

Problema de Optimización Multiobjetivo

Modelo a Escala Genómica de la Bacteria *Bizionia argentinensis*

Ramos Matías Hernán^{a, b}

^a Departamento de Ingeniería Química, Universidad Nacional Del Sur (UNS), Argentina

^b Planta Piloto de Ingeniería Química, PLAPIQUI (UNS-CONICET), Bahía Blanca, Argentina
Bahía Blanca, Argentina

Contact: ramosmatíashernan@gmail.com & mramos@plapiqui.edu.ar

RESUMEN

En este trabajo se propone el desarrollo de un modelo de optimización multiobjetivo aplicando el método de las épsilon restricciones con el fin de lograr una clara visualización de las diferentes soluciones en términos de productividad de biomasa y succinato para la bacteria *Bizionia argentinensis*. Se obtiene el frente de Pareto y se concluye que cerca del extremo ($v_{biomasa} = 39.99 \text{ min}^{-1}$; $v_{succinato} = 208.86 \text{ gDW biomasa.min}^{-1}$) se concentran las soluciones más atractivas desde el punto de vista de las productividades pretendidas.

I. INTRODUCCIÓN

La tesis está orientada al diseño y optimización de bioprocesos con un enfoque de Ingeniería de Sistemas Metabólicos (SME, Systems Metabolic Engineering). Dentro de los diferentes segmentos que componen la matriz productiva mundial, el sector biotecnológico es uno de los que más invierte en Investigación y Desarrollo (I+D). Así como en su momento la tecnología a vapor tuvo un rol central en la revolución industrial, y las tecnologías de la información y la comunicación (TICs) están jugando un rol trascendente en la actualidad, hay fuertes indicios de que la biotecnología (junto con la nanotecnología) serán las tecnologías que estarán en el centro de la revolución que viene. En el marco de la tesis se pretende estudiar y validar diferentes metodologías basadas en modelado matemático para facilitar los procesos de innovación en el sector biotecnológico.

Uno de los pilares fundamentales del éxito de una empresa biotecnológica cuyos productos se obtienen en base a cultivos microbianos es contar con un microorganismo de producción a escala industrial. Para lograr esto, en la actualidad se realizan modificaciones genéticas a los microorganismos mediante enfoques convencionales de ingeniería metabólica resultando en un proceso lento y costoso, siendo así un cuello de botella en el desarrollo u optimización de bioprocesos.

En este contexto, en la tesis se busca validar como solución a esta problemática la formulación de modelos metabólicos a escala genómica que permitan determinar las estrategias óptimas de ingeniería metabólica para el diseño de “fábricas celulares”.

Para el proyecto final del curso “Herramientas de Optimización Multiobjetivo”, se trabajará con un caso de

estudio de un modelo metabólico de la bacteria *Bizionia argentinensis*, desarrollado en el marco de la tesis para el estudio de la optimización de la producción de compuestos de interés.

II. MARCO TEÓRICO

Una de las bases fundamentales para el desarrollo de nuevos bioprocesos bajo el concepto de SME son los modelos matemáticos de redes metabólicas a escala genómica (GENome-scale Metabolic models, GEMs). Éstos incluyen información del genoma completo del microorganismo con todas las reacciones y metabolitos relacionados al mismo y proveen una visión integrada del metabolismo permitiendo el estudio de la distribución de flujos de las distintas vías metabólicas, la predicción de tasas de crecimiento y producción de los compuestos de interés bajo diferentes condiciones ambientales considerando una multitud de componentes y sus interacciones al mismo tiempo. Permiten, además, el diseño de cepas industriales más robustas y eficientes a través de estrategias óptimas de ingeniería metabólica que generalmente no resultan intuitivas [1].

El enfoque de Modelado Basado en Restricciones (Constraint-Based Modeling, CBM), que combina el análisis estequiométrico con el uso de técnicas de programación matemática para el estudio de los GEMs, constituye actualmente la herramienta mejor establecida y más poderosa para el modelado de redes metabólicas a gran escala para el estudio y optimización de “fábricas celulares” [2] y son aplicables prácticamente a cualquier especie con genoma secuenciado.

De acuerdo a lo anterior, la hipótesis de trabajo es que a través de la aplicación de herramientas computacionales es posible diseñar y optimizar bioprocesos de manera eficiente con un enfoque SME.

III. METODOLOGÍA

Los modelos matemáticos de redes metabólicas se formulan a partir de balances de masa para cada metabolito de la red y se pueden expresar en forma matricial de la siguiente manera:

$$\frac{dC}{dt} = Sv \quad (1)$$

donde C representa la concentración de cada metabolito dentro de la célula, S representa la matriz estequiométrica de $i \times j$, con i correspondiendo al conjunto de los metabolitos (M) y j al conjunto de las reacciones (R), y v representa el vector de flujos metabólicos

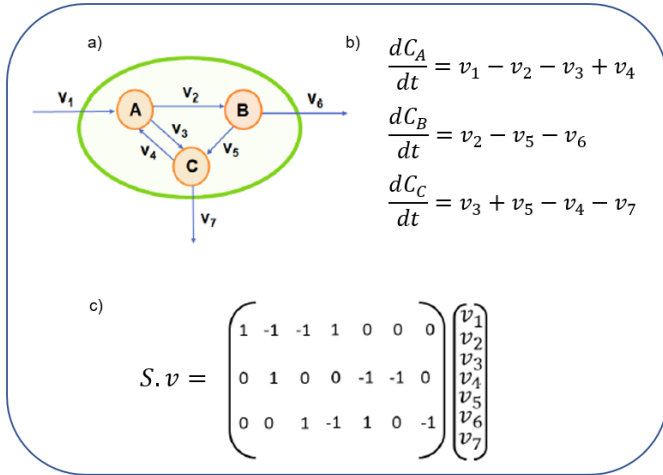


Figura 1. Modelo sencillo de red metabólica. a) Red metabólica con 3 metabolitos (A, B y C) y 7 reacciones. b) Balances de masa dinámicos para cada metabolito. c) Matriz estequiométrica S .

o velocidades de reacción. Se puede observar un ejemplo sencillo en la Figura 1.

Considerando que los transitorios intracelulares son del orden de milisegundos, y que el tiempo de crecimiento celular es del orden de horas, podemos considerar un cuasi estado estacionario, obteniendo el siguiente sistema de ecuaciones lineales:

$$Sv = 0 \quad (2)$$

Típicamente el número de reacciones es mayor que el número de metabolitos ($j > i$), por lo tanto, la ec. (2) no determina la distribución de flujos por sí sola, si no que define un sistema lineal indeterminado que tiene infinitas soluciones (distribuciones de flujo). El enfoque más empleado es la formulación de un problema de Programación Lineal (LP) en el cual se toma como objetivo la maximización de velocidad de formación de biomasa (BOF, Biomass Objective Function). Este enfoque se denomina Análisis de Balance de Flujo (FBA, Flux Balance Analysis) y constituye una herramienta básica para el estudio de redes metabólicas, permitiendo determinar la distribución de flujos para un estado estacionario de la red mediante la resolución de un problema de Programación Lineal (LP) [3].

El problema LP resultante es:

$$\begin{aligned} & \max v_{biomasa} \\ & \text{s.t.} \\ & \sum_{j \in R} S_{i,j} v_j = 0 \quad \forall i \in M \end{aligned} \quad (3)$$

$$LB_j \leq v_j \leq UB_j \quad \forall j \in R \quad (4)$$

$$v_j \in R$$

donde LB_j y UB_j son cotas inferior y superior, respectivamente, para todos los flujos de la red, mediante los cuales pueden fijarse valores para el ingreso de nutrientes.

Resolviendo el LP se obtiene la distribución de flujos para un estado de la red determinado por los límites inferior y superior de las variables, y las restricciones de tipo (3). Se pueden plantear problemas LP con otras funciones objetivo, como podría ser la maximización del flujo metabólico hacia un producto de interés del microorganismo, o también la maximización del ATP (energía celular).

En este sentido, como caso de estudio se propone trabajar con un modelo metabólico a escala genómica de la bacteria *Bizionia argentinensis*, para la cual se representó matemáticamente el modelo metabólico empleando las estrategias previamente mencionadas, y se seleccionan 2 funciones objetivo: maximización del crecimiento celular (flujo metabólico de la biomasa) y maximización de la producción de succinato (flujo metabólico de la vía generadora de succinato):

$$\max v_{biomasa}$$

$$\max v_{succinato}$$

s.t.

$$\sum_{j \in R} S_{i,j} v_j = 0 \quad \forall i \in M$$

$$LB_j \leq v_j \leq UB_j \quad \forall j \in R$$

$$v_j \in R$$

IV. RESULTADOS Y DISCUSIONES

En la Figura 2 se pueden observar algunas de las reacciones y metabolitos del GEM de la bacteria *B. argentinensis* a modo de ejemplo para ayudar a la comprensión de cómo se describen estos modelos metabólicos. El GEM desarrollado actualmente para *B. argentinensis* contempla 1328 reacciones y 1011 metabolitos (Modelo "A"). Al mismo tiempo, se considera un segundo modelo que representa una versión condensada del GEM de *B. argentinensis*, con el fin de comprobar la eficiencia del enfoque multiobjetivo que se emplee para resolver el problema. Este último contiene las reacciones metabólicas centrales del organismo. En su totalidad contiene 77 reacciones y 63 metabolitos (Modelo "B").

Para poder el modelo A, el cual contiene 1011 restricciones de balances de masa, 2022 restricciones de cotas inferiores y superiores, 1328 variables y 2 funciones objetivo ($v_{biomasa}$ y $v_{succinato}$); y el modelo B que presenta 63 restricciones de balances de masa, 154 restricciones de cotas inferiores y superiores, 77 variables y las mismas 2 funciones objetivo, se utiliza Pyomo en jupyter Notebook mediante el planteo de un modelo abstracto. Luego, para crear instancias del modelo abstracto, es decir, transformarlo en un modelo concreto, se incorpora un archivo de extensión ".dat" en donde se encuentran guardados todos los parámetros requeridos, como lo son los elementos de la matriz estequiométrica, las cotas inferiores y superiores para cada variable de flujo, y los parámetros correspondientes a cada función objetivo.

| Reaction | Protein | Lower bound | Upper bound | EC Number |
|--|------------------------------------|-------------|-------------|-----------|
| $o2[C_c] + 2\text{mbcoa}[C_c] \rightarrow 2\text{h2o}[C_c] + 2\text{mb2coa}[C_c]$ | (WP_038530051_1 or WP_038531106_1) | 0 | 1000 | |
| $fad[C_c] + \text{hxcoa}[C_c] \rightleftharpoons fadh2[C_c] + \text{hx2coa}[C_c]$ | WP_038531747_1 | -1000 | 1000 | 1.3.99.3 |
| $fad[C_c] + \text{occoa}[C_c] \rightleftharpoons fadh2[C_c] + \text{oc2coa}[C_c]$ | WP_038531747_1 | -1000 | 1000 | 1.3.99.3 |
| $nad[C_c] + \text{dccoa}[C_c] \rightleftharpoons h[C_c] + \text{nadh}[C_c] + \text{dc2coa}[C_c]$ | WP_038531747_1 | -1000 | 1000 | |
| $fad[C_c] + \text{dcacoa}[C_c] \rightleftharpoons fadh2[C_c] + \text{dc2coa}[C_c]$ | WP_038531747_1 | -1000 | 1000 | 1.3.99.3 |

Nº de reacciones: 1328

Nº de metabolitos: 1011

| Abbreviation | Metabolite | Formula |
|--------------|---------------------------|---------------|
| 10fth[C_c] | 10-Formyltetrahydrofolate | C20H21N7O7 |
| adp[C_c] | ADP | C10H12N5O10P2 |
| atp[C_c] | ATP | C10H12N5O13P3 |
| glu_L[C_c] | L-Glutamate | C5H8NO4 |

Figura 2. Modelo sencillo de red metabólica. a) Red metabólica con 3 metabolitos (A, B y C) y 7 reacciones. b) Balances de masa dinámicos parabolito. c) Matriz estequiométrica S.

En el siguiente paso, se realiza un ordenamiento lexicográfico para obtener los valores mínimos y máximos de cada función objetivo (crecimiento celular y producción de succinato) en el frente de Pareto.

A continuación, se aplica el método de las épsilon restricciones [4] para obtener las diferentes soluciones en el frente de Pareto. En la Figura 3 y 4 se presentan los frentes de Pareto resultantes del problema de optimización multiobjetivo, para los 2 modelos disponibles.

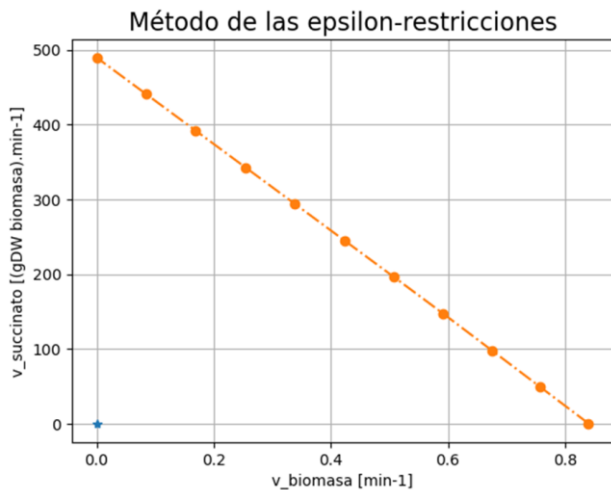


Figura 3. Frente de Pareto del MOP para el modelo B. Se busca maximizar el crecimiento celular y la producción del compuesto de interés succinato.

Una vez obtenido el frente de Pareto, se cuenta con una herramienta muy útil para la toma de decisiones. En el caso del estudio presente, se requiere de soluciones más orientadas a maximizar la producción de succinato ($v_{succinato} \geq 50$ gDW biomasa.min⁻¹) evitando aquellas con crecimiento celular aproximadamente nulo ($v_{biomasa} \cong 0$ min⁻¹). Analizando el frente de Pareto correspondiente al modelo A, se puede observar que no hay soluciones óptimas que en las que el crecimiento celular sea nulo, siendo el mínimo valor posible $v_{biomasa} = 39.99$ min⁻¹ ($v_{succinato} = 208.86$ gDW biomasa

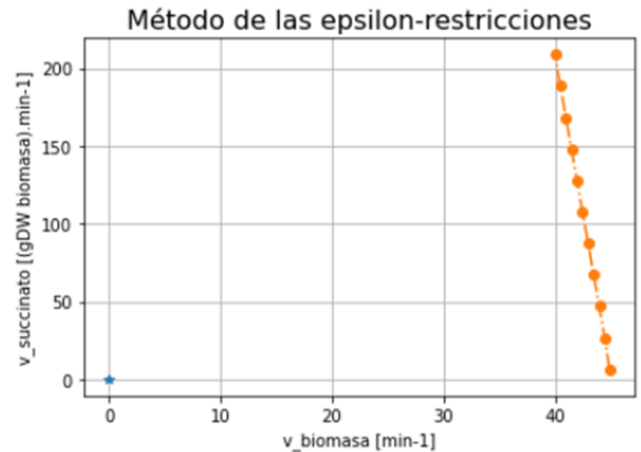


Figura 4. Frente de Pareto del MOP para el modelo A. Se busca maximizar el crecimiento celular y la producción del compuesto de interés succinato.

. min⁻¹), considerado apropiado en el contexto analizado. En el otro extremo del frente se alcanza un valor de $v_{biomasa}$ de 44.91 min⁻¹ y $v_{succinato} = 6.69$ gDW biomasa.min⁻¹. Esta solución extrema mejora en un 12.30 % respecto de la anterior, en términos del crecimiento celular ($v_{biomasa}$). Sin embargo, en términos de la productividad de succinato, decrece en un 96.80 %, resultando en una solución poco atractiva para ser seleccionada. Si bien existen soluciones intermedias del frente que serían buenas alternativas, se concluye que las mejores soluciones se encuentran alrededor del extremo ($v_{biomasa} = 39.99$ min⁻¹; $v_{succinato} = 208.86$ gDW biomasa.min⁻¹). El flujo metabólico hacia la producción de biomasa de las soluciones cercanas a dicho extremo es suficientemente elevado, y para determinar con precisión cual solución seleccionar, se debe realizar un estudio detallado de la cantidad de productos intracelulares (% de la biomasa), como pigmentos, se pueden extraer con el fin de agregar valor a los procesos que empleen la cepa *B. argentinensis*.

Respecto a la resolución del MOP para el modelo condensado B, se puede apreciar un comportamiento similar en el frente de Pareto respecto del modelo completo A, ya que en

uno de los extremos donde el crecimiento celular es máximo ($v_{biomasa} = 0.84 \text{ min}^{-1}$), la producción de succinato es nula. Si bien para el modelo A la producción de succinato no era nula, en términos de productividad deseados, es despreciable el valor alcanzado de $6.69 \text{ gDW biomasa.min}^{-1}$. Además, el otro extremo del frente para el modelo B, el crecimiento celular es nulo, mientras la productividad del succinato es de $489.60 \text{ gDW biomasa.min}^{-1}$. Para ambos modelos, se observa una tendencia al incremento abrupto de la productividad del succinato a cambio de pequeñas disminuciones en el crecimiento celular del microorganismo.

Las variaciones ya mencionadas entre ambos modelos, y otras como las diferencia en las magnitudes de las funciones objetivo en los extremos de los frentes de Pareto, se corresponden a la cantidad de reacciones no contempladas en

el modelo B, donde solo se consideran las esenciales y primarias.

REFERENCES

- [1] Laerke Hansen, A.S., Lennen, R.M., Sonnenschein, N. & Herrgard, M.J. (2017). Systems biology solutions for biochemical production challenges. *Curr. Opin. Biotechnol.* 45: 85–91.
- [2] Nilsson, A. & Nielsen, J. (2017). Genome scale metabolic modeling of cancer. *Metabolic Engineering*, 43: 103 - 112.
- [3] Savinell, J.M. & Palsson B.Ø. (1992). Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism. *J. Theor. Biol.*, 154: 421–454.
- [4] Mavrotas G. (2009). Effective implementation of the ϵ -constraint method in multi objective mathematical programming problems *Applied Mathematics and Computation* 213(2), 455-465.