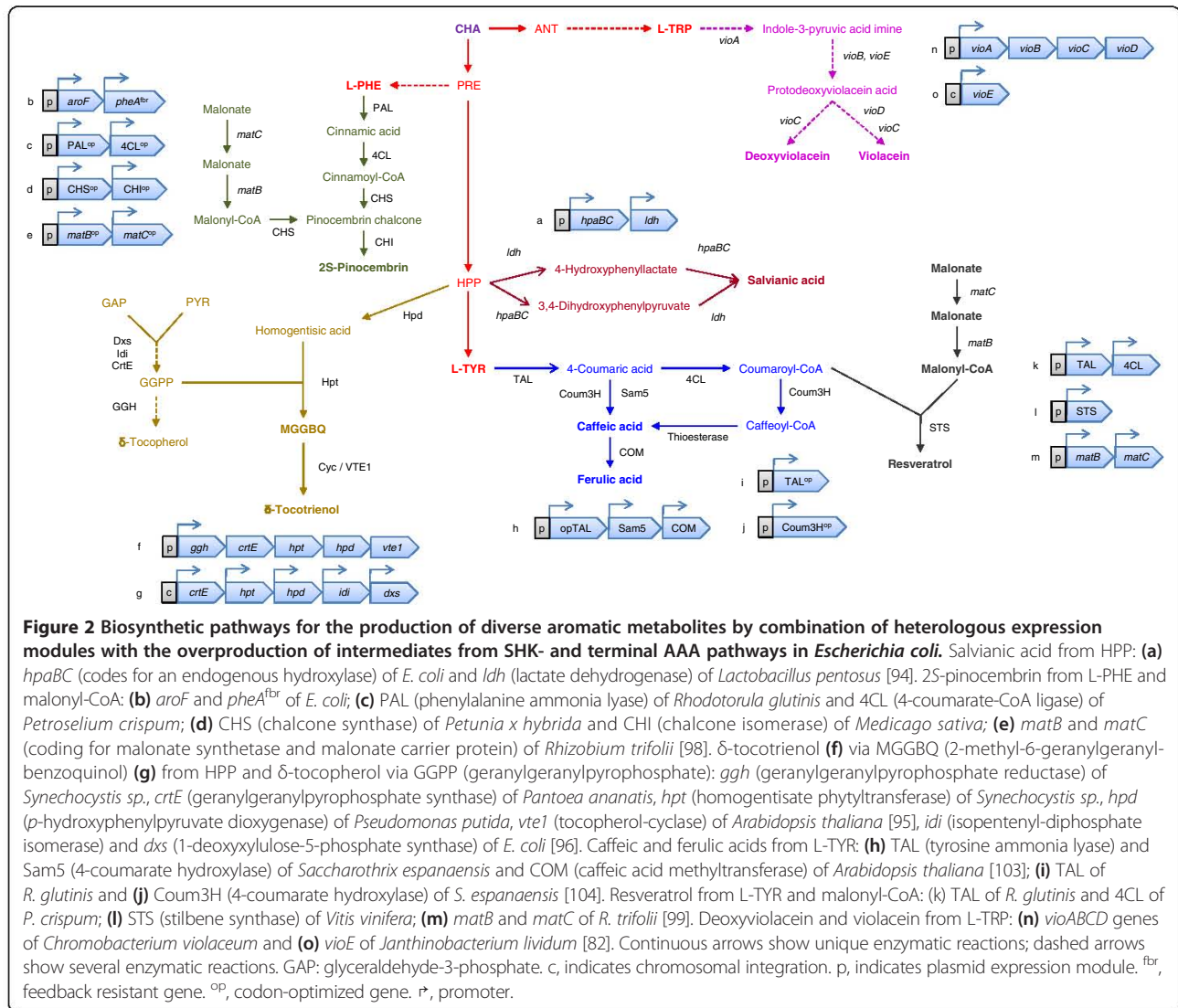
Recombinant pathways are presented in Figure 2 and Figure 3.

slight modifications were introduced to increase malonyl-CoA and L-TYR availability. This work revealed large variations in the concentration of produced metabolites with respect to small variations in the genetic constructs [99] (Figure 2). In a similar approach, the introduction of a foreign pathway succeeded in deviating carbon flow from 3-dehydroshikimate towards the synthesis of 2-pyrone-4,6-dicarboxylic acid (PDC) (Figure 3). Strains overexpressing six different genes from three different plasmids were able to produce the desired compound with a 17.3% yield from glucose [100].

It is also interesting to consider other strategies to generate and screen metabolic diversity, such as modifications of the global transcription machinery coupled to high-throughput screening for metabolite production

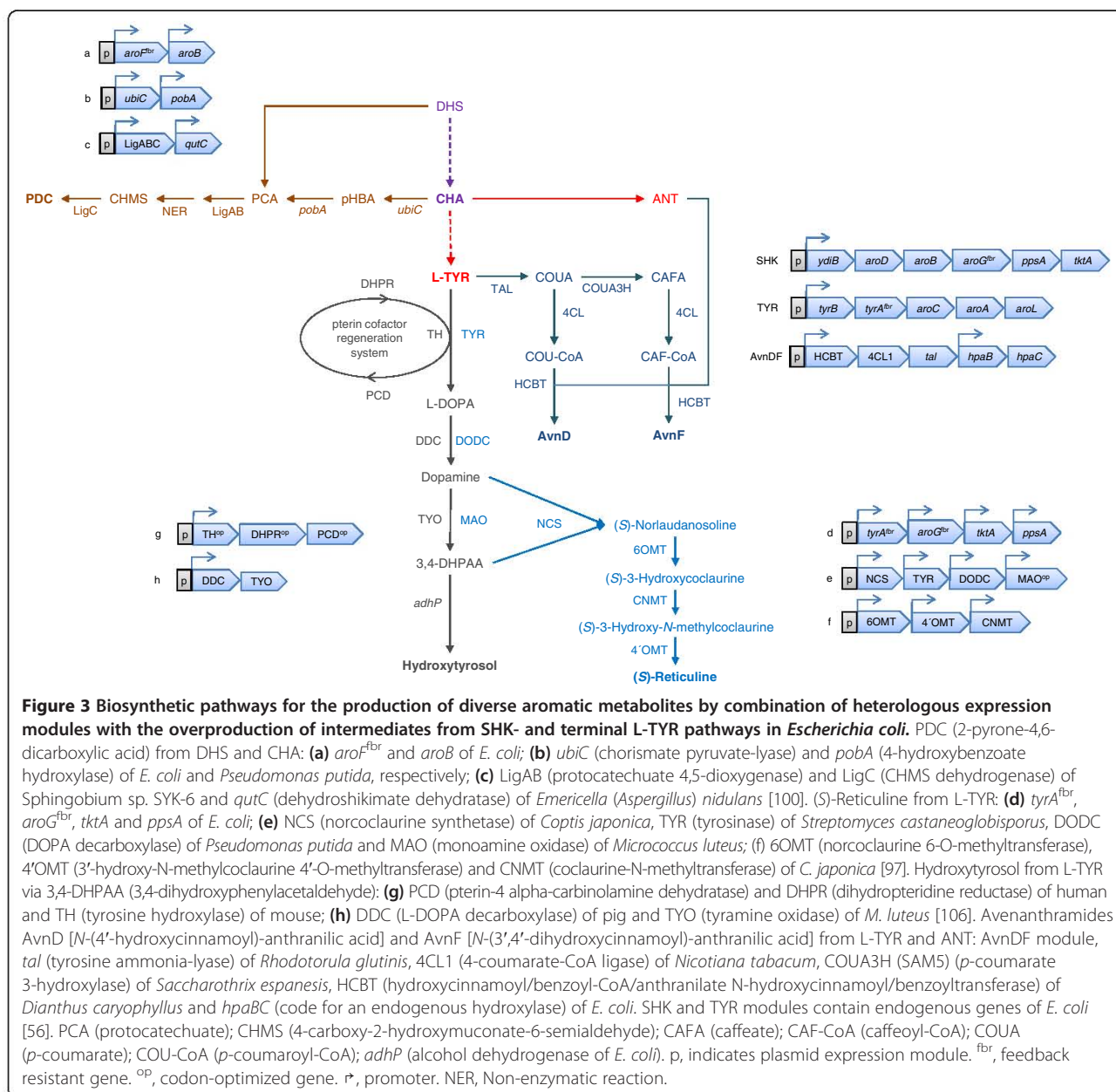
[101,102]. By merging these approaches with combinatorial techniques for gene overexpression, a 114% increase in L-TYR production from a previously engineered strain was reported [101].

Strains with the capability to overproduce L-TYR from simple carbon sources have been used as a backbone for production of more structurally complex compounds. For example, the construction of codon-optimized heterologous gene clusters with a wide span of strengths in promoter and ribosome binding sequences (RBS) has allowed the generation of *E. coli* strains capable of producing phenylpropanoic acids such as caffeic acid, coumaric acid and ferulic acid [103-105] (Figure 2), as well as hydroxycinnamoyl anthranilates [87] and other derivatives, such as hydroxytyrosol [106] (Figure 3). Another combinatorial



technique applied in the generation and isolation of strains with an increased production of indigo (a compound that can be obtained from the L-TRP biosynthetic intermediate indole) is coselection MAGE (multiplex automated genome engineering). This method relies on a cyclical oligo-mediated allelic replacement to modify genomic targets [107] that was later improved by linking the process with the recovery of an inactivated selection marker, enhancing the size and efficiency of insertions [108]. With this approach, the authors were able to insert T7 promoters upstream of 12 genes or operons associated with the AAA pathway in a strain modified to produce indigo and recovered 80 unique derivatives with variable promoter insertions. As a result, it was possible to identify strains with more than a fourfold improvement in indigo production over the ancestor strain, as well as synergetic interactions of expressed genes [108].

The application of synthetic RNA devices with the goal of increasing AAA production in *E. coli* has recently attracted attention. In particular, artificial riboswitches coupling the binding of L-TRP to growth under a selective pressure have been constructed and tested *in vivo*. By modulating the expression of gene *aroG* under this scheme, strains with superior capabilities for L-TRP production could be linked to the increased growth rates after rounds of selective improvement [109]. In another report, a synthetic sRNA library was constructed for targeted gene expression silencing. The authors demonstrated the applicability of this approach in the production of L-TYR with the plasmid-based expression of genes *ppsA*, *tktA*, *aroF*, *aroK*, *tyrC*, *aroG* and *tyrA*, and the simultaneous silencing of genes *tyrR*, *csrA*, *pgi* and *ppc* in several *E. coli* strains. With an easily transferrable gene-regulation platform, the combination of expression levels and genetic backgrounds led to the selection of a strain that can



accumulate up to 21.9 g/L of L-TYR in high-density cultures [110].

### Integration and application of data: systems-based approaches to the production of AAA

Even with the relative success obtained so far regarding the overproduction of aromatic compounds in *E. coli*, insights into the global metabolic state of engineered strains under production conditions are still scarce. Moreover, the effects of targeted strain modifications are typically underestimated, since they do not always result in significant differences in cell growth or production of specific metabolites. Combination of techniques such as

genomics, transcriptomics, proteomics, metabolomics and fluxomics can unravel the particular cellular state during a defined condition by providing snapshots of different levels of metabolism [111]. However, in order to turn this information into knowledge of new potential engineering targets, adequate comparisons must be established. Since it is not trivial to define the type and extent of data to be extracted and compared, systems biology approaches are needed to manage holistic information at different levels of cellular functions [112,113].

Although the systematic integration of -omics approaches have been applied to characterize and reverse engineer bacterial strains producing several amino acids [6,114-117],



there are still relatively few reports on the use of these techniques with AAA overproducers. For example, one study reports the effect of inactivating genes coding for PEP-consuming enzymes (PTS, PykF and PykA) over the flux distributions in the central carbon metabolism as an attempt to increase the availability of this AAA precursor [33]. The net result of either inactivation was a flux increase to biomass formation pathways, but several differences on important CCM nodes were also found between all conditions. Furthermore, PTS inactivation revealed a carbon recycling response between PEP and OAA combined with a reduced glycolytic flux. When these strains were transformed with plasmids encoding enzymes to promote the production of L-PHE, a 19-, 14-, and 25- fold increase on the yield of this amino acid was observed for the PTS, PTS<sup>-</sup> *pykA*, and PTS<sup>-</sup> *pykF* mutants, respectively [33].

Targeted proteomics and metabolite profiling analyses are also very valuable to provide feedback about expression systems used in the production of AAA. One report describes such approaches on a collection of L-TYR producing strains with different gene-expression arrays, allowing the authors to identify and improve sub-optimally expressed genes. After a second engineering round of the synthetic expression modules a strain was constructed which can produce L-TYR from glucose with 80% of the theoretical yield, estimated as 0.55 g/g in strains with a functional PTS [58]. A related work characterized the impact on SHK pathway enzyme levels resulting from the removal of TyrR regulator, along with the use of a feedback-resistant TyrA and deletion of the *pheA* gene on L-TYR producing strains. The results showed that small changes in protein levels caused by the genetic alterations can have a big impact on metabolite production, as a 250-fold span of L-TYR concentrations were detected [118]. A different work found many proteins differentially expressed as a response to the sole inactivation of the *pykF* gene, including DAHP synthase (AroG), SHK dehydrogenase (AroE), SHK kinase I (AroK), CHA synthase (AroC), prephenate dehydratase (PheA), anthranilate synthase (TrpD, TrpE) and L-TRP synthase (TrpA), as compared to the wild type strain [119].

In another example, transcriptional analysis and whole genome sequencing studies were performed on L-TYR producing strains obtained by combinatorial and targeted approaches, coupled to high-throughput screening, in an attempt to discover the changes that led to higher L-TYR production [101]. The transcriptional analysis revealed upregulation of genes related to acid stress resistance and global reductions in the expression of several pathways such as ribosomal protein and RNA formation, fatty acid elongation, de novo purine/pyrimidine biosynthesis and DNA replication, which imply a cellular shift from proliferation and growth to maintenance and stress

survival. Genomic analyses revealed differential single base-pair changes between the studied strains. When these mutations were reintroduced on a parental strain background higher L-TYR production was observed, showing their contributions to the overproduction phenotype. Finally, a reverse engineered strain was constructed, which gave a titer of 902 mg/L and an L-TYR yield on glucose of 0.18 g/g on a genetically-defined background [101]. Other works have also characterized the global transcriptional response to the presence of high levels of L-PHE or SHK in simple and complex media [120,121] or starvation conditions [122], revealing metabolic information that can be used for further improvement of the strains and cultivation conditions.

Along with data obtained by high-throughput systems, modeling of metabolism by mathematical approaches has become an important tool for analyzing cell responses and unravel the metabolic regulation between the cell information/control systems [111]. Moreover, genome-scale models of metabolism have been analyzed by constraint-based approaches [123]. Gene deletion effects over flux distributions have also been studied in order to find the combination that provides the best metabolic performance on a given condition. For example, the deletion impact of 1261 genes was modeled using a reconstruction of biochemical interactions, resulting in 195 genes exerting high impact on flux distributions in various metabolic subsystems [124]. A strategy developed to circumvent the need for kinetic parameters of enzymes present in a metabolic network is ensemble modeling, which uses phenotypic data obtained from overexpression and deletion of enzymes to screen out flux distributions from an initial ensemble of solutions derived from elementary reaction models [125]. This method has been used to model the AAA pathway for DAHP production with data obtained from the overexpression of CCM genes. A subset of flux distributions was found capable of describing the phenotypic characteristics of the strains and rendered information about the kinetic and stoichiometric limitations around PEP and E4P nodes [125]. As more genomic, transcriptomic and proteomic functional interactions continue to be unraveled, similar approaches will become powerful tools to model specific metabolic outcomes related to AAA biosynthesis.

#### **Bioprocess engineering: optimization of AAA compound production**

In order to create economically viable products, the processes developed and tested at laboratory scale have to be adapted to larger operational volumes. Although engineered strains should ideally perform equally in 1 L scale as in industrial scales (going from 500–10,000 L for fine chemicals to more than 100,000 L for commodity chemicals), a significant reduction in performance as a result of scale-up is often observed [126]. Therefore, it

is important to apply strategies to prevent physiological changes caused by heterogeneities of fermentation parameters during scale-up processes with *E. coli*. Stress-mediated cellular responses to chemical and physical factors can negatively impact as much as 60% over the productivity, the biomass and product yields when a strain is exposed to large-scale production conditions [127,128].

Fed-batch cultivations have been a popular method to produce aromatic compounds since they promote high cellular densities, offer tight control over the  $\mu$  and substrate concentrations, and permit a better management of dissolved oxygen tension (DOT) to prevent the activation of fermentative pathways [16,129]. In one example, a fed-batch strategy improved violacein production from arabinose (through an expanded pathway from L-TRP) by adjusting the  $\mu$  at  $0.011 \text{ h}^{-1}$  [82]. With this procedure, cellular concentrations with optical density values up to 70 were reached, producing 710 mg/L of violacein and avoiding acetate accumulation in the medium, a known inhibitor of growth and pigment production.

Another work studied the impact of different feeding strategies over the production of L-TRP in a recombinant strain [16]. An increase in the volumetric productivity of this compound was reached by a novel feeding strategy with a highly concentrated glucose solution (800 g/L) after the exhaustion of the initial glucose. By using a combined pseudo-exponential feeding at the exponential phase and a glucose-stat feeding after the exponential phase, an efficient control over the  $\mu$  was achieved (below the acetic acid production threshold), reaching 38.8 g/L of L-TRP. This represented an increase of 19.9% due to reduced acetic acid accumulation [16].

Even if feeding strategies can cope to some extent with the problems derived from acetic acid production, a combination of these with genetic modifications has also been tested for the production of aromatic compounds. In a recent report, the effect of inactivating the gene coding for the enzyme phosphotransacetylase (Pta) over the production of L-TRP was assessed. By combining this modification with the use of a DO-stat for controlling inflow rate at a suitable DOT, the authors were able to increase the production of L-TRP and biomass while maintaining the growth rate and reducing the accumulation of acetic acid [129].

Substrate characteristics can also be optimization variables for the production of AAA pathway intermediates. One example is the evaluation of glycerol for L-PHE production [130,131]. The low cost of glycerol coupled with its higher degree of carbon reduction when compared to other sugars such as glucose, could result in high energy yield per carbon and hence be advantageous for AAA production processes. However, it is important to characterize the influence of fermentation parameters such as DOT, temperature and pH, as well as

the availability of substrates, over the growth and product formation rates. In one report, variations in oxygen supply (by changing aeration rates and impeller speeds) were tested over the L-PHE production capabilities of a recombinant strain growing on glycerol [130]. With this approach, a direct correlation between biomass and L-PHE production rates were found at impeller speeds up to 400 rpm, being this the maximum operational value before shear stress starts to diminish strain capabilities. After setting the impeller speed to 400 rpm, aeration optimization resulted in the highest product yield obtained, 0.58 g/g, which is 20% higher than the yields obtained before optimization of oxygen supply. Interestingly, the authors report this high yield with a strain in which the only recombinant measure taken is the heterologous expression of a phenylalanine dehydrogenase gene [130]. On the other hand, another group has recently reported the production of 13 g/L of L-PHE from glycerol and a yield of 0.15 g/g using a multi-phase fed-batch process with a strain containing several genetic modifications [131] (Table 1).

Product characteristics should also be taken into account when developing an efficient bioprocess. For example, L-TYR exhibits low solubility in typical fermentation conditions, triggering its precipitation when saturation is reached. This characteristic would normally be beneficial for a fermentation process, as a precipitated compound can be easily recovered and it is not expected to affect strain physiology and production capabilities. Interestingly, one report described that the L-TYR crystals can stabilize foam, causing operational problems during fermentation [17]. Consequently, this foaming process was studied on 10 and 200 L fed-batch fermentations for L-TYR production to assess the effect of pH, antifoam concentration, cooling rate, L-PHE concentration and seeding level on foam production. It was determined that high concentrations of L-PHE or antifoam, as well as low pH and low seeding, are the preferred conditions to avoid detrimental foaming production. With this approach it was possible to produce L-TYR from glucose with a yield of 0.3 g/g and titers as high as 55 g/L on a 200 L scale [17]. Moreover, this study revealed important data for the design of an economically feasible process for the production of L-TYR.

Process optimization could also be concerned with an enhancement of the strain ability to withstand high concentrations of aromatic compounds, not only for toxic final products but for harmful intermediates or byproducts, which often accumulate as a consequence of the suboptimal alleviation of control levels in the biosynthetic pathways. This is a commonly-encountered problem with many of the intermediates and final products in the AAA. For example, one group reported the optimization of L-TRP production by modifying the export

and import capabilities of a modified strain in order to minimize its intracellular concentration and avoid feedback control by product accumulation [79]. This group constructed a strain featuring the plasmid-based overexpression of the AAA exporter YddG, resulting in a production increase of 12.6% compared to the parental strain on a 30 L fermentation. Another example of the successful combination of genetic and fermentation procedures involves the construction of a strain for L-PHE overproduction with a PTS-independent glucose transport and expression of feedback-resistant versions of AroG and PheA. By overexpressing genes *ydiB*, *aroK* and *tyrB* with a temperature-dependent system, as well as *yddG* in a TyrA<sup>-</sup> background, the authors were able to produce up to 47 g/L of L-PHE with a yield of 0.25 g/g from glucose in a 15 L fed-batch process [132].

Finally, bioprocess design is also an important factor in optimization of the production of aromatic pathway derivatives. The bioconversion of phenylpyruvate (PPN) to L-PHE was studied with an immobilized cell bioprocess [133]. This technique has several advantages such as the ability to reuse the immobilized cells, the capacity to utilize high cell densities and improved stability of the system. For example, a mixed-gel surface composed of k-carrageenan and gelatin, together with the optimization of its composition to enhance the mechanical strength and reduce the toxicity and solidification point was used as biomass carrier for the production of L-PHE [133]. Studies on the effect of pH, temperature, Mg<sup>+2</sup> and trehalose presence resulted in the implementation of a process showing an improvement of 80% on the L-PHE conversion from PPN after 15 successive batch experiments.

## Conclusions

The present review aims to provide a panorama of the current achievements and newly found goals related to the production of aromatic compounds in *E. coli*. The AAA pathway and the metabolic changes resulting from its deregulation have attracted the interest of metabolic engineers for many years and remain important research targets on several organisms. It is evident that the establishment of efficient bioprocesses on this topic requires the design and implementation of multidisciplinary strategies, taking advantage of the fast-paced developments coming from nearly all biotechnological fields but particularly from those related with information technologies, such as systems and synthetic biology. The works compiled here are a good example of the benefits obtained when new ideas and viewpoints are introduced to an established field in order to cope with long-known problems. From the comparisons presented, it is noticeable that the use of rational and combinatorial approaches powered by the ability to develop complex genetic circuits and high-throughput screenings of new producers has set new trends

when dealing with the production of aromatic compounds in *E. coli*. The benefits of the integral application of these technologies can already be observed, not only from the improved production processes for AAA and pathway intermediates with large and established markets, but also with the generation of novel derivative compounds with important pharmaceutical applications.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

All authors participated in the preparation of this contribution. AR had a major role in writing and editing the manuscript. All authors have read and approved the final version.

## Acknowledgements

This work was supported by CONACYT 105782, 177568 and DGAPA-PAPIIT UNAM IN206812 grants.

Received: 30 May 2014 Accepted: 17 August 2014

Published online: 09 September 2014

## References

1. Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martinez C, Fulcher C, Huerta AM, Kothari A, Krummenacker M, Latendresse M, Muniz-Rascado L, Ong Q, Paley S, Schroder I, Shearer AG, Subhraveti P, Travers M, Weerasinghe D, Weiss V, Collado-Vides J, Gunsalus RP, Paulsen I, Karp PD: **EcoCyc: fusing model organism databases with systems biology.** *Nucleic Acids Res* 2013, **41**:D605–D612.
2. Sprenger G: **Aromatic Amino Acids.** In *Amin Acid Biosynth - Pathways, Regul Metab Eng.* Edited by Wendisch VF. Berlin, Heidelberg: Springer; 2007:418 [Microbiology Monographs, vol. 5].
3. Herrmann KM, Weaver LM: **The shikimate pathway.** *Annu Rev Plant Biol* 1999, **50**:473–503.
4. Richards TA, Dacks JB, Campbell SA, Blanchard JL, Foster PG, McLeod R, Roberts CW: **Evolutionary origins of the eukaryotic shikimate pathway: gene fusions, horizontal gene transfer, and endosymbiotic replacements.** *Eukaryot Cell* 2006, **5**:1517–1531.
5. Lütke-Eversloh T, Santos CNS, Stephanopoulos G: **Perspectives of biotechnological production of L-tyrosine and its applications.** *Appl Microbiol Biotechnol* 2007, **77**:751–762.
6. Becker J, Wittmann C: **Systems and synthetic metabolic engineering for amino acid production - the heartbeat of industrial strain development.** *Curr Opin Biotechnol* 2012, **23**:718–726.
7. Ikeda M, Takeno S: **Amino acid production by *Corynebacterium glutamicum*.** In *Corynebacterium glutamicum. Volume 23.* Edited by Yukawa H, Inui M. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013:107–147.
8. Ajinomoto Co. Inc: *FY2013 Market and Other Information.* 2014.
9. Li Z, Ji X, Kan S, Qiao H, Jian M, Lu D, Wang J, Huang H, Jia H, Ouyang P, Ying H: **Past, Present and Future Industrial Biotechnology in China.** In *Biotechnol China II Chem Energy Environ.* Edited by Tsao GT, Ouyang P, Chen J. Berlin, Heidelberg: Springer; 2010:1–42.
10. 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.
11. Sprenger GA: **From scratch to value: Engineering *Escherichia coli* wild type cells to the production of L-phenylalanine and other fine chemicals derived from chorismate.** *Appl Microbiol Biotechnol* 2007, **75**:739–749.
12. Pittard J, Yang J: **Biosynthesis of the Aromatic Amino Acids.** *Eco Sal Plus* 2008, **1**.
13. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, Palsson BØ: **A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism-2011.** *Mol Syst Biol* 2011, **7**.
14. Klein-Marcuschamer D, Santos CNS, Yu H, Stephanopoulos G: **Mutagenesis of the bacterial RNA polymerase alpha subunit for improvement of complex phenotypes.** *Appl Environ Microbiol* 2009, **75**:2705–2711.
15. Jeong J, Cho N, Jung D, Bang D: **Genome-scale genetic engineering in *Escherichia coli*.** *Biotechnol Adv* 2013, **31**:804–810.

16. Cheng L-K, Wang J, Xu Q-Y, Xie X-X, Zhang Y-J, Zhao C-G, Chen N: **Effect of feeding strategy on L-tryptophan production by recombinant *Escherichia coli***. *Ann Microbiol* 2012, **62**:1625–1634.
17. Patnaik R, Zolanz RR, Green DA, Kraynie DF: **L-Tyrosine production by recombinant *Escherichia coli* : Fermentation optimization and recovery**. *Biotechnol Bioeng* 2008, **99**:741–752.
18. Ikeda M: **Amino acid production processes**. In *Microb Prod L-amino acids*. Berlin, Heidelberg: Springer; 2003:1–35.
19. Gosset G: **Production of aromatic compounds in bacteria**. *Curr Opin Biotechnol* 2009, **20**:651–658.
20. Westerhoff HW, Palsson BO: **The evolution of molecular biology into systems biology**. *Nat Biotechnol* 2004, **22**:1249–1252.
21. Blazcek J, Alper H: **Systems metabolic engineering: Genome-scale models and beyond**. *Biotechnol J* 2010, **5**:647–659.
22. Postma PW, Lengeler JW, Jacobson GR: **Phosphoenolpyruvate: carbohydrate phosphotransferase systems of bacteria**. *Microbiol Rev* 1993, **57**:543–594.
23. Gosset G: **Improvement of *Escherichia coli* production strains by modification of the phosphoenolpyruvate: sugar phosphotransferase system**. *Microb Cell Fact* 2005, **4**:14.
24. Escalante A, Salinas-Cervantes A, Gosset G, Bolivar F: **Current knowledge of the *Escherichia coli* phosphoenolpyruvate-carbohydrate phosphotransferase system: peculiarities of regulation and impact on growth and product formation**. *Appl Microbiol Biotechnol* 2012, **94**:1483–1494.
25. Sauer U, Eikmanns BJ: **The PEP – pyruvate – oxaloacetate node as the switch point for carbon flux distribution in bacteria**. *FEMS Microbiol Rev* 2005, **29**:765–794.
26. Sprenger G, Siewe R, Martin K, Sonke T: *Microbial Preparation of Substances from Aromatic Metabolism/I. WO patent 98/18936*. 1998.
27. Frost JW, Knop DR: *Biocatalytic Synthesis of Shikimic Acid. US patent 6,472,169 B1*. 2002.
28. 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**:808–814.
29. Yi J, Draths KM, Li K, Frost JW: **Altered glucose transport and shikimate pathway product yields in *E. coli***. *Biotechnol Prog* 2003, **19**:1450–1459.
30. Balderas-Hernandez VE, Sabido-Ramos A, Silva P, Cabrera-Valladares N, Hernandez-Chavez G, Baez-Viveros JL, Martinez A, Bolivar F, Gosset G: **Metabolic engineering for improving anthranilate synthesis from glucose in *Escherichia coli***. *Microb Cell Fact* 2009, **8**:19.
31. Flores N, Xiao J, Bolivar F, Valle F: **Pathway engineering for the production of aromatic compounds in *Escherichia coli***. *Nat Biotechnol* 1996, **14**:620–623.
32. Valle F, Mejia N, Berry A: *Application of Glucose Transport Mutants for Production of Aromatic Pathway Compounds. WO patent 96/34961*. 1996.
33. Meza E, Becker J, Bolivar F, Gosset G, Wittmann C: **Consequences of phosphoenolpyruvate: sugar phosphotransferase system and pyruvate kinase isozymes inactivation in central carbon metabolism flux distribution in *Escherichia coli***. *Microb Cell Fact* 2012, **11**:127.
34. Sabido A, Sigala JC, Hernández-Chávez G, Flores N, Gosset G, Bolivar F: **Physiological and transcriptional characterization of *Escherichia coli* strains lacking interconversion of phosphoenolpyruvate and pyruvate when glucose and acetate are coutilized**. *Biotechnol Bioeng* 2014, **111**:1150–1160.
35. Patnaik R, Liao JC: **Engineering of *Escherichia coli* central metabolism for aromatic metabolite production with near theoretical yield**. *Appl Environ Microbiol* 1994, **60**:3903–3908.
36. Liao JC: *Microorganisms and Methods for Overproduction of DAHP by Cloned pps Gene. WO patent 96/08567*. 1996.
37. 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**:1141–1148.
38. Liao JC, Hou SY, Chao YP: **Pathway analysis, engineering, and physiological considerations for redirecting central metabolism**. *Biotechnol Bioeng* 1996, **52**:129–140.
39. Gulevich AY, Biryukova IV, Zimenkov DV, Skorokhodova AY, Kivero AD, Belareva AV, Mashko SV: *Method for Producing An L-amino Acid Using A Bacterium Having Enhanced Expression of the pckA Gene. US Patent 2006/0035348 A1*. 2006.
40. Tatarko M, Romeo T: **Disruption of a global regulatory gene to enhance central carbon flux into phenylalanine biosynthesis in *Escherichia coli***. *Curr Microbiol* 2001, **43**:26–32.
41. Yakandawala N, Romeo T, Friesen AD, Madhyastha S: **Metabolic engineering of *Escherichia coli* to enhance phenylalanine production**. *Appl Microbiol Biotechnol* 2008, **78**:283–291.
42. Zhao G, Winkler ME: **An *Escherichia coli* K-12 tktA tktB mutant deficient in transketolase activity requires pyridoxine (vitamin B6) as well as the aromatic amino acids and vitamins for growth**. *J Bacteriol* 1994, **176**:6134–6138.
43. Nakahigashi K, Toya Y, Ishii N, Soga T, Hasegawa M, Watanabe H, Takai Y, Honma M, Mori H, Tomita M: **Systematic phenome analysis of *Escherichia coli* multiple-knockout mutants reveals hidden reactions in central carbon metabolism**. *Mol Syst Biol* 2009, **5**:306.
44. Draths KM, Pompliano DL, Conley DL, Frost JW, Berry A, Disbrow GL, Staversky RJ, Lievens JC: **Biocatalytic synthesis of aromatics from D-Glucose: the role of transketolase**. *J Am Chem Soc* 1992, **114**:3956–3962.
45. Patnaik R, Spitzer RG, Liao JC: **Pathway engineering for production of aromatics in *Escherichia coli*: confirmation of stoichiometric analysis by independent modulation of AroG, TktA, and Pps activities**. *Biotechnol Bioeng* 1995, **46**:361–370.
46. Báez JL, Bolivar F, Gosset G: **Determination of 3-deoxy-D-arabino-heptulosonate 7-phosphate productivity and yield from glucose in *Escherichia coli* devoid of the glucose phosphotransferase transport system**. *Biotechnol Bioeng* 2001, **73**:530–535.
47. Lu J, Liao JC: **Metabolic engineering and control analysis for production of aromatics: role of transaldolase**. *Biotechnol Bioeng* 1997, **53**:132–138.
48. Mascarenhas D, Ashworth DJ, Chen CS: **Deletion of *pgi* alters tryptophan biosynthesis in a genetically engineered strain of *Escherichia coli***. *Appl Environ Microbiol* 1991, **57**:2995–2999.
49. Ahn J, Chung BKS, Lee D, Park M, Karimi IA, Jung J, Lee H: **NADPH-dependent *pgi*-gene knockout *Escherichia coli* metabolism producing shikimate on different carbon sources**. *FEMS Microbiol Lett* 2011, **324**:10–16.
50. 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 Fact* 2013, **12**:86.
51. Li K, Frost JW: **Microbial synthesis of 3-dehydroshikimic acid: a comparative analysis of D-xylose, L-arabinose, and D-glucose carbon sources**. *Biotechnol Prog* 1999, **15**:876–883.
52. Martínez K, De Anda R, Hernández G, Escalante A, Gosset G, Ramírez OT, Bolivar 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 Fact* 2008, **7**:1.
53. Ahn JO, Lee HW, Saha R, Park MS, Jung JK, Lee DY: **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**:1773–1784.
54. Chen K, Dou J, Tang S, Yang Y, Wang H, Fang H, Zhou C: **Deletion of the *aroK* gene is essential for high shikimic acid accumulation through the shikimate pathway in *E. coli***. *Bioresour Technol* 2012, **119**:141–147.
55. Báez-Viveros JL, Osuna J, Hernández-Chávez G, Soberón X, Bolivar F, Gosset G: **Metabolic engineering and protein directed evolution increase the yield of L-phenylalanine synthesized from glucose in *Escherichia coli***. *Biotechnol Bioeng* 2004, **87**:516–524.
56. Liu S-P, Xiao M-R, Zhang L, Xu J, Ding Z-Y, Gu Z-H, Shi G-Y: **Production of L-phenylalanine from glucose by metabolic engineering of wild type *Escherichia coli* W3110**. *Process Biochem* 2013, **48**:413–419.
57. Chavez-Bejar MI, Lara AR, Lopez H, Hernandez-Chavez G, Martinez A, Ramirez OT, Bolivar F, Gosset G: **Metabolic engineering of *Escherichia coli* for L-tyrosine production by expression of genes coding for the chorismate mutase domain of the native chorismate mutase-prephenate dehydratase and a cyclohexadienyl dehydrogenase from *Zymomonas mobilis***. *Appl Environ Microbiol* 2008, **74**:3284–3290.
58. Juminaga D, Baidoo EEK, Redding-Johanson AM, Batth TS, Burd H, Mukhopadhyay A, Petzold CJ, Keasling JD: **Modular engineering of L-tyrosine production in *Escherichia coli***. *Appl Environ Microbiol* 2012, **78**:89–98.
59. Zhao Z-J, Zou C, Zhu Y-X, Dai J, Chen S, Wu D, Wu J, Chen J: **Development of L-tryptophan production strains by defined genetic modification in *Escherichia coli***. *J Ind Microbiol Biotechnol* 2011, **38**:1921–1929.
60. Shen T, Liu Q, Xie X, Xu Q, Chen N: **Improved production of tryptophan in genetically engineered *Escherichia coli* with TktA and PpsA overexpression**. *J Biomed Biotechnol* 2012, **2012**:1–8.

61. Wang J, Cheng L-K, Wang J, Liu Q, Shen T, Chen N: **Genetic engineering of *Escherichia coli* to enhance production of L-tryptophan.** *Appl Microbiol Biotechnol* 2013, **97**:7587–7596.
62. Doroshenko VG, Tsyrenzhapova IS, Krylov AA, Kiseleva EM, Ermishev VY, Kazakova SM, Biryukova IV, Mashko SV: **Pho regulon promoter-mediated transcription of the key pathway gene *aroGFbr* improves the performance of an L-phenylalanine-producing *Escherichia coli* strain.** *Appl Microbiol Biotechnol* 2010, **88**:1287–1295.
63. Dell KA, Frost JW: **Identification and removal of impediments to biocatalytic synthesis of aromatics from D-Glucose: rate-limiting enzymes in the common pathway of aromatic amino acid biosynthesis.** *J Am Chem Soc* 1993, **115**:11581–11589.
64. Krämer M, Bongaerts J, Bovenberg R, Kremer S, Müller U, Orf S, Wubbolts M, Raeven L: **Metabolic engineering for microbial production of shikimic acid.** *Metab Eng* 2003, **5**:277–283.
65. Oldiges M, Kunze M, Degenring D, Sprenger GA, Takors R: **Stimulation, monitoring, and analysis of pathway dynamics by metabolic profiling in the aromatic amino acid pathway.** *Biotechnol Prog* 2004, **20**:1623–1633.
66. Escalante A, Calderón R, Valdivia A, De Anda R, Hernández G, Ramírez OT, Gosset G, Bolívar F: **Metabolic engineering for the production of shikimic acid in an evolved *Escherichia coli* strain lacking the phosphoenolpyruvate: carbohydrate phosphotransferase system.** *Microb Cell Fact* 2010, **9**:21.
67. Lutke-Eversloh T, Stephanopoulos G: **Feedback inhibition of chorismate mutase/prephenate dehydrogenase (*TyrA*) of *Escherichia coli*: Generation and characterization of tyrosine-insensitive mutants.** *Appl Environ Microbiol* 2005, **71**:7224–7228.
68. 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 Fact* 2014, **13**:21.
69. Johansson L, Lindskog A, Silfversparre G, Cimander C, Nielsen KF, Lidén G: **Shikimic acid production by a modified strain of *E. coli* (W3110.shik1) under phosphate-limited and carbon-limited conditions.** *Biotechnol Bioeng* 2005, **92**:541–552.
70. 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.
71. Ghosh S, Chisti Y, Banerjee UC: **Production of shikimic acid.** *Biotechnol Adv* 2012, **30**:1425–1431.
72. Rawat G, Tripathi P, Saxena RK: **Expanding horizons of shikimic acid: Recent progresses in production and its endless frontiers in application and market trends.** *Appl Microbiol Biotechnol* 2013, **97**:4277–4287.
73. Tripathi P, Rawat G, Yadav S, Saxena RK: **Fermentative production of shikimic acid: a paradigm shift of production concept from plant route to microbial route.** *Bioprocess Biosyst Eng* 2013, **36**:1665–1673.
74. Estevez A, Estevez R: **A short overview on the medicinal chemistry of (–)-shikimic acid.** *Mini Rev Med Chem* 2012, **12**:1443–1454.
75. Salgado H, Peralta-Gil M, Gama-Castro S, Santos-Zavaleta A, Muñoz-Rascado L, García-Sotelo JS, Weiss V, Solano-Lira H, Martínez-Flores I, Medina-Rivera A, Salgado-Orsorio G, Alquicira-Hernández S, Alquicira-Hernández K, López-Fuentes A, Porrón-Sotelo L, Huerta AM, Bonavides-Martínez C, Balderas-Martínez YI, Pannier L, Olvera M, Labastida A, Jiménez-Jacinto V, Vega-Alvarado L, Del Moral-Chávez V, Hernández-Alvarez A, Morett E, Collado-Vides J: **RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more.** *Nucleic Acids Res* 2013, **41**(Database issue):D203–D213.
76. Báez-Viveros J, Flores N, Juárez K, Castillo-España P, Bolívar F, Gosset G: **Metabolic transcription analysis of engineered *Escherichia coli* strains that overproduce L-phenylalanine.** *Microb Cell Fact* 2007, **6**:30.
77. Doroshenko VG, Shakulov RS, Kazakova SM, Kivero AD, Yampolskaya TA, Mashko SV: **Construction of an L-phenylalanine-producing tyrosine-prototrophic *Escherichia coli* strain using *tyrA* *ssrA*-like tagged alleles.** *Biotechnol Lett* 2010, **32**:1117–1121.
78. Doroshenko V, Airich L, Vitushkina M, Kolokolova A, Livshits V, Mashko S: **YddG from *Escherichia coli* promotes export of aromatic amino acids.** *FEMS Microbiol Lett* 2007, **275**:312–318.
79. Liu Q, Cheng Y, Xie X, Xu Q, Chen N: **Modification of tryptophan transport system and its impact on production of L-tryptophan in *Escherichia coli*.** *Bioresour Technol* 2012, **114**:549–554.
80. Gu P, Yang F, Li F, Liang Q, Qi Q: **Knocking out analysis of tryptophan permeases in *Escherichia coli* for improving L-tryptophan production.** *Appl Microbiol Biotechnol* 2013, **97**:6677–6683.
81. Gu P, Yang F, Kang J, Wang Q, Qi Q: **One-step of tryptophan attenuator inactivation and promoter swapping to improve the production of L-tryptophan in *Escherichia coli*.** *Microb Cell Fact* 2012, **11**:30.
82. Rodrigues AL, Trachtmann N, Becker J, Lohanatha AF, Blotenberg J, Bolten CJ, Korneli C, De Souza Lima AO, Porto LM, Sprenger GA, Wittmann C: **Systems metabolic engineering of *Escherichia coli* for production of the antitumor drugs violacein and deoxyviolacein.** *Metab Eng* 2013, **20**:29–41.
83. Koma D, Yamanaka H, Moriyoshi K, Ohmoto T, Sakai K: **Production of aromatic compounds by metabolically engineered *Escherichia coli* with an expanded shikimate pathway.** *Appl Environ Microbiol* 2012, **78**:6203–6216.
84. Muñoz AJ, Hernández-Chávez G, Anda R, Martínez A, Bolívar F, Gosset G: **Metabolic engineering of *Escherichia coli* for improving L-3,4-dihydroxyphenylalanine (L-DOPA) synthesis from glucose.** *J Ind Microbiol Biotechnol* 2011, **38**:1845–1852.
85. Sun Z, Ning Y, Liu L, Liu Y, Sun B, Jiang W, Yang C, Yang S: **Metabolic engineering of the L-phenylalanine pathway in *Escherichia coli* for the production of S- or R-mandelic acid.** *Microb Cell Fact* 2011, **10**:71.
86. Rodrigues AL, Becker J, De Souza Lima AO, Porto LM, Wittmann C: **Systems metabolic engineering of *Escherichia coli* for gram scale production of the antitumor drug deoxyviolacein from glycerol.** *Biotechnol Bioeng* 2014, **20**:1–31.
87. Eudes A, Juminaga D, Baidoo EE, Collins FW, Keasling JD, Loqué D: **Production of hydroxycinnamoyl anthranilates from glucose in *Escherichia coli*.** *Microb Cell Fact* 2013, **12**:62.
88. Lim CG, Fowler ZL, Hueller T, Schaffer S, Koffas MAG: **High-yield resveratrol production in engineered *Escherichia coli*.** *Appl Environ Microbiol* 2011, **77**:3451–3460.
89. Lynch SA, Gill RT: **Synthetic biology: New strategies for directing design.** *Metab Eng* 2012, **14**:205–211.
90. Luo Y, Lee J-K, Zhao H: **Challenges and opportunities in synthetic biology for chemical engineers.** *Chem Eng Sci* 2013, **103**:115–119.
91. Boyle PM, Silver PA: **Parts plus pipes: Synthetic biology approaches to metabolic engineering.** *Metab Eng* 2012, **14**:223–232.
92. Yadav VG, De Mey M, Giaw Lim C, Kumaran Ajikumar P, Stephanopoulos G: **The future of metabolic engineering and synthetic biology: Towards a systematic practice.** *Metab Eng* 2012, **14**:233–241.
93. Lütke-Eversloh T, Stephanopoulos G: **Combinatorial pathway analysis for improved L-tyrosine production in *Escherichia coli*: identification of enzymatic bottlenecks by systematic gene overexpression.** *Metab Eng* 2008, **10**:69–77.
94. Yao Y-F, Wang C-S, Qiao J, Zhao G-R: **Metabolic engineering of *Escherichia coli* for production of salvanic acid A via an artificial biosynthetic pathway.** *Metab Eng* 2013, **19**:79–87.
95. Albermann C, Ghanegaonkar S, Lemuth K, Vallon T, Reuss M, Armbruster W, Sprenger G: **Biosynthesis of the vitamin E compound delta-tocotrienol in recombinant *Escherichia coli* cells.** *Chembiochem* 2008, **9**:2524–2533.
96. Ghanegaonkar S, Conrad J, Beifuss U, Sprenger G, Albermann C: **Towards the in vivo production of tocotrienol compounds: engineering of a plasmid-free *Escherichia coli* strain for the heterologous synthesis of 2-methyl-6-geranylgeranyl-benzoquinol.** *J Biotechnol* 2012, **164**:238–247.
97. Nakagawa A, Minami H, Kim J-S, Koyanagi T, Katayama T, Sato F, Kumagai H: **A bacterial platform for fermentative production of plant alkaloids.** *Nat Commun* 2011, **2**:326.
98. Wu J, Du G, Zhou J, Chen J: **Metabolic engineering of *Escherichia coli* for (2S)-pinocembrin production from glucose by a modular metabolic strategy.** *Metab Eng* 2013, **16**:48–55.
99. Wu J, Liu P, Fan Y, Bao H, Du G, Zhou J, Chen J: **Multivariate modular metabolic engineering of *Escherichia coli* to produce resveratrol from L-tyrosine.** *J Biotechnol* 2013, **167**:404–411.
100. Nakajima M, Nishino Y, Tamura M, Mase K, Masai E, Otsuka Y, Nakamura M, Sato K, Fukuda M, Shigehara K, Ohara S, Katayama Y, Kajita S: **Microbial conversion of glucose to a novel chemical building block, 2-pyrone-4,6-dicarboxylic acid.** *Metab Eng* 2009, **11**:213–220.
101. Santos CNS, Xiao W, Stephanopoulos G: **Rational, combinatorial, and genomic approaches for engineering L-tyrosine production in *Escherichia coli*.** *Proc Natl Acad Sci* 2012, **109**:13538–13543.
102. Alper H, Stephanopoulos G: **Global transcription machinery engineering: A new approach for improving cellular phenotype.** *Metab Eng* 2007, **9**:258–267.

103. Kang S-Y, Choi O, Lee JK, Hwang BY, Uhm T-B, Hong Y-S: **Artificial biosynthesis of phenylpropanoic acids in a tyrosine overproducing *Escherichia coli* strain.** *Microb Cell Fact* 2012, **11**:153.
104. Zhang H, Stephanopoulos G: **Engineering *E. coli* for caffeic acid biosynthesis from renewable sugars.** *Appl Microbiol Biotechnol* 2013, **97**:3333–3341.
105. Huang Q, Lin Y, Yan Y: **Caffeic acid production enhancement by engineering a phenylalanine over-producing *Escherichia coli* strain.** *Biotechnol Bioeng* 2013, **110**:3188–3196.
106. Satoh Y, Tajima K, Munekata M, Keasling JD, Lee TS: **Engineering of L-tyrosine oxidation in *Escherichia coli* and microbial production of hydroxytyrosol.** *Metab Eng* 2012, **14**:603–610.
107. Wang HH, Isaacs FJ, Carr P a, Sun ZZ, Xu G, Forest CR, Church GM: **Programming cells by multiplex genome engineering and accelerated evolution.** *Nature* 2009, **460**:894–898.
108. Wang HH, Kim H, Cong L, Jeong J, Bang D, Church GM: **Genome-scale promoter engineering by coselection MAGE.** *Nat Methods* 2012, **9**:591–593.
109. Yang J, Seo SW, Jang S, Shin S-I, Lim CH, Roh T-Y, Jung GY: **Synthetic RNA devices to expedite the evolution of metabolite-producing microbes.** *Nat Commun* 2013, **4**:1413.
110. Na D, Yoo SM, Chung H, Park H, Park JH, Lee SY: **Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs.** *Nat Biotechnol* 2013, **31**:170–174.
111. Shimizu K: **Toward systematic metabolic engineering based on the analysis of metabolic regulation by the integration of different levels of information.** *Biochem Eng J* 2009, **46**:235–251.
112. Bro C, Nielsen J: **Impact of “ome” analyses on inverse metabolic engineering.** *Metab Eng* 2004, **6**:204–211.
113. Park JH, Lee SY, Kim TY, Kim HU: **Application of systems biology for bioprocess development.** *Trends Biotechnol* 2008, **26**:404–412.
114. Park JH, Lee SY: **Towards systems metabolic engineering of microorganisms for amino acid production.** *Curr Opin Biotechnol* 2008, **19**:454–460.
115. Bartek T, Blombach B, Lang S, Eikmanns BJ, Wiechert W, Oldiges M, Nöh K, Noack S: **Comparative <sup>13</sup>C metabolic flux analysis of pyruvate dehydrogenase complex-deficient, L-valine-producing *Corynebacterium glutamicum*.** *Appl Environ Microbiol* 2011, **77**:6644–6652.
116. Becker J, Zelder O, Häfner S, Schröder H, Wittmann C: **From zero to hero — Design-based systems metabolic engineering of *Corynebacterium glutamicum* for L-lysine production.** *Metab Eng* 2011, **13**:159–168.
117. Van Ooyen J, Noack S, Bott M, Reth A, Eggeling L: **Improved L-lysine production with *Corynebacterium glutamicum* and systemic insight into citrate synthase flux and activity.** *Biotechnol Bioeng* 2012, **109**:2070–2081.
118. Singh P, Bath TS, Juminaga D, Dahl RH, Keasling JD, Adams PD, Petzold CJ: **Application of targeted proteomics to metabolically engineered *Escherichia coli*.** *Proteomics* 2012, **12**:1289–1299.
119. Kedar P, Colah R, Shimizu K: **Proteomic investigation on the pyk-F gene knockout *Escherichia coli* for aromatic amino acid production.** *Enzyme Microb Technol* 2007, **41**:455–465.
120. Polen T, Krämer M, Bongaerts J, Wubbolts M, Wendisch VF: **The global gene expression response of *Escherichia coli* to L-phenylalanine.** *J Biotechnol* 2005, **115**:221–237.
121. Cortés-Talalpa L, Gutiérrez-Ríos RM, Martínez LM, De Anda R, Gosset G, Bolívar F, Escalante A: **Global transcriptomic analysis of an engineered *Escherichia coli* strain lacking the phosphoenolpyruvate: carbohydrate phosphotransferase system during shikimic acid production in rich culture medium.** *Microb Cell Fact* 2014, **13**:28.
122. Johansson L, Lidén G: **Transcriptome analysis of a shikimic acid producing strain of *Escherichia coli* W3110 grown under carbon- and phosphate-limited conditions.** *J Biotechnol* 2006, **126**:528–545.
123. Joyce AR, Reed JL, White A, Edwards R, Osterman A, Baba T, Mori H, Lesely SA, Palsson BO, Agarwalla S: **Experimental and computational assessment of conditionally essential genes in *Escherichia coli*.** *J Bacteriol* 2006, **188**:8259–8271.
124. Xu Z, Sun X, Yu S: **Genome-scale analysis to the impact of gene deletion on the metabolism of *E. coli*: constraint-based simulation approach.** *BMC Bioinformatics* 2009, **10**(Suppl 1):S62.
125. Rizk ML, Liao JC: **Ensemble modeling for aromatic production in *Escherichia coli*.** *PLoS One* 2009, **4**:e6903.
126. Takors R: **Scale-up of microbial processes: Impacts, tools and open questions.** *J Biotechnol* 2012, **160**:3–9.
127. Lara AR, Galindo E, Ramirez OT, Palomares LA: **Living with heterogeneities in bioreactors. Understanding the effects of environmental gradients on cells.** *Mol Biotechnol* 2006, **34**:355–381.
128. Noorman H: **An industrial perspective on bioreactor scale-down: What we can learn from combined large-scale bioprocess and model fluid studies.** *Biotechnol J* 2011, **6**:934–943.
129. Wang J, Huang J, Shi J, Xu Q, Xie X, Chen N: **Fermentation characterization of an L-tryptophan producing *Escherichia coli* strain with inactivated phosphotransacetylase.** *Ann Microbiol* 2013, **63**:1219–1224.
130. Khamduang M, Packdibamrung K, Chutmanop J, Chisti Y, Srinophakun P: **Production of L-phenylalanine from glycerol by a recombinant *Escherichia coli*.** *J Ind Microbiol Biotechnol* 2009, **36**:1267–1274.
131. Weiner M, Albermann C, Gottlieb K, Sprenger GA, Weuster-Botz D: **Fed-batch production of L-phenylalanine from glycerol and ammonia with recombinant *Escherichia coli*.** *Biochem Eng J* 2014, **83**:62–69.
132. Liu SP, Liu RX, Xiao MR, Zhang L, Ding ZY, Gu ZH, Shi GY: **A systems level engineered *E. coli* capable of efficiently producing L-phenylalanine.** *Process Biochem* 2014, **49**:751–757.
133. Hu Y, Tang T, Yang W, Zhou H: **Bioconversion of phenylpyruvic acid to L-phenylalanine by mixed-gel immobilization of *Escherichia coli* EP8-10.** *Process Biochem* 2009, **44**:142–145.

doi:10.1186/s12934-014-0126-z

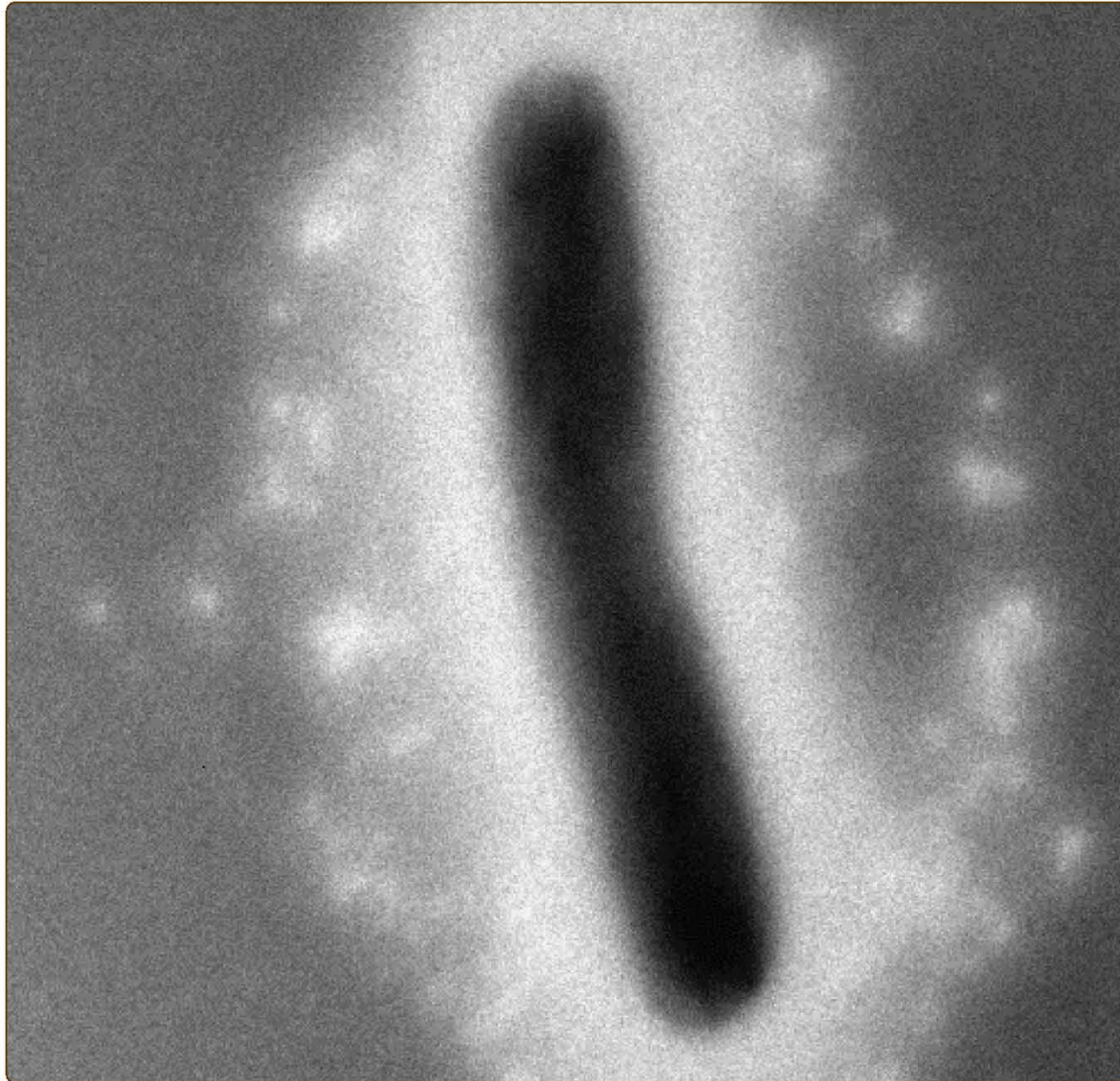
**Cite this article as:** Rodriguez et al.: Engineering *Escherichia coli* to overproduce aromatic amino acids and derived compounds. *Microbial Cell Factories* 2014 **13**:126.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit





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*

Rodriguez *et al.*

RESEARCH

Open Access

# 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*

Alberto Rodríguez<sup>1</sup>, Juan A Martínez<sup>1</sup>, José L Báez-Viveros<sup>2</sup>, Noemí Flores<sup>1</sup>, Georgina Hernández-Chávez<sup>1</sup>, Octavio T Ramírez<sup>3</sup>, Guillermo Gosset<sup>1\*</sup> and Francisco Bolívar<sup>1</sup>

## Abstract

**Background:** During the last two decades many efforts have been directed towards obtaining efficient microbial processes for the production of shikimic acid (SA); however, feeding high amounts of substrate to increase the titer of this compound has invariably rendered low conversion yields, leaving room for improvement of the producing strains. In this work we report an alternative platform to overproduce SA in a laboratory-evolved *Escherichia coli* strain, based on plasmid-driven constitutive expression of six genes selected from the pentose phosphate and aromatic amino acid pathways, artificially arranged as an operon. Production strains also carried inactivated genes coding for phosphotransferase system components (*ptsHlcr*), shikimate kinases I and II (*aroK* and *aroL*), pyruvate kinase I (*pykF*) and the lactose operon repressor (*lacI*).

**Results:** The strong and constitutive expression of the constructed operon permitted SA production from the beginning of the cultures, as evidenced in 1 L batch-mode fermentors starting with high concentrations of glucose and yeast extract. Inactivation of the *pykF* gene improved SA production under the evaluated conditions by increasing the titer, yield and productivity of this metabolite compared to the isogenic *pykF*<sup>+</sup> strain. The best producing strain accumulated up to 43 g/L of SA in 30 h and relatively low concentrations of acetate and aromatic byproducts were detected, with SA accounting for 80% of the produced aromatic compounds. These results were consistent with high expression levels of the glycolytic pathway and synthetic operon genes from the beginning of fermentations, as revealed by transcriptomic analysis. Despite the consumption of 100 g/L of glucose, the yields on glucose of SA and of total aromatic compounds were about 50% and 60% of the theoretical maximum, respectively. The obtained yields and specific production and consumption rates proved to be constant with three different substrate concentrations.

**Conclusions:** The developed production system allowed continuous SA accumulation until glucose exhaustion and eliminated the requirement for culture inducers. The obtained SA titers and yields represent the highest reported values for a high-substrate batch process, postulating the strategy described in this report as an interesting alternative to the traditionally employed fed-batch processes for SA production.

**Keywords:** Shikimic acid, Synthetic operon, *Escherichia coli*, *pykF*, Aromatic compounds

\* Correspondence: gosset@ibt.unam.mx

<sup>1</sup>Departamento de Ingeniería Celular y Biocatálisis, Instituto de Biotecnología, Universidad Nacional Autónoma de México (UNAM), Apdo. Postal 510-3, Cuernavaca, Morelos 62250, Mexico

Full list of author information is available at the end of the article

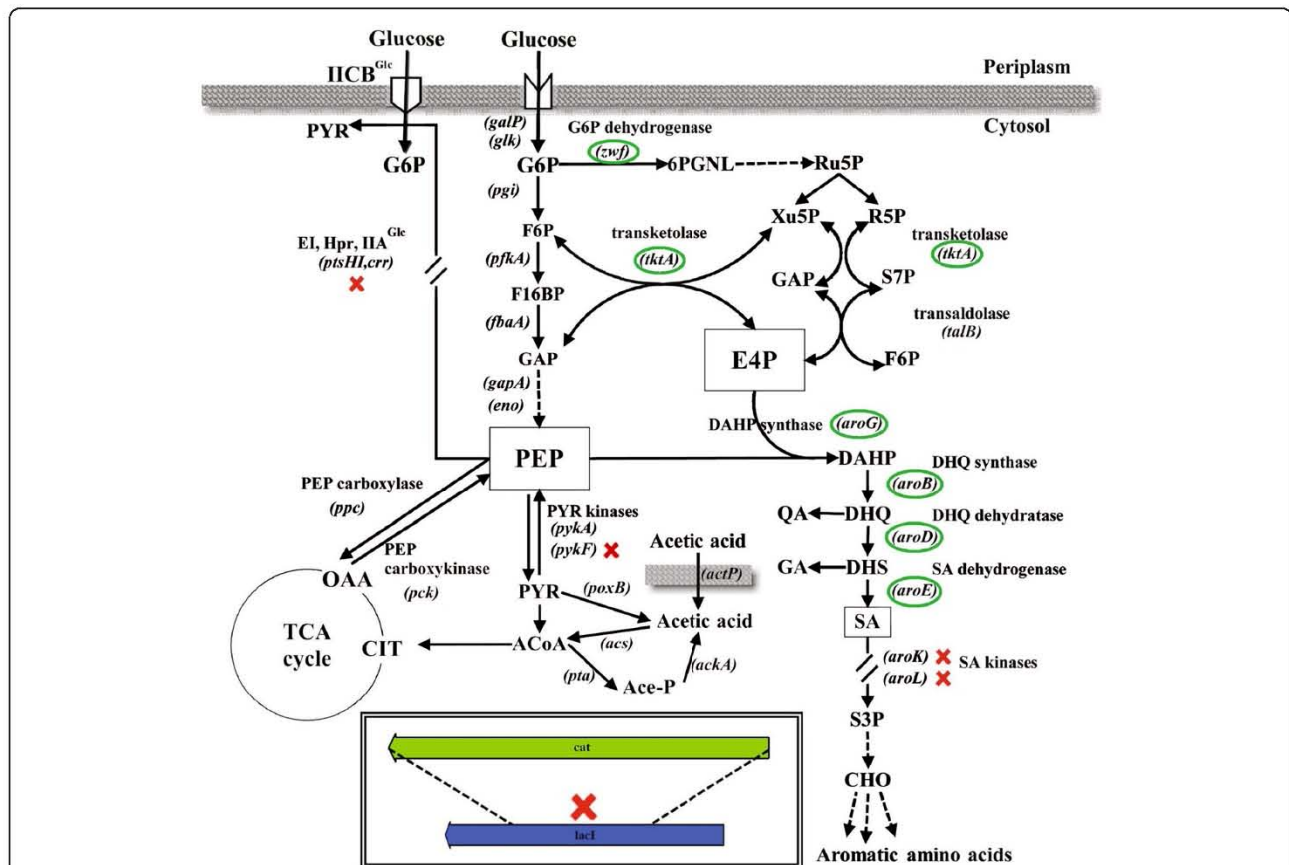


## Background

Shikimic acid (SA) is an intermediate compound in the aromatic amino acid (AAA) biosynthetic pathway in plants and bacteria (Figure 1). This metabolite is utilized as starting material in the chemical synthesis of oseltamivir phosphate (Tamiflu), used for influenza treatment [1-3]. Several genetic strategies have been reported for improving SA productivity and yield in *Escherichia coli*. These strategies aim to increase the availability of the direct precursors of the AAA pathway, erythrose 4-phosphate (E4P) and phosphoenolpyruvate (PEP), by genetic alterations that promote a convenient redistribution of carbon fluxes in the central metabolism [4,5]. Complementary approaches include the interruption of the AAA pathway after SA formation by inactivation of the genes coding for

shikimate kinases (*aroK* and *aroL*), as well as enhancements in carbon channeling towards SA by overexpression of feedback-resistant DAHP synthases, shikimate dehydrogenase, transketolase, and DHQ synthase enzymes (coded by *aroFGH<sup>fbr</sup>*, *aroE*, *tktA*, and *aroB*, respectively) (Figure 1) [6-9]. In an attempt to further increase the intracellular availability of PEP, strains overexpressing PEP synthase (coded by *ppsA*), or lacking the PEP:carbohydrate phosphotransferase system (PTS) and the pyruvate kinase isozymes (coded by *pykF* and *pykA*), have also been evaluated (Figure 1) [10-13].

Although the implementation of these modifications along with bioengineering strategies has led to diverse *E. coli* strains capable of accumulating SA, the yields obtained to date are still far from the theoretical maximum



**Figure 1** Genetic modifications applied in this report to enhance the production of shikimic acid from glucose in the laboratory-evolved *E. coli* strain PB12. Inactivated genes are indicated with a cross and plasmid-expressed genes are circled (see Methods for details). Dashed arrows indicate more than one catalytic step. G6P = glucose 6-phosphate; F6P = fructose 6-phosphate; GAP = glyceraldehyde 3-phosphate; 6PGNL = 6-phosphogluconolactone; Ru5P = ribulose 5-phosphate; R5P = ribose 5-phosphate; Xu5P = xylulose 5-phosphate; S7P = sedoheptulose 7-phosphate; E4P = erythrose 4-phosphate; PEP = phosphoenolpyruvate; PYR = pyruvate; ACoA = acetyl-coenzyme A; Ace-P = acetyl phosphate; CIT = citrate; OAA = oxaloacetate; DAHP = 3-deoxy-D-arabinoheptulosonate 7-phosphate; DHQ = 3-dehydroquinic acid; DHS = 3-dehydroshikimic acid; QA = quinic acid; GA = gallic acid; SA = shikimic acid; S3P = shikimate 3-phosphate; CHO = chorismate; IICBGlc = membrane component of glucose-specific PTS permease; E1 = PTS enzyme 1; Hpr = PTS histidine protein; IIA<sub>Glc</sub> = cytosolic component of glucose-specific PTS permease. Genes coding for enzymes not named in the figure: *galP*, galactose permease; *glk*, glucokinase; *pgi*, phosphoglucose isomerase; *pfkA*, 6-phosphofructokinase I; *fbpA*, fructose biphosphate aldolase class II; *gapA*, glyceraldehyde 3-phosphate dehydrogenase; *eno*, enolase; *actP*, acetate permease; *acs*, acetyl-coenzyme A synthetase; *pta*, phosphate acetyltransferase; *ackA*, acetate kinase; *poxB*, pyruvate oxidase.

[10,11,14,15]. This can be partially attributed to the fact that most expression systems used involve genes controlled by a mixture of inducible and native promoters of variable strengths, contained in more than one type of plasmid. These imbalances often cause a metabolic burden and heterogeneities on the intensity and temporality of gene expression, which may translate into suboptimal production capabilities of the recombinant strains, resulting in low productivity and yield of SA [16-19]. Consequently, optimized DNA expression systems and genetic backgrounds are needed for promoting a more efficient carbon channeling towards SA formation.

With the goal of producing aromatic compounds, our group has constructed and characterized strains lacking PTS, the major glucose transport system [20]. One of such strains is PB11, which grows poorly on glucose due to the inactivation of PTS [21,22]. Strain PB12, a derivative of PB11 with a 400% increased growth rate, was isolated in a short laboratory adaptive evolution process to foster derivatives growing in glucose [21,22]. This strain can simultaneously utilize glucose and other carbon sources (acetate, glycerol and various carbohydrates) in minimal medium due to the lack of catabolite repression exerted by PTS [21,23]. Whole genome analysis allowed the identification of the genetic changes that occurred in PB12, suggesting that the deletion of 12 genes, including *rppH*, *galR* and *mutH*, is the main reason for its rapid growth on glucose [24].

It was reported that PB12, which assimilates glucose by the non-PTS symporter GalP [25], can be engineered to accumulate SA in culture media containing glucose (Glc) and yeast extract (YE). For instance, when PB12 was transformed with two plasmids encoding four biosynthetic genes, the variant with both functional pyruvate kinases accumulated the highest SA concentration (up to 7 g/L), but the highest yield of aromatic compounds was achieved by a derivative with an inactivated *pykF* gene [11]. This result may be related to other reported effects caused by the inactivation of *pykF*, such as an increase in plasmid copy number per cell [26], low acetate production due to less glycolytic overflux [13,27,28], or higher concentrations of the AAA pathway enzymes [29]. Interestingly, in spite of the aforesaid features that can be beneficial for SA production, the metabolic engineering efforts to overproduce this compound have been mainly applied to strains with a *pykF*<sup>+</sup> background, probably because of their typically higher glucose consumption rates compared to the *pykF* counterparts [11,30].

Here, we propose that a PTS<sup>-</sup> *pykF*<sup>-</sup> strain has the potential to increase the yield and titer of SA when compared to an isogenic *pykF*<sup>+</sup> strain, provided that the gene expression system permits an appropriate temporal coordination in the synthesis of the enzymes required to channel the carbon towards SA, while reducing

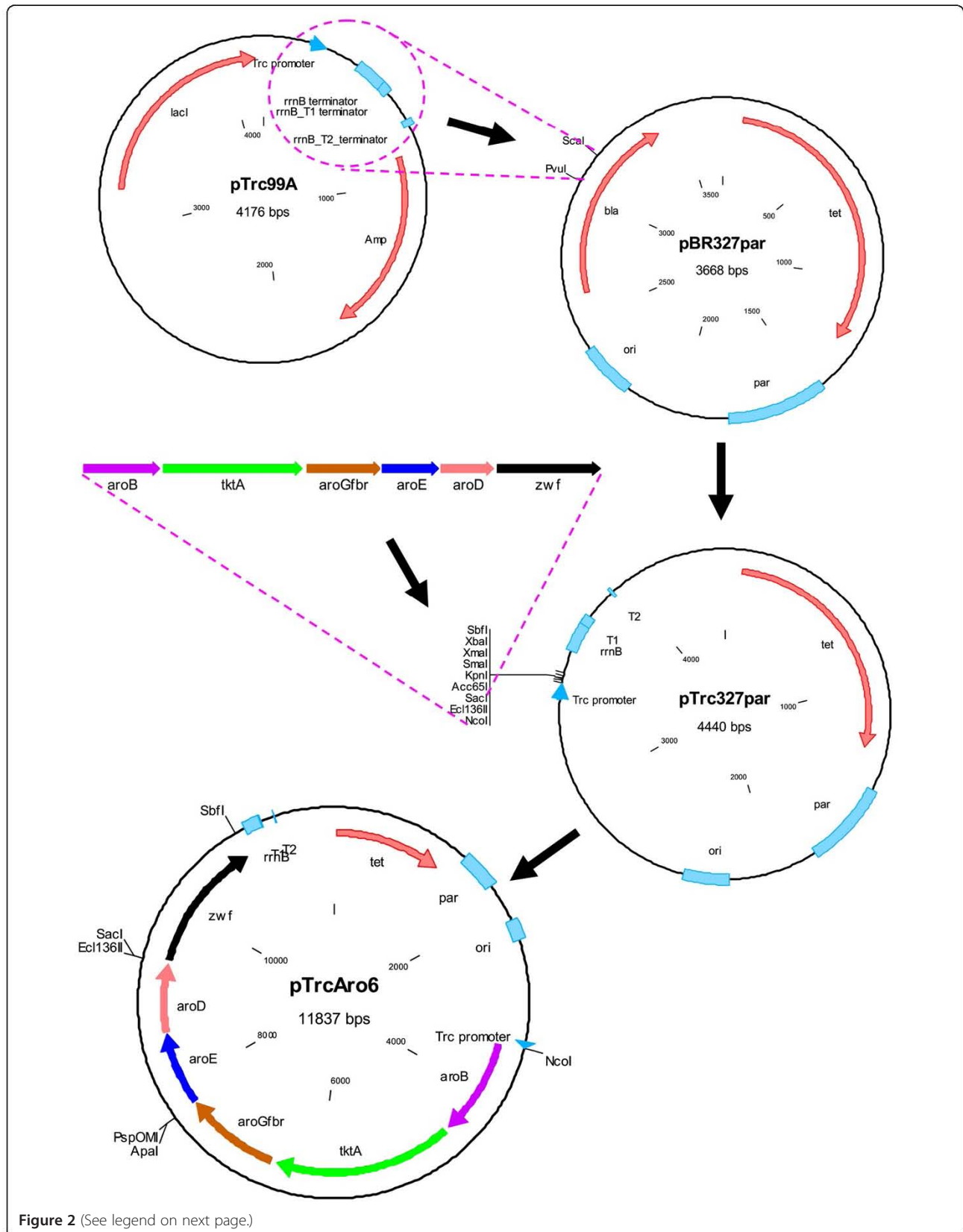
the accumulation of acetate and intermediate compounds in the AAA pathway. In order to accomplish this goal a synthetic operon was constructed containing the coding sequences of six genes selected from the pentose phosphate (PPP) and AAA pathways (Figure 1), controlled by a single constitutive Trc promoter [31] and inserted it into a high-copy plasmid containing a region that confers segregational stability (Figure 2) [32]. The resulting plasmid was transformed into a modified PB12 strain with inactive *aroK*, *aroL*, *pykF*, and *lacI* genes, and was cultured in fermentors using mineral media containing Glc and supplemented with YE. Overall, the strategy proposed in this report allowed the overproduction of SA from the beginning of the culture, resulting in a high titer and yield of SA with relatively low accumulation of acetate and aromatic byproducts. It was also found that, under the high-substrate conditions tested, the SA titer was independent of the YE concentration and the maximum biomass produced depended exclusively on the initial YE concentration but not on the amount of glucose.

## Results and discussion

### Construction of strains derived from PB12 *aroK*<sup>-</sup> *aroL*<sup>-</sup> containing a plasmid designed for the constitutive expression of a synthetic operon used in the production of shikimic acid

Unpublished evidence from our laboratory indicates that the production of aromatic compounds in the laboratory-evolved strain PB12 can attain higher levels when the transcriptional induction of the genes involved in canalizing carbon flux into the AAA pathway occurs at the beginning of fermentations. Taking into account this observation, a new strategy was developed for optimizing the production of SA in PB12 carrying inactive *aroK* and *aroL* genes (Figure 1). This strategy included the design and construction of a plasmid for the strong and stable expression of six key genes arranged in the form of a synthetic operon, controlled exclusively by a single Trc promoter. In order to reduce metabolic burden, a single plasmid derived from pBR327 carrying the *par* locus for increased plasmid stability was utilized as the vector [32], after incorporating a fragment containing the promoter, polylinker, and transcriptional terminators from pTrc99A (Figure 2).

The initial part of the operon was constructed by sequential amplification and ligation of the first 4 coding sequences (*aroB*, *tktA*, *aroG*<sup>fbt</sup>, and *aroE*) into the polylinker of plasmid pBRINT-Ts Cm, used as a cloning scaffold (see Methods). Later, the 4-gene construction was transferred to the hybrid plasmid pTrc327par in conjunction with 2 more genes (*aroD* and *zwf*), leading to an 8Kb operon contained in a 12Kb plasmid (Figure 2). The resulting plasmid, termed pTrcAro6, was transformed into the PB12 *aroK*<sup>-</sup> *aroL*<sup>-</sup> strain devoid of the *lacI*<sup>f</sup> gene,



**Figure 2** (See legend on next page.)

(See figure on previous page.)

**Figure 2 Simplified scheme of the steps required in the construction of plasmid pTrcAro6, carrying 6 synthetic genes under the control of the Trc promoter.** First, a segment of pTrc99A was amplified and ligated into pBR327par, creating plasmid pTrc327par. A synthetic operon comprising the required genes (*aroB*, *tktA*, *aroG<sup>fbr</sup>*, *aroE*, *aroD*, and *zwf*) was assembled separately and transferred to pTrc327par, generating pTrcAro6. The dotted lines indicate the site and orientation of some of the performed ligation reactions. Only the relevant restriction sites are displayed. A more detailed scheme of the constructions is presented in Additional file 3.

allowing constitutive expression of the genes of interest (Table 1). For simplicity, the generated PB12 *aroK<sup>-</sup> aroL<sup>-</sup> lacI<sup>-</sup>* strain was termed AR2. After the *pykF* gene was inactivated in AR2, the resulting strain was named AR3. Strains derived from AR2 and AR3 carrying plasmid pTrcAro6 were named AR26 and AR36, respectively (Table 1).

The spatial arrangement of the coding sequences that constitute the synthetic operon in pTrcAro6, flanked by the Trc promoter and transcriptional terminators, is shown in Figure 2. *aroB* is the first gene in the operon since several evidences indicate that its low expression is one of the limiting steps in the production of aromatic compounds [33-35]. Plasmid pTrcAro6 also carries the *tktA* and *aroG<sup>fbr</sup>* genes, whose products are involved in E4P synthesis and its condensation with PEP to form DAHP, the first aromatic compound (Figure 1). *aroD* and *aroE* genes were also included to promote an efficient conversion of DHQ to SA. Additionally, this plasmid carries

the *zwf* gene, coding for the first enzyme of the PPP (Figure 1). The decision to include this gene was based on the following observations: 1) the overexpression of *zwf* substantially recovered the growth rate loss due to plasmid metabolic load in strain JM101 growing on glucose as only carbon source [36]; 2) it has been reported that strain PB12 displays a particularly low carbon flux partition at the glucose 6-phosphate (G6P) node towards the PPP (5% of the consumed G6P compared to 22% in the parental strain JM101) [25]. Therefore, an overexpression of this gene should increase NADPH availability, required in catalytic amounts by the enzyme shikimate dehydrogenase (AroE), and may alleviate potential growth affectations by redirecting more G6P towards nucleotide and amino acid biosynthesis in strains derived from PB12 [37]. However, the experiments presented in this report did not aim to dissect the specific effect of any utilized gene but instead sought to characterize the consequences of expressing all of them as an operon.

**Table 1 *Escherichia coli* strains and plasmids utilized in this report**

Strains		
Name	Characteristics	Reference
JM101	F' <i>traD36 proA<sup>+</sup> proB<sup>+</sup> lacIq lacZΔM15/supE thi Δ(lac-proAB) rpoS(33 am)</i>	Messing [41]
PB11	JM101 <i>Δ(ptsH, ptsI, crr):kan</i>	Flores et al. [21]; Flores et al. [22]
PB12	PB11, PTS <sup>-</sup> Glc <sup>+</sup> ; laboratory-evolved strain	Flores et al. [21]; Flores et al. [22]
AR2	PB12 <i>lacI<sup>-</sup> aroK<sup>-</sup> aroL<sup>-</sup></i>	This work
AR3	PB12 <i>lacI<sup>-</sup> aroK<sup>-</sup> aroL<sup>-</sup> pykF</i>	This work
AR26	AR2 + pTrcAro6 ( <i>Trc/aroB<sup>+</sup>tktA<sup>+</sup>aroGfbr<sup>+</sup>aroE<sup>+</sup>aroD<sup>+</sup>zwf<sup>+</sup></i> )	This work
AR36	AR3 + pTrcAro6 ( <i>Trc/aroB<sup>+</sup>tktA<sup>+</sup>aroGfbr<sup>+</sup>aroE<sup>+</sup>aroD<sup>+</sup>zwf<sup>+</sup></i> )	This work
AR2e	AR2 + pTrc327par (plasmid vector without synthetic operon)	This work
AR3e	AR3 + pTrc327par (plasmid vector without synthetic operon)	This work
Plasmids		
Name	Characteristics	Reference
pKD3	PCR template for amplification of chloramphenicol resistance gene flanked by homologous sequences	Datsenko and Wanner [42]
pKD46	Plasmid expressing λ-Red recombinase system with thermosensitive origin of replication	Datsenko and Wanner [42]
pCP20	FLP recombinase expression plasmid	Cherepanov and Wackernagel [43]
pBR327par	Derivative of pBR322 exhibiting increased copy number and segregational stability	Zurita et al. [32]
pTrc99A	Multipurpose expression plasmid bearing <i>lacI</i> gene and a polylinker in front of Trc promoter	Amann et al. [31]
pTrc327par	Contains the promoter, polylinker, and terminators of pTrc99A, and <i>par</i> and <i>ori</i> regions of pBR327par	This work (Figure 2)
pTrcAro6	pTrc327par containing a 6-gene synthetic operon to enhance the production of shikimate	This work (Figure 2)

In order to promote an efficient translation of every gene, each coding sequence was amplified using designated primers that introduced a consensus Shine-Dalgarno sequence located 8 bp upstream of the translation start site. The nucleotide sequence of the constructed operon is presented in Additional file 1.

#### Assessment of the effects caused by *pykF* inactivation in strains expressing the Aro6 operon

To evaluate the effects caused by *pykF* inactivation on the production of SA, the performance of production strains AR26 (*pykF*<sup>+</sup>) and AR36 (*pykF*<sup>-</sup>) was compared using shake flasks containing 15 g/L of Glc and 5 g/L of YE. As a control, the same strains containing an empty pTrc327par plasmid (without the Aro6 operon), AR2e and AR3e, were also included.

Even though SA accumulated in all cases, as expected for mutants in *aroK* and *aroL*, the strains containing pTrcAro6 reached higher SA concentrations than the ones with an empty plasmid (Figure 3b). Moreover, the SA titer was almost two times higher in AR36 than in AR26 (6.1 g/L vs. 3.3 g/L). A decrease in Glc consumption was observed in strain AR26 after approximately 18 h of culture, correlating with high acetate concentration and an arrest in the production of SA. In contrast, strain AR36 exhibited constant Glc consumption and negligible amounts of acetate were produced (Figure 3c, 3d). These results demonstrate that the genes present in the artificial operon are functional and promote the production of SA since the beginning of the culture. Their constitutive expression diminished the specific growth rate ( $\mu$ ) by 25% in the *pykF*<sup>+</sup> background, and marginally increased it in the *pykF*<sup>-</sup> variant, but did not cause significant changes to the maximum biomass produced ( $X_{\max}$ ) compared to strains with an empty plasmid (Figure 3a). Remarkably, in the operon-expressing strains under these growth conditions, the inactivation of the *pykF* gene increased the production of SA, eliminated the accumulation of acetate, and allowed steady Glc consumption.

To determine if the higher acetate production and lower SA production in AR26 compared to AR36 is a consequence of the inherently low oxygen availability and acidification of the medium in shake flask cultures, both strains were cultured in 1 L batch fermentors under controlled conditions of pH and dissolved oxygen tension (DOT). As an approach to increase the SA titer, the initial concentration of Glc in these experiments was raised to 100 g/L, and the YE concentration was concomitantly increased to 15 g/L to allow higher biomass generation.

Under these conditions strain AR36 produced 42 g/L of SA in 60 h, consuming all the Glc, and accumulating 12 g/L of acetate. In contrast, after 47 h strain AR26 produced a maximum of 13 g/L of SA, did not exhaust the

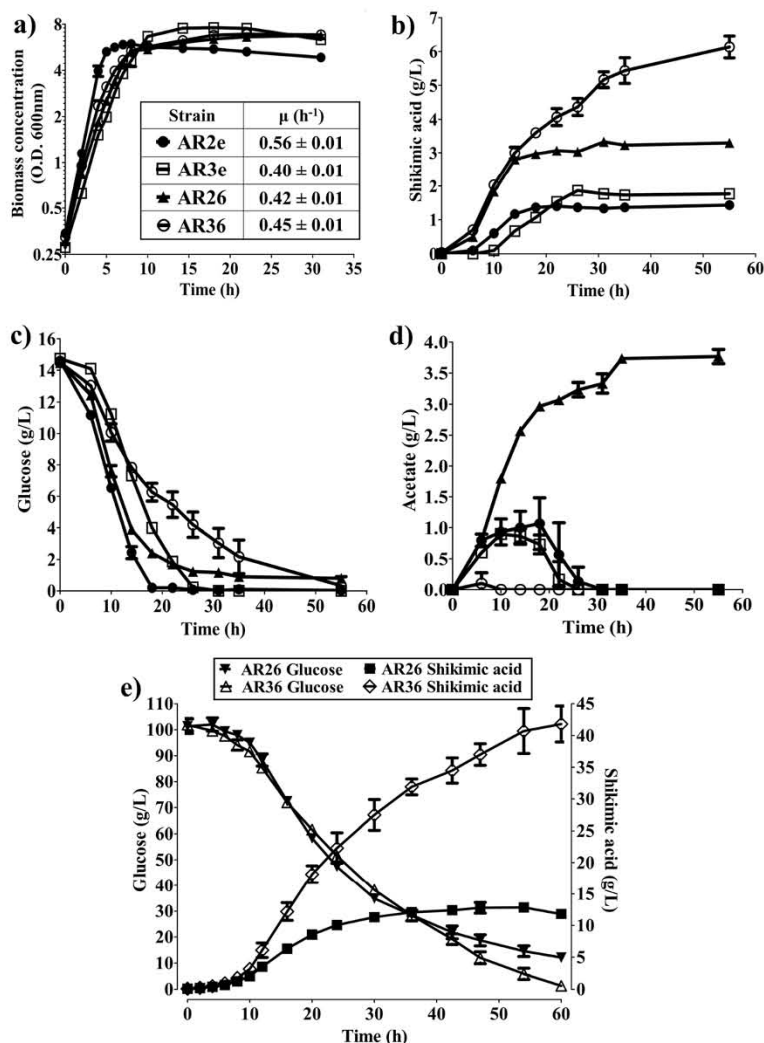
Glc, and accumulated 29 g/L of acetate (Figure 3e and Table 2). Regardless of the controlled conditions in the fermentors, where the pH was kept at 7 and the DOT was higher than 20% at all times, the production profiles of both strains resembled the behavior observed in shake flasks, with AR26 producing more acetate and less SA. Even when the global volumetric Glc consumption rate ( $Q_{s_{\text{global}}}$ ),  $\mu$  and  $X_{\max}$  attained by both strains were similar, the productivity, yield, and titer were more than twofold higher in AR36 than in AR26 (Figure 3e and Table 2).

It is remarkable that such large differences in acetate and SA production were observed by disrupting only one gene, which demonstrates the advantages of the combined inactivation of PTS and *pykF* when using a constitutive expression system in an evolved *E. coli* strain. To account for the observed improvements in SA production, we suggest that the early and constant expression of enzymes encoded in the operon could maintain a steady consumption of glycolytic intermediates throughout the cultures, preventing high fluctuations in their intracellular concentrations. We hypothesize that the combination of this steady metabolic state with a reduced flux from PEP to pyruvate caused by the inactivation of the *pykF* gene may increase the availability of PEP and other glycolytic precursors for SA production without decreasing the Glc consumption rate. However, we acknowledge that in the absence of measured intracellular metabolite concentrations, these remarks are speculative.

#### Fermentation profiles of AR36 in batch cultures

Taking into account the previous results, AR36 was selected for further characterization of its kinetic and stoichiometric performance in 1 L fermentors. To accomplish such purpose, the production of SA was tested with three different high-substrate culture conditions. Growth, Glc and byproducts were measured for each case, which in turn allowed a comparison of the productivities and yields.

First, 50 g/L of Glc and 15 g/L of YE were utilized (Figure 4a). Growth occurred during the first 10 h, generating 6.3 g/L of dry cell weight with a  $\mu$  of 0.53 h<sup>-1</sup>. Under this condition, 24 g/L of SA were produced in 32 h. Glc consumption and SA production occurred since the beginning of the fermentation and lasted until Glc exhaustion, although the specific Glc consumption rate ( $q_s$ ) and specific SA productivity ( $q_p$ ) were higher in exponential phase (Table 3). The resulting yield of SA on Glc ( $Y_{SA/Glc}$ ) was 0.47 mol/mol and the global volumetric SA productivity ( $Q_{p_{\text{global}}}$ ) was 0.74 gSA/L\*h (Table 3). With respect to the accumulation of byproducts in the SA pathway, concentrations of 2.4 g/L of DAHP, 2.1 g/L of DHS, 1.4 g/L of QA, 0.4 g/L of GA, and 0.3 g/L of DHQ, were present in the supernatant at the end of the fermentation (Figure 5a). Under these conditions, virtually no acetate was produced during



**Figure 3** Behavior of strains AR26, AR36, and their empty-plasmid derivatives, AR2e (*pykF*<sup>+</sup>) and AR3e (*pykF*), using shake flasks containing 15 g/L of Glc and 5 g/L of YE (a,b,c,d), and 1 L fermentors containing 100 g/L of Glc and 15 g/L of YE (e). a) Growth; b) SA production; c) Glc consumption; d) acetate production; e) Glc consumption and SA production of AR26 and AR36 in fermentors. Error bars represent standard deviation.

the course of the fermentation, reaching a maximum concentration of 1.5 g/L after 32 h (Figure 4a).

Considering that 50 g/L of Glc were consumed completely, a second batch experiment was initiated with 100 g/L of Glc and 15 g/L of YE. As stated in the comparison with AR26 in the previous section, AR36 grown under these conditions produced approximately 42 g/L of SA in 60 h (Figure 4b). In this case, after consuming about 100 g/L of glucose and attaining the maximum concentration of SA, the strain produced 12 g/L of acetate. The values obtained for  $Y_{SA/Glc}$ ,  $Q_{p_{global}}$ ,  $Q_{S_{global}}$ ,  $X_{max}$ , and  $\mu$ , were similar to those obtained with 50 g/L of Glc and 15 g/L of YE (Table 3). These experiments show that when using the same YE concentration, twice the amount of Glc is consumed in almost twice the time, indicating that

the average glucose consumption rate is maintained between both culture conditions. Concentrations of 4.8 g/L of DAHP, 2.8 g/L of DHS, 3.4 g/L of QA, 0.7 g/L of GA, and 0.9 g/L of DHQ, were present in the supernatant after 60 h (Figure 5b). Interestingly, when doubling the Glc concentration the intermediate products of the AAA pathway increased in a fairly proportional manner with the SA, indicating that the consumption of 100 g/L of Glc did not apparently generate new carbon flux bottlenecks. As a result, the amount of SA formed with respect to the total aromatic compounds produced was close to 80% in both experiments (Figure 6).

The effect of increasing the YE on SA productivity was investigated with a third set of experiments, using 100 g/L of Glc and 30 g/L of YE. Although the biomass was

**Table 2 Comparative data from 1 L batch fermentations of strains AR26 and AR36, using 100 g/L of Glc and 15 g/L of YE as substrates**

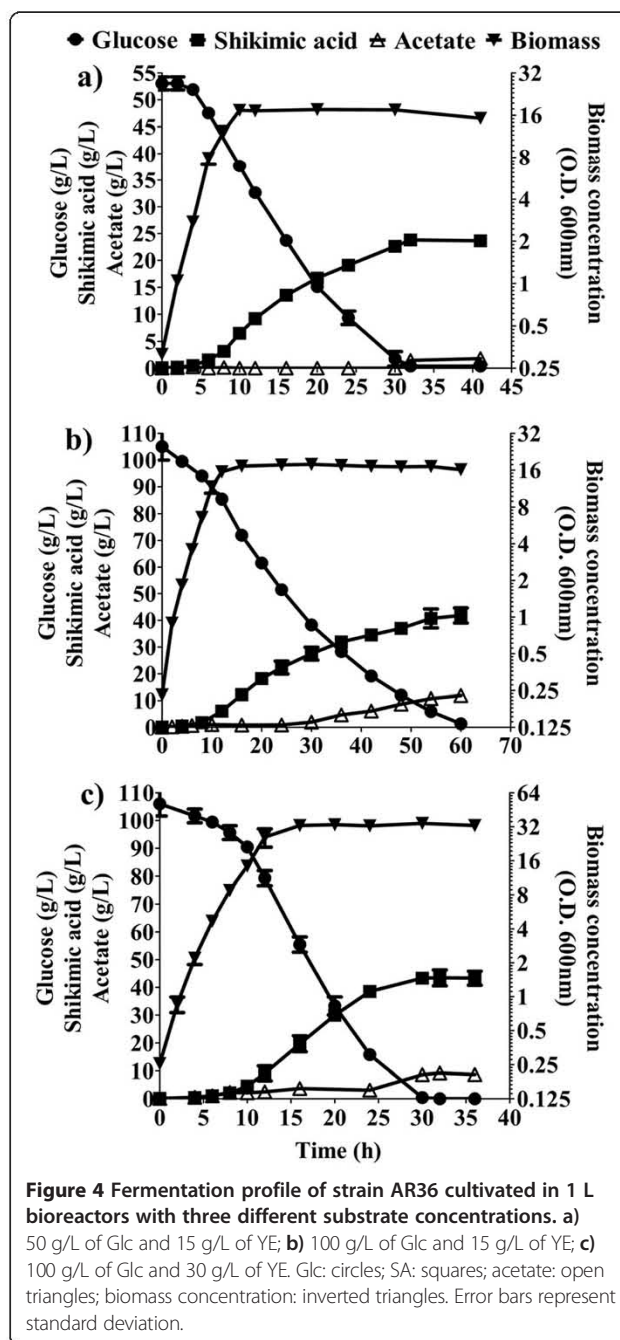
Strain	AR26	AR36
SA titer (g/L)	12.95 ± 0.64	41.80 ± 2.83
Glc consumed (g/L)	82.65 ± 4.88	103.70 ± 6.79
Duration of culture (h)	47	60
$Y_{SA/Glc}$ (mol/mol)	0.16 ± 0.02	0.42 ± 0.00
Acetate titer (g/L)	29.35 ± 0.21	11.90 ± 0.14
$X_{max}$ (g/L)	6.18 ± 0.10	6.54 ± 0.09
$\mu$ ( $h^{-1}$ )	0.45 ± 0.01	0.45 ± 0.02
$Q_{p_{global}}$ (gSA/L*h)	0.27 ± 0.02	0.75 ± 0.07
$Q_{s_{global}}$ (gGlc/L*h)	-1.76 ± 0.10	-1.73 ± 0.11

$Y_{SA/Glc}$  = SA yield from Glc;  $X_{max}$  = maximum attained biomass expressed as dry cell weight;  $Q_{p_{global}}$  = volumetric SA productivity for the entire fermentation;  $Q_{s_{global}}$  = volumetric Glc consumption rate for the entire fermentation;  $\mu$  = specific growth rate.

doubled when using twice the concentration of YE, the SA titer,  $\mu$  and  $Y_{SA/Glc}$  were very similar to those obtained in the culture with 100 g/L of Glc and 15 g/L of YE (Figure 4b and Figure 4c). In conjunction with data obtained from the other two conditions, these findings suggest that the amount of YE primarily determines the maximum biomass that can be achieved. Additionally, an increment in the initial YE concentration did not alter the SA titer, and supports the observation that SA is mainly being produced from glucose. The direct relation between the initial YE concentration and the maximum biomass generated, regardless of the initial Glc concentration tested in these growth conditions, suggests that one or more limiting nutrients are being supplied by the YE. It would also appear that such nutrients cannot be synthesized from Glc, hence their depletion from YE limits growth long before Glc is exhausted. It is expected that the aromatic amino acids and vitamins present in the YE that are needed to counteract AR36 auxotrophy will become limiting; however, other compounds in this complex media may also play a role in growth limitation over time.

For a starting YE concentration of 30 g/L, a total of 106 g/L of Glc and 43 g/L of SA were consumed and produced, respectively, in approximately half the time than the fermentation with 15 g/L of YE. With 30 g/L of YE, the  $Q_{s_{global}}$  and  $Q_{p_{global}}$  increased twofold, in comparison with the fermentations with 15 g/L of YE, even though the SA titer remained unchanged (Table 3). Since the biomass also increased twofold, the calculated qp and qs were similar between the three experiments, both in exponential and stationary phases, exhibiting the metabolic robustness of the engineered strain under the tested conditions.

In addition, the results showed that an increase in YE concentration did not increase considerably the



concentration of SA pathway intermediates (Figure 5c). In this respect, it has been acknowledged that the presence of high quantities of pathway intermediates can negatively impact the recovery of SA from the fermentation broth [7,38]. This concern has directed some efforts into the subject, leading to the testing of culture conditions, genetic backgrounds, and the use of non-metabolizable glucose analogs, as attempts to minimize byproduct generation [39].

In these experiments, a high proportion of SA relative to byproducts was detected without applying any further

**Table 3 Comparison of measured metabolites and calculated kinetic and stoichiometric parameters between three fermentations of strain AR36 with different substrate concentrations**

Strain	AR36	AR36	AR36
<b>Culture conditions</b>	Batch 50 g/L Glc + 15 g/L YE	Batch 100 g/L Glc + 15 g/L YE	Batch 100 g/L Glc + 30 g/L YE
<b>SA titer (g/L)</b>	23.80 ± 0.00	41.80 ± 2.83	43.30 ± 0.57
<b>Glc consumed (g/L)</b>	52.65 ± 1.20	103.70 ± 6.79	105.55 ± 4.45
<b>Duration of culture (h)</b>	32	60	30
<b>Y<sub>SA/Glc</sub> (mol/mol)</b>	0.47 ± 0.01	0.42 ± 0.00	0.42 ± 0.01
<b>Acetate titer (g/L)</b>	1.45 ± 0.00	11.90 ± 0.14	8.65 ± 0.92
<b>X<sub>max</sub> (g/L)</b>	6.30 ± 0.09	6.54 ± 0.09	12.54 ± 0.06
<b>μ (h<sup>-1</sup>)</b>	0.53 ± 0.03	0.45 ± 0.02	0.45 ± 0.00
<b>Q<sub>p</sub>global (gSA/L*h)</b>	0.74 ± 0.00	0.75 ± 0.07	1.44 ± 0.02
<b>Q<sub>s</sub>global (gGlc/L*h)</b>	-1.65 ± 0.04	-1.74 ± 0.11	-3.52 ± 0.15
<b>q<sub>p</sub>exp (gSA/gDCW*h)</b>	0.38 ± 0.00	0.34 ± 0.03	0.46 ± 0.06
<b>q<sub>s</sub>exp (gGlc/gDCW*h)</b>	-1.14 ± 0.15	-1.11 ± 0.00	-1.18 ± 0.03
<b>q<sub>p</sub>sta (gSA/gDCW*h)</b>	0.16 ± 0.01	0.19 ± 0.02	0.20 ± 0.01
<b>q<sub>s</sub>sta (gGlc/gDCW*h)</b>	-0.36 ± 0.01	-0.42 ± 0.01	-0.42 ± 0.03

Y<sub>SA/Glc</sub> = SA yield from Glc; X<sub>max</sub> = maximum attained biomass expressed as dry cell weight; Q<sub>p</sub>global = volumetric SA productivity for the entire fermentation; Q<sub>s</sub>global = volumetric Glc consumption rate for the entire fermentation; q<sub>p</sub>exp = specific SA productivity for exponential phase; q<sub>s</sub>exp = specific Glc consumption rate for exponential phase; q<sub>p</sub>sta = specific SA productivity for stationary phase; q<sub>s</sub>sta = specific Glc consumption rate for stationary phase; μ = specific growth rate.

modification to the strain or process. The concentration of each pathway intermediate was compared against the sum of all aromatic intermediates, and their percentages were used to calculate the molar ratio of SA to each byproduct at the end of the fermentations (Figure 6). The ratio of SA turned out to be higher than 10 for DHS, QA, or DAHP, and higher than 40 for GA or DHQ for all the substrate concentrations tested. Remarkably, in all the conditions the obtained SA yields were close to 50% of the theoretical maximum and the yields of total aromatic compounds (TAC) were above 60% of the theoretical maximum, estimated as 0.86 mol<sub>TAC</sub>/mol<sub>Glc</sub> (see Methods and Figure 6). This reflects the efficient redirection of glucose towards the AAA pathway in strain AR36, even when using high-glucose batch cultures. The ratio of SA to byproducts, as well as the obtained SA and TAC yields are fairly constant for all the conditions evaluated, and represent to our knowledge the highest reported values for a SA production fermentation process. These improvements can be justified by taking into account that the platform present in the engineered strain allows a more homogeneous expression of the necessary enzymes on an efficient genetic background. This, in contrast with other expression systems where the required genes are expressed from separate plasmids, under different promoters, or in strains not optimized for efficient use of high levels of Glc. In addition to the advantages concerning the dynamics of gene expression, the fact that IPTG is not needed to induce the Aro6 operon represents an

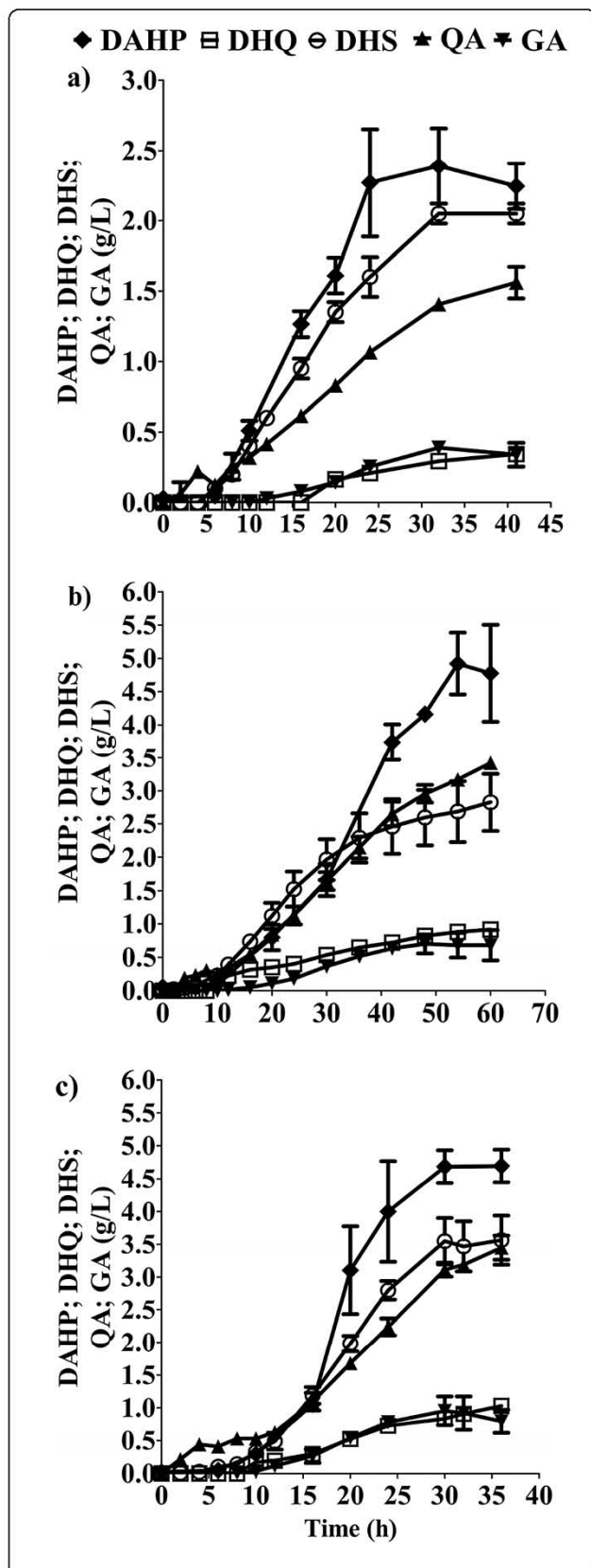
important economical benefit for the production process, since the high price of IPTG restricts its use in large-scale fermentations.

#### Insights on the glycolytic and acetate metabolisms of strain AR36 by RT-qPCR

To gain a deeper insight of the metabolic changes induced by the constitutive expression of the Aro6 synthetic operon in strain AR36, transcript levels of several genes were measured at three different growth stages in cultures with 50 g/L of Glc and 15 g/L of YE. As detailed in Methods, data obtained from early exponential phase (EE), late exponential phase (LE), and stationary phase (ST) were normalized against the values measured from strain AR3e at EE, grown under the same culture conditions.

The results indicate that the presence and expression of the operon in strain AR36 increases the transcriptional levels of several genes coding for glycolytic enzymes during the EE and LE phases (Figure 1 and Figure 7a). The rise in expression of genes *galP* and *glk* is particularly interesting because it has been reported that their products control the import and phosphorylation of glucose in PB12, the parental strain of AR36 [21,25]. Furthermore, there is a significant increase in the transcriptional levels of *pgi* and *eno*, but not *pykA*. These changes may translate into higher availability of PEP and fructose 6-P (which can be directly converted into E4P by plasmid-encoded transketolase), increasing the yield of aromatic compounds. We theorize that the observed



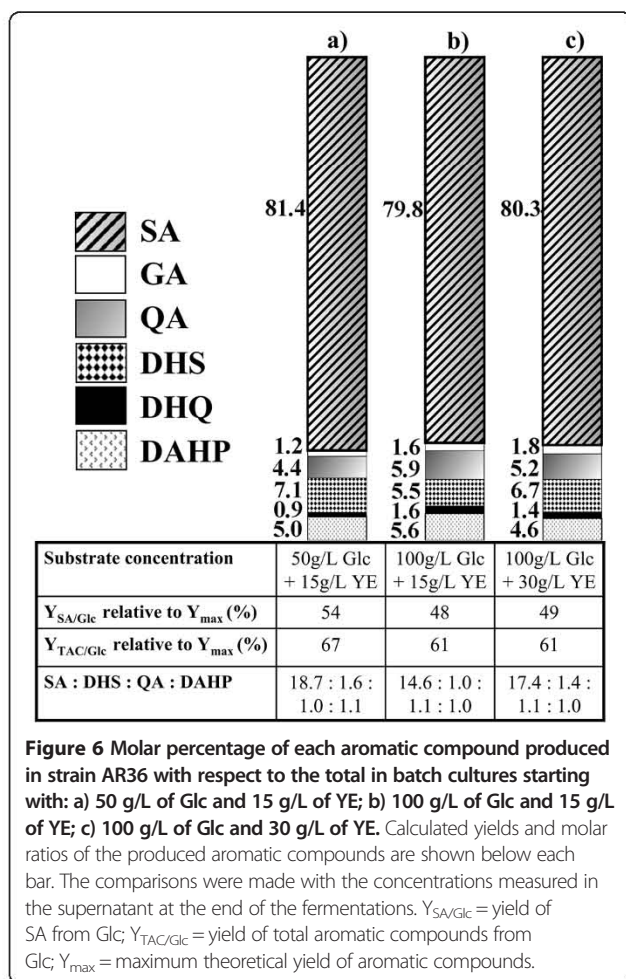


**Figure 5 Aromatic byproducts of the SA pathway detected in 1 L fermentor cultures of strain AR36 using three different substrate concentrations.** a) 50 g/L of Glc and 15 g/L of YE; b) 100 g/L of Glc and 15 g/L of YE; c) 100 g/L of Glc and 30 g/L of YE. Diamonds: DAHP (3-deoxy-D-arabinoheptulosonate 7-phosphate); squares: DHQ (3-dehydroquinic acid); circles: DHS (3-dehydroshikimic acid); triangles: QA (quinic acid); inverted triangles: GA (gallic acid). Error bars represent standard deviation.

upregulation of glycolytic genes in strain AR36 could be one of the consequences to low levels of some glycolytic intermediates (glucose 6-phosphate, fructose 6-phosphate and PEP), caused by the strong and constitutive expression of the operon-encoded enzymes that consume these metabolites.

On the other side, the transcriptional levels of genes coding for enzymes involved in acetate biosynthesis (*poxB*, *ackA* and *pta*) were not modified by the presence of the synthetic operon, while *actP* and *acs*, coding for enzymes involved in acetate assimilation, were strongly upregulated in the EE and LE phases (Figure 7b). Upregulation of *actP* and *acs* genes has also been detected in the exponential growth phase in the parental strain PB12 that is capable of co-utilizing Glc and acetate in minimal medium [21]. These findings correlate with the low levels of acetate in the assayed growth condition (Figure 4a). Importantly, the transcriptional values of these genes involved in acetate assimilation were low in ST phase (Figure 7b). If this response is representative of the other growth conditions used, it could partially explain the acetate accumulation observed in fermentations with 100 g/L of Glc, which consume higher amounts of Glc during stationary phase (Figure 4b and Figure 4c). These results highlight *actP* and *acs* as potential gene targets to artificially increase their expression in late culture stages, taking advantage of the expected capabilities of strain AR36 to utilize simultaneously Glc and acetate, present in its parental strain PB12 [21,40].

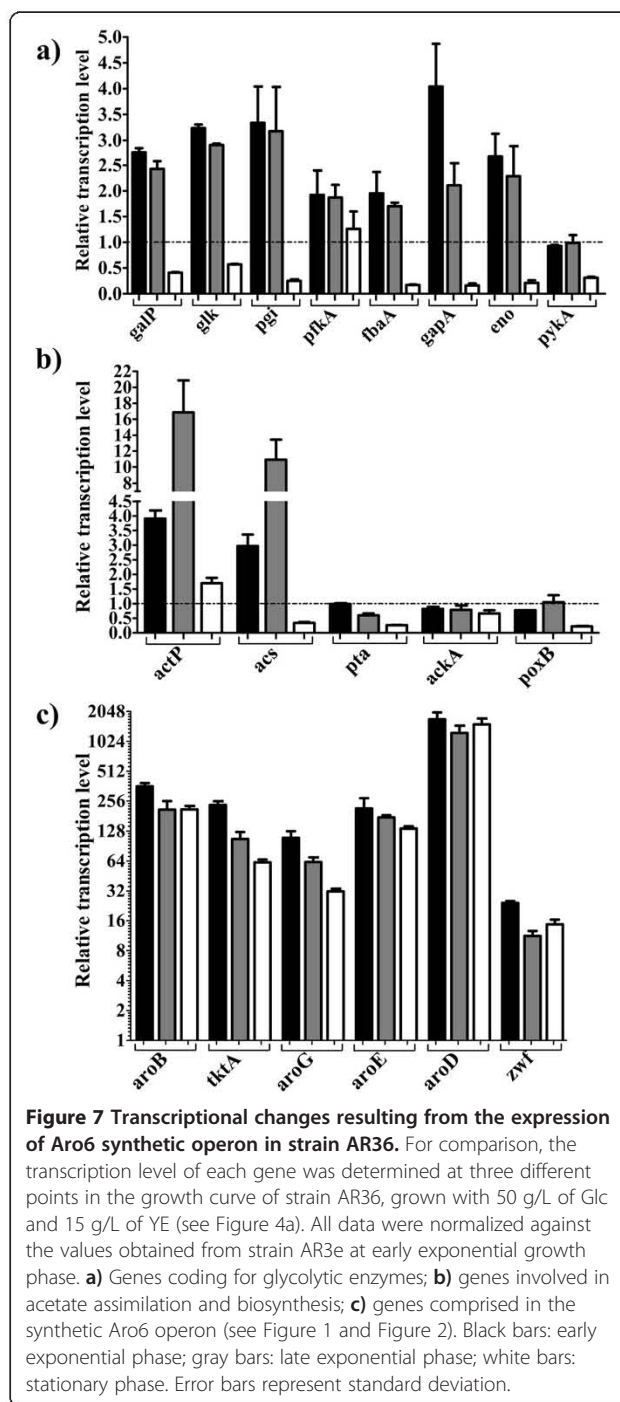
The genes present in the synthetic operon showed very strong expression levels (even in stationary phase), reflecting the constitutive nature of the promoter and high copy number of the plasmid (Figure 7c). These results correlate with the uninterrupted Glc consumption and SA production observed during the entire fermentation (Figure 4a), suggesting that the enzymes coded by the genes in the operon are present throughout the cultivation time. It can be seen in Figure 7c that the transcript levels of *aroD* and *zwf* are comparatively higher and lower, respectively, than the other four genes in the operon. This observation should be taken with caution because the six genes in the operon are being compared to the ones present in the chromosome of reference strain AR3e. Since the values obtained for the six genes are not normalized



between them, variations amongst their chromosomal expression in strain AR3e can alter the relative comparisons with strain AR36. Nevertheless, the transcriptomic data is consistent with the high ratio of SA to aromatic intermediates obtained in the tested conditions, which is to be expected if all the genes in the operon were adequately expressed. Together with kinetic and stoichiometric data, these results highlight the benefits of employing a constitutively-expressed synthetic operon as an alternate strategy to increase the yield of SA from Glc in an evolved strain that lacks PTS and *pykF*.

### Conclusions

*E. coli* is the microorganism that has given the best results for SA production, with engineered strains that can accumulate up to 85 g/L using 10 L fermentors in fed-batch processes [10]. In this report, we showed that the constitutive and synchronous expression of a six-gene synthetic operon, in a laboratory-evolved strain bearing simultaneous PTS and *pykF* inactivations, resulted in a competitive process for the production of SA. The expression of Aro6 operon in the PTS<sup>-</sup> *pykF* derivative



resulted in higher  $Q_{p,global}$  and similar  $Q_{s,global}$  than its PTS<sup>-</sup> *pykF*<sup>+</sup> counterpart. In addition, the Glc consumption and SA production profiles of strain AR36 are consistent with the observed increase in transcription levels of glycolytic genes as a response to the constitutive expression of the operon in this strain. These features translated into significant improvements in growth and production parameters in strain AR36 (producing 0.74 gSA/L\*h with 54% yield, using 50 g/L of Glc + 15 g/L of YE), compared

to the PTS<sup>-</sup> *pykF* strain reported by Escalante et al. in 2010 (producing 0.11 gSA/L\*h with 26% yield, using 25 g/L of Glc + 15 g/L YE).

Albeit fed-batch fermentations in the past have given the best results with respect to SA production, here we report that under the appropriate conditions, batch cultures of strain AR36 with an initially high concentration of Glc can also be efficiently used to produce SA. However, the production profiles obtained suggest that fed-batch fermentations could also yield good results with this strain. The fact that the SA production ceases only when the glucose is exhausted suggests that higher titers could be achieved by adequate Glc feeding strategies and improvements in the acetate uptake capabilities of this strain, considering that it lacks PTS and could co-utilize Glc and acetate in these growing conditions. The fermentations reported here yielded an elevated ratio of SA to other byproducts of the pathway (10 times more SA than the main byproducts generated). Besides increasing the SA yield, this behavior is relevant for purifying SA from the culture broth. In fact, preliminary experiments concerning SA purification from the broth obtained from these cultures, resulted in an almost quantitative purification process (unpublished results). Furthermore, the highest yield of total aromatic compounds obtained represents 67% of the theoretical maximum, demonstrating the efficient redirection of carbon to the AAA pathway by strain AR36. Nevertheless, the relatively low cellular concentration present in the cultures, even when administering high concentrations of YE, represents a significant problem to this production system because it restricts the productivity of the process. Other strategies need to be utilized in order to increase the biomass concentration without increasing the supplemented YE, which will constitute an important improvement for scaling-up the process. Minimizing the metabolic load imposed by a high-copy plasmid while maintaining a sufficient gene dosage of the operon, should improve the distribution of resources that are directed towards biomass generation and SA production.

## Methods

### Construction of *Escherichia coli* derivatives and plasmids

The laboratory-evolved strain PB12, a derivative of PB11 (obtained by the inactivation of PTS in strain JM101 [41]), was the receptor of the genetic modifications described in this work [21,22,24]. The strains and plasmids used in this report are listed in Table 1. The chromosomal inactivations of *aroK*, *aroL*, and *lacI* genes were performed sequentially by homologous recombination of PCR products [42]. In all cases, plasmid pKD3 was used as a PCR template in conjunction with tailored oligonucleotides containing 45 bp homology with the target chromosomal sequence (Additional file 2). Plasmid pKD46 expressed the

Red recombinase system of bacteriophage lambda, and plasmid pCP20 allowed removal of the chloramphenicol resistance cassette after each event [43]. Every step was verified by PCR, identifying the clones that presented the expected amplicon sizes when using different sets of oligonucleotides and chromosomal DNA as a template (Additional file 2). The *pykF* inactivated gene was transduced to PB12 *aroK* *aroL*<sup>-</sup> using a P1 phage lysate obtained from strain PB28 (*pykF*::gen) [44]. Transductants were selected on gentamycin plates (10 µg/ml), and the inactivation was confirmed by PCR.

The construction of the Aro6 operon and expression vector was accomplished in several steps (Figure 2 and Additional file 3). First, *aroB*, *aroG*<sup>fbt</sup>, *tktA* and *aroE* genes were amplified by PCR using Pfu DNA polymerase and ligated sequentially into the polylinker of plasmid pBRINT-Ts Cm [45]. Chromosomal DNA from strain JM101 was used as a template for amplification of the required genes, with the exception of *aroG*<sup>fbt</sup>, which was amplified from plasmid pJLBaroG<sup>fbt</sup>*tktA* [46]. Different sets of oligonucleotides were employed for the amplification of each gene (Additional file 2), which also generated flanking restriction sites and consensus Shine-Dalgarno sequences (AGGAGG) situated 8 bp upstream of the start of each coding sequence. The PCR products were inserted into the polylinker in the following order: *aroB* in the *SmaI* site, *aroG*<sup>fbt</sup> in the *XhoI* site, *tktA* in the *EcoRV* site, and *aroE* in the *ApaI* site. Simultaneously, plasmid pTrc327par*lacI*<sup>+</sup> (Additional file 3) was built by ligating a PCR-amplified fragment containing the *lacI* gene, Trc promoter, polylinker, and transcriptional terminators of pTrc99A [31], into the *ScaI* and *PvuII* sites of pBR327par [32]. The 4-gene operon present in plasmid pBRINT-Ts Cm was amplified by PCR with a unique set of oligonucleotides (Additional file 1) and ligated into pTrc327par*lacI*<sup>+</sup> after digesting both with *SacI* and *XbaI*. Later, *aroD* was amplified by PCR (flanked by *NheI* sites) and ligated into compatible *XbaI* site of pTrc327par*lacI*<sup>+</sup>. Because of our interest in expressing the operon in a constitutive manner, a *lacI* derivative of the initial pTrc327par*lacI*<sup>+</sup> plasmid (without synthetic operon) was generated, and called pTrc327par (Additional file 3). The 5-gene operon was then transferred into *SacI* and *NcoI* sites of pTrc327par, giving rise to pTrcAro5. Finally, the *zwf* gene was inserted into the *XbaI* site, creating a 6-gene operon in pTrc327par. The resulting plasmid was named pTrcAro6 and transformed into AR2 and AR3 strains, generating AR26 and AR36, respectively (Table 1). The transformed strains were selected in LB plates supplemented with tetracycline (30 µg/ml).

Each step in the gene cloning and plasmid construction schemes was screened by endonuclease digestion and PCR, visualized with gel electrophoresis, and verified by DNA sequencing (3730, Perkin-Elmer/Applied Biosystems, USA). All the enzymes and reagents used in the molecular biology

procedures were purchased from Fermentas (USA) and New England Biolabs (USA). When required, kits for the purification of PCR, plasmid, and agarose-embedded DNA were utilized (Roche, Switzerland). TOP10 cells (Invitrogen, USA) were used as a host for screening of DNA ligations during intermediate steps in vector construction.

### Cultivation media and growth conditions

#### Composition of production medium

SA production medium (adjusted to pH 7.0 with 10 N NaOH) contained per liter:  $K_2HPO_4$  (7.5 g),  $KH_2PO_4$  (7.5 g), citric acid monohydrate (2.1 g), ammonium iron (III) citrate (0.3 g), concentrated  $H_2SO_4$  (1.2 ml),  $MgSO_4$  (0.64 g),  $CaCl_2$  (0.06 g),  $(NH_4)_6(Mo_7O_{24})$  (0.0037 g),  $ZnSO_4$  (0.0029 g),  $H_3BO_3$  (0.0247 g),  $CuSO_4$  (0.0025 g),  $MnCl_2$  (0.0158 g),  $CoCl_2$  (0.00129 g), thiamine (0.001 g), and betaine (0.234 g) as an osmoprotectant. Tetracycline (30  $\mu$ g/ml) was added to inocula and cultures whenever needed for plasmid maintenance. Glucose (filter-sterilized) and yeast extract (added before autoclaving) were supplied at the concentrations indicated for each experiment. The glucose was purchased from JT Baker (USA) and the autolysed yeast extract from BD Difco (USA).

#### Shake flask cultures

The inoculum preparation for the shake flask cultures started by the addition of 1 ml frozen aliquots to 250 ml shake flasks containing 25 ml of production medium supplemented with glucose (25 g/L) and yeast extract (15 g/L). The inoculum was grown at 37°C and 300 rpm until mid-exponential phase, and approximately 5% of the final volume was transferred to the test shake flasks and incubated under the same controlled conditions with media containing 15 g/L of glucose and 5 g/L of yeast extract. Cell growth was measured by monitoring the optical density at 600 nm ( $OD_{600}$ ) in a DU700 spectrophotometer (Beckman, USA), and samples were taken periodically, centrifuged, and the supernatant was stored at -20°C for metabolite analysis. These experiments were performed at least in triplicate. All cultures started at approximately 0.3  $OD_{600}$ .

#### Fermentor cultures

Batch cultures were performed at least in duplicate using 1 L autoclavable glass bioreactors (Applikon, The Netherlands) with 500 ml of working volume. Bioreactors were connected to an Applikon ADI 1010 BioController and ADI 1025 controllers to monitor temperature, pH, impeller speed, and dissolved oxygen tension (DOT). The pH was kept at 7.0 by the addition of  $H_3PO_3$  (3.3%) and  $NH_4OH$  (10%). DOT in the culture medium was maintained by a continuous supply of filtered air (1 vvm), and by manually controlling the impeller speed (ranging from 500 to 1000 rpm) to ensure that DOT was

kept above 20% at all times. The inoculum preparation for the fermentors started by the addition of 1 ml frozen aliquots to 500 ml shake flasks containing 50 ml of production medium supplemented with glucose (25 g/L) and yeast extract (15 g/L). The strains were grown at 37°C and 300 rpm until mid-exponential phase and approximately 5% of the final volume was transferred from each inoculum to previously prepared bioreactors containing the production medium. All fermentations were performed in presence of tetracycline (30  $\mu$ g/ml). Cell growth was measured by monitoring optical density at 600 nm ( $OD_{600}$ ) in a spectrophotometer (DU700, Beckman, USA), and samples were taken periodically, centrifuged, and the supernatant was stored at -20°C for metabolite analysis. All the fermentations started at approximately 0.3  $OD_{600}$ .

#### Metabolite quantification

The supernatant from each sample was properly diluted and filtered through 0.45  $\mu$ m nylon membranes. Shikimic acid (SA), 3-dehydroshikimic acid (DHS), 3-dehydroquinic acid (DHQ), quinic acid (QA), gallic acid (GA), acetic acid, and glucose (Glc) concentrations were determined by HPLC using a Waters system (600E quaternary pump, 717 automatic injector, 2410 refraction index, and 996 photodiode array detectors; USA) equipped with an Aminex HPX-87H column (300  $\times$  7.8 mm; 9  $\mu$ m; Bio-Rad, USA). The mobile phase was 5 mM  $H_2SO_4$ , with a flow rate of 0.5 ml/min, maintained at 50°C. 3-deoxy-D-arabinoheptulose 7-phosphate (DAHP) concentrations were determined colorimetrically by the thiobarbituric acid assay [47]. This method does not distinguish between DAHP and its unphosphorylated form, DAH, therefore in this work DAHP levels correspond to the sum of both compounds.

#### Data analysis and calculations

The measured concentrations of metabolites and biomass were normalized to the starting volume conditions to account for changes derived from pH control in fermentors. Data from independent experiments were averaged and presented in the corresponding graphs, where the error bars indicate the standard deviation for each point. Biomass concentration ( $X$ ) was determined with a calibration curve between dry cellular weight and  $OD_{600}$ , resulting in the equation  $X = 0.3587 * OD_{600}$ . Specific growth rate ( $\mu$ ) was determined by linearly fitting the biomass concentration to time during exponential phase with the following equation:  $\ln X = \ln X_0 + \mu * t$  (where  $t$  is time, and  $X_0$  is the biomass concentration at initial time), displaying  $R^2$  values >0.97. The yield of SA from Glc ( $Y_{SA/Glc}$ ) was calculated with the average molar concentrations of SA and Glc, produced and consumed, respectively, at the point of highest SA concentration. The yield of total aromatic compounds from glucose ( $Y_{TAC/Glc}$ ) was calculated with the combined molar

yields of DAHP, DHQ, DHS, SA, QA, and GA at the point of highest SA concentration. The maximum theoretical yield of aromatic compounds was previously estimated as  $0.86 \text{ mol}_{\text{TAC}}/\text{mol}_{\text{Glc}}$  for a PTS<sup>-</sup> strain growing on glucose as only carbon source [10]. The global volumetric SA productivity ( $Q_{\text{P}_{\text{global}}}$ ) and the global volumetric Glc consumption rate ( $Q_{\text{S}_{\text{global}}}$ ) were calculated taking into account the time needed to reach the maximum SA concentration. Besides the previous calculations, linearizations were made to obtain apparent biomass on substrate ( $Y_{\text{X/S}}$ ) and product on biomass ( $Y_{\text{P/X}}$ ) yields. Although these apparent yields do not take into account the yeast extract consumption, correlation values for linearizations in all experiments were found to be  $>0.95$ , allowing comparisons between them. These yields were used to calculate the specific productivity and specific consumption rate on the exponential phase ( $q_{\text{P}_{\text{exp}}}$  and  $q_{\text{S}_{\text{exp}}}$ , respectively) with the following equations:  $q_{\text{P}_{\text{exp}}} = Y_{\text{P/X}} \mu$ ;  $q_{\text{S}_{\text{exp}}} = \mu/Y_{\text{X/S}}$ .

The volumetric productivity and volumetric Glc consumption rate in stationary phase were determined by linearization of the first concentration data points at this stage versus time. The volumetric rates were utilized for calculation of specific production and consumption rates at stationary phase ( $q_{\text{P}_{\text{sta}}}$  and  $q_{\text{S}_{\text{sta}}}$ , respectively) by dividing them by the average biomass concentration.

#### RNA extraction, cDNA synthesis and RT-qPCR analysis

Samples from batch fermentations of strain AR36 with 50 g/L of Glc and 15 g/L of YE were collected for RNA extraction at early exponential phase, EE (2 h ~ 1 OD<sub>600</sub>), late exponential phase, LE (8 h ~ 12 OD<sub>600</sub>), and stationary phase, ST (24 h ~ 17 OD<sub>600</sub>), to determine gene expression levels. For comparison of the data, samples from early exponential phase (3.5 h ~ 1 OD<sub>600</sub>) of strain AR3e (bearing an empty pTrc327par plasmid carrying only the tetracycline-resistance gene, Figure 2) cultured under the same conditions were also collected and processed. RNA was extracted using hot phenol equilibrated with water, and cDNA synthesis was performed using RevertAid H First Strand cDNA Synthesis kit (Fermentas, USA) and a mixture of specific DNA primers, as reported previously [21,24]. qPCR experiments were performed with the ABI Prism 7300 Real Time PCR System (Applied Biosystems, USA) using Maxima SYBRGreen PCR Master Mix (Fermentas, USA) and reaction conditions previously described [21,24]. The quantification technique used to compare data was the  $2^{-\Delta\Delta\text{CT}}$  method [48] and the results were normalized using the *ihfB* gene as an internal control. The same reproducible expression level for this gene was detected in all the strains and conditions analyzed [24]. All qPCR experiments complied with the MIQE guidelines for publication of quantitative real-time PCR experiments [49]. Using cells from two separate

fermentations, RNA extraction and cDNA synthesis reactions were performed for each biological replicate at the indicated times and the gene expression values were measured by triplicate for each sample. Average values were graphed, with error bars representing standard deviation. Standard deviation was less than 30% in all cases.

#### Additional files

**Additional file 1: Nucleotide sequence of the synthetic operon constructed in this work and present in plasmid pTrcAro6.** The *aroG<sup>fb</sup>* gene included in this construction was a gift from DuPont™-Genencor®, therefore its coding sequence cannot be disclosed. Each nucleotide of *aroG<sup>fb</sup>* is indicated with an “n” except for the ones corresponding to its start and stop codons.

**Additional file 2: Oligonucleotides utilized in this work.**

**Additional file 3: Detailed scheme of the construction of plasmid pTrcAro6.**

#### Competing interests

The authors have declared no competing interests.

#### Authors' contributions

AR, JLB and FB conceived the study and designed the experiments. AR and JAM carried out all fermentation experiments and data analysis. JAM performed the calculation of fermentation parameters. NF carried out RNA extraction and RT-qPCR experiments. GH was involved in metabolite quantification and analysis. AR performed the molecular biology and strain construction procedures and wrote the manuscript. JAM, OTR, GG and FB critically revised the results and the manuscript. All authors have read and approved the final manuscript.

#### Acknowledgements

The authors thank Ramón de Anda for technical support, José González for assistance with strain construction, Enrique Morett for helpful discussions, Paul Gaytán, Jorge Yáñez and Eugenio López for the synthesis of oligonucleotides and sequencing of DNA (Instituto de Biotecnología-UNAM, México), as well as Mercedes Enzaldo and Aurelia González for technical assistance. A. Rodríguez was supported by grants 215044 and 290604 from Consejo Nacional de Ciencia y Tecnología (CONACyT, México). This work was supported by grants 105782 and 177568 (CONACyT), 44126 and 126793 (FONSEC/SSA/ISSSTE/CONACyT), and IN224709 and IN205811 (DGAPA-PAPIIT-UNAM). The funding sources had no involvement in study design, collection and analysis of data, writing of the report or decision to publish.

#### Author details

<sup>1</sup>Departamento de Ingeniería Celular y Biotecnología, Instituto de Biotecnología, Universidad Nacional Autónoma de México (UNAM), Apdo. Postal 510-3, Cuernavaca, Morelos 62250, Mexico. <sup>2</sup>Centro de Investigación en Biotecnología, Universidad Autónoma del Estado de Morelos (UAEM), Av. Universidad 2000, Cuernavaca, Morelos 62250, Mexico. <sup>3</sup>Departamento de Medicina Molecular y Bioprocesos, Instituto de Biotecnología, Universidad Nacional Autónoma de México (UNAM), Apdo. Postal 510-3, Cuernavaca, Morelos 62250, Mexico.

Received: 21 June 2013 Accepted: 28 August 2013

Published: 30 September 2013

#### References

1. Farina V, Brown JD: **Tamiflu: the supply problem.** *Angew Chem Int Ed Engl* 2006, **45**:7330–7334.
2. Kim HK, Park KJ: **A new efficient synthesis of oseltamivir phosphate (Tamiflu) from (-)-shikimic acid.** *Tetrahedron Lett* 2012, **53**:1561–1563.
3. Nie LD, Ding W, Shi XX, Quan N, Lu X: **A novel and high-yielding asymmetric synthesis of oseltamivir phosphate (Tamiflu) starting from (-)-shikimic acid.** *Tetrahedron: Asymmetry* 2012, **23**:742–747.

4. Patnaik R, Spitzer RG, Liao JC: **Pathway engineering for production of aromatics in *Escherichia coli*: confirmation of stoichiometric analysis by independent modulation of AroG, TktA, and Pps activities.** *Biotechnol Bioeng* 1995, **46**:361–370.
5. Gosset G, Yong-Xiao J, Berry A: **A direct comparison of approaches for increasing carbon flow to aromatic biosynthesis in *Escherichia coli*.** *J Ind Microbiol* 1996, **17**:47–52.
6. Draths KM, Knop DR, Frost JW: **Shikimic acid and quinic acid: replacing isolation from plant sources with recombinant microbial biocatalysis.** *J Am Chem Soc* 1999, **121**:1603–1604.
7. Johansson L, Lindskog A, Silfversparre G, Cimander C, Nielsen KF, Lidén G: **Shikimic acid production by a modified strain of *E. coli* (W3110.shik1) under phosphate-limited and carbon-limited conditions.** *Biotechnol Bioeng* 2005, **92**:541–552.
8. Chen K, Dou J, Tang S, Yang Y, Wang H, Fang H, Zhou C: **Deletion of the *aroK* gene is essential for high shikimic acid accumulation through the shikimate pathway in *E. coli*.** *Bioresour Technol* 2012, **119**:141–147.
9. Flores N, Xiao J, Berry A, Bolivar F, Valle F: **Pathway engineering for the production of aromatic compounds in *Escherichia coli*.** *Nat Biotechnol* 1996, **14**:620–623.
10. 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**:808–814.
11. Escalante A, Calderón R, Valdivia A, De Anda R, Hernández G, Ramírez OT, Gosset G, Bolivar F: **Metabolic engineering for the production of shikimic acid in an evolved *Escherichia coli* strain lacking the phosphoenolpyruvate: carbohydrate phosphotransferase system.** *Microb Cell Fact* 2010, **9**:21.
12. Juminaga D, Baidoo EEK, Redding-Johanson AM, Bath TS, Burd H, Mukhopadhyay A, Petzold CJ, Keasling JD: **Modular engineering of L-tyrosine production in *Escherichia coli*.** *Appl Environ Microbiol* 2012, **78**:89–98.
13. Meza E, Becker J, Bolivar F, Gosset G, Wittmann C: **Consequences of phosphoenolpyruvate:sugar phosphotransferase system and pyruvate kinase isozymes inactivation in central carbon metabolism flux distribution in *Escherichia coli*.** *Microb Cell Fact* 2012, **11**:127.
14. Ahn JO, Lee HW, Saha R, Park MS, Jung JK, Lee DY: **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**:1773–1784.
15. Ghosh S, Chisti Y, Banerjee UC: **Production of shikimic acid.** *Biotechnol Adv* 2012, **30**:1425–1431.
16. Glick BR: **Metabolic load and heterologous gene expression.** *Biotechnol Adv* 1995, **13**:247–261.
17. Keasling JD: **Gene-expression tools for the metabolic engineering of bacteria.** *Trends Biotechnol* 1999, **17**:452–460.
18. Silva F, Queiroz JA, Domingues FC: **Evaluating metabolic stress and plasmid stability in plasmid DNA production by *Escherichia coli*.** *Biotechnol Adv* 2012, **30**:691–708.
19. Carneiro S, Ferreira EC, Rocha I: **Metabolic responses to recombinant bioprocesses in *Escherichia coli*.** *J Biotechnol* 2013, **164**:396–408.
20. Postma PW, Lengeler JW, Jacobson GR: **Phosphoenolpyruvate: carbohydrate phosphotransferase systems.** In *Escherichia coli and Salmonella: Cellular and Molecular Biology*. 2nd edition. Edited by Neidhart FC. USA: ASM Press; 1996:1149–1174.
21. Flores N, Flores S, Escalante A, De Anda R, Leal L, Malpica R, Georgellis D, Gosset G, Bolivar 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.
22. Flores N, Leal L, Sigala JC, De Anda R, Escalante A, Martínez A, Ramírez OT, Gosset G, Bolivar F: **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* 2007, **13**:105–116.
23. Martínez K, De Anda R, Hernández G, Escalante A, Gosset G, Ramírez OT, Bolivar F: **Couitilization of glucose and glycerol enhances the production of aromatic compounds in an *Escherichia coli* strain lacking the phosphoenolpyruvate: carbohydrate phosphotransferase system.** *Microb Cell Fact* 2008, **7**:1.
24. Aguilar C, Escalante A, Flores N, De Anda R, Riveros-McKay F, Gosset G, Morett E, Bolivar 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**:385.
25. Flores S, Gosset G, Flores N, De Graaf AA, Bolivar F: **Analysis of carbon metabolism in *Escherichia coli* strains with an inactive phosphotransferase system by (<sup>13</sup>C) labeling and NMR spectroscopy.** *Metab Eng* 2002, **4**:124–137.
26. Cunningham DS, Liu Z, Domagalski N, Koepsel RR, Ataii MM, Domach MM: **Pyruvate kinase-deficient *Escherichia coli* exhibits increased plasmid copy number and cyclic AMP levels.** *J Bacteriol* 2009, **191**:3041–3049.
27. Ponce E: **Effect of growth rate reduction and genetic modifications on acetate accumulation and biomass yields in *Escherichia coli*.** *J Biosci Bioeng* 1999, **87**:775–780.
28. Zhu T, Phalakornkule C, Koepsel RR, Domach MM, Ataii MM: **Cell growth and by-product formation in a pyruvate kinase mutant of *E. coli*.** *Biotechnol Prog* 2001, **17**:624–628.
29. Kedar P, Colah R, Shimizu K: **Proteomic investigation on the *pyk-F* gene knockout *Escherichia coli* for aromatic amino acid production.** *Enzyme Microb Technol* 2007, **41**:455–465.
30. Al Zaid Siddiquee K, Arauzo-Bravo MJ, Shimizu K: **Metabolic flux analysis of *pykF* gene knockout *Escherichia coli* based on 13 C-labeling experiments together with measurements of enzyme activities and intracellular metabolite concentrations.** *Appl Microbiol Biotechnol* 2004, **63**:407–417.
31. Amann E, Ochs B, Abel K: **Tightly regulated *tac* promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*.** *Gene* 1988, **69**:301–315.
32. Zurita M, Bolivar F, Soberon X: **Construction and characterization of new cloning vehicles VII. Construction of plasmid pBR327par, a completely sequenced, stable derivative of pBR327 containing the *par* locus of pSC101.** *Gene* 1984, **28**:119–122.
33. Dell KA, Frost JW: **Identification and removal of impediments to biocatalytic synthesis of aromatics from D-Glucose: rate-limiting enzymes in the common pathway of aromatic amino acid biosynthesis.** *J Am Chem Soc* 1993, **115**:11581–11589.
34. Oldiges M, Kunze M, Degenring D, Sprenger GA, Takors R: **Stimulation, monitoring, and analysis of pathway dynamics by metabolic profiling in the aromatic amino acid pathway.** *Biotechnol Prog* 2004, **20**:1623–1633.
35. Lütke-Eversloh T, Stephanopoulos G: **Combinatorial pathway analysis for improved L-tyrosine production in *Escherichia coli*: identification of enzymatic bottlenecks by systematic gene overexpression.** *Metab Eng* 2008, **10**:69–77.
36. Flores S, De Anda-Herrera R, Gosset G, Bolivar F: **Growth-rate recovery of *Escherichia coli* cultures carrying a multicopy plasmid, by engineering of the pentose-phosphate pathway.** *Biotechnol Bioeng* 2004, **87**:485–494.
37. Sprenger GA: **Genetics of pentose-phosphate pathway enzymes of *Escherichia coli* K-12.** *Arch Microbiol* 1995, **164**:324–330.
38. Knop DR, Draths KM, Chandran SS, Barker JL, Von Daeniken R, Weber W, Frost JW: **Hydroaromatic equilibration during biosynthesis of shikimic acid.** *J Am Chem Soc* 2001, **123**:10173–10182.
39. Krämer M, Bongaerts J, Bovenberg R, Kremer S, Müller U, Orf S, Wubbolts M, Raeven L: **Metabolic engineering for microbial production of shikimic acid.** *Metab Eng* 2003, **5**:277–283.
40. Gimenez R, Nuñez MF, Badia J, Aguilar J, Baldoma L: **The gene *yjcG*, cotranscribed with the gene *acs*, encodes an acetate permease in *Escherichia coli*.** *J Bacteriol* 2003, **185**:6448–6455.
41. Messing J: **Recombinant DNA technical bulletin.** *NIH Publ* 1979, **2**:43–48.
42. Datsenko K, Wanner B: **One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products.** *Proc Natl Acad Sci U S A* 2000, **97**:6640–6645.
43. Cherepanov PP, Wackernagel W: **Gene disruption in *Escherichia coli*: Tc<sup>R</sup> and Km<sup>R</sup> cassettes with the option of FLP-catalyzed excision of the antibiotic-resistance determinant.** *Gene* 1995, **158**:9–14.
44. Ponce E, Flores N, Martínez A, Valle F, Bolivar F: **Cloning of the two pyruvate kinase isoenzyme structural genes from *Escherichia coli*: the relative roles of these enzymes in pyruvate biosynthesis.** *J Bacteriol* 1995, **177**:5719–5722.
45. Le Borgne S, Palmeros B, Valle F, Bolivar F, Gosset G: **pBRINT-Ts: a plasmid family with a temperature-sensitive replicon, designed for chromosomal integration into the *lacZ* gene of *Escherichia coli*.** *Gene* 1998, **223**:213–219.

46. Balderas-Hernández VE, Sabido-Ramos A, Silva P, Cabrera-Valladares N, Hernández-Chávez G, Baez-Viveros JL, Martínez A, Bolívar F, Gosset G: **Metabolic engineering for improving anthranilate synthesis from glucose in *Escherichia coli***. *Microb Cell Fact* 2009, **8**:19.
47. Weissbach A, Hurwitz J: **The Formation of 2-Keto-3-deoxyheptonic Acid in Extracts of *Escherichia coli* B**. *J Biol Chem* 1958, **234**:705–709.
48. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  method**. *Methods* 2001, **25**:402–408.
49. Bustin SA, Benes V, Garson JA, Hellems J, Hugget J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT: **The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments**. *Clin Chem* 2009, **55**:611–622.

doi:10.1186/1475-2859-12-86

**Cite this article as:** Rodriguez *et al.*: 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*. *Microbial Cell Factories* 2013 **12**:86.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

