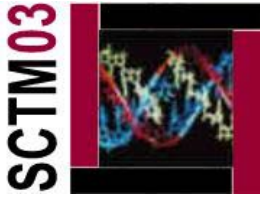


## **Modelización matemática y optimización de bioprocesos: Métodos y aplicaciones**



*Néstor V. Torres Darias*

Profesor Titular de Bioquímica y Biología Molecular  
Departamento de Bioquímica y Biología Molecular  
Universidad de La Laguna

### ***1. Introducción: El papel de los modelos matemáticos en las biociencias***

El objetivo de cualquier disciplina científica es la comprensión y el control del comportamiento de los sistemas objeto de estudio en términos de sus elementos constituyentes. Esta comprensión implica *mucho más que la mera descripción fenomenológica de las estructuras o los comportamientos*. En el caso de los sistemas biológicos, este objetivo es particularmente ambicioso y por ello difícil, dada la enorme complejidad estructural y de relaciones que éstos presentan.

Así planteado un proyecto de investigación científica constará necesariamente de dos etapas. La primera fase, etapa reduccionista, estará dirigida a la identificación y estudio individualizado de las estructuras constituyentes. A esta fase le debe seguir la etapa integradora, en la cual la información acumulada sobre esos elementos se integra en descripciones cuantitativas que nos describen el sistema y permiten no sólo comprender sino también controlar dicho sistema. La Historia de la Biología y la de las muchas subdisciplinas que comprende (entre las que no son excepción la Bioquímica, la Biomedicina y la Biotecnología) ha consistido básicamente en el desarrollo de la fase reduccionista. El desarrollo de esta fase previa se reveló desde el primer momento como una herramienta muy eficaz, y así lo prueba el progreso experimentado en los últimos cien años por la biociencias. Pero este desarrollo empieza a ser mucho más lento (cuando no a estancarse) en muchos ámbitos, lo que no es sino una manifestación de la aproximación reduccionista. Y, efectivamente, muy pronto resultó evidente que sólo el conocimiento de las propiedades individuales de los sistemas biológicos (enzimas, proteínas y ácidos nucleicos) *no bastaba* para comprender el comportamiento del conjunto, sino que se precisaba atender la etapa reconstruccionista, integradora del proyecto científico.

Hasta hace muy poco tiempo la integración de información en modelos comprensivos ha estado seriamente limitada por dos hechos, a saber: la falta de información detallada sobre los elementos constitutivos y la escasa disponibilidad de herramientas de cálculo suficientemente potentes, necesarias éstas para la integración y cuantificación del comportamiento de los sistemas. Sin embargo, en su mayor parte estas limitaciones han desaparecido en los últimos años. En algunas áreas de investigación (Bioquímica, Biotecnología o numerosas disciplinas biomédicas) la acumulación de datos sobre las bases moleculares de los sistemas ha experimentado un gran crecimiento. Esta disponibilidad de información unida a la ubicuidad y potencia de los ordenadores ha puesto de manifiesto que se dan las condiciones necesarias para acometer la fase integradora de la empresa científica. Es en este punto en el que un viejo conocido de la comunidad científica, el modelo matemático, adquiere un protagonismo sin precedentes en la Historia de la Biología. El modelo matemático es el único que permite una representación cuantitativa, precisa y absolutamente libre de interpretaciones subjetivas de la

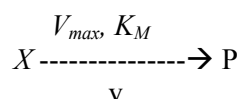
realidad. Por otra parte, la elaboración de un modelo fuerza a la organización y correcta ordenación de datos e ideas sobre los sistemas objeto de estudio. Finalmente, los modelos permiten predecir la evolución y dinámica de los sistemas frente a perturbaciones o condiciones ambientales cambiantes; es decir, podemos realizar con ellos experimentos que de otra manera no serían posibles.

¿Qué se necesita, pues, para construir modelos matemáticos útiles de los sistemas? Ya hemos indicado que los requerimientos básicos son la disponibilidad de datos sobre los mecanismos y elementos constituyentes y de herramientas de computación eficaces. A estas condiciones necesarias hay que sumarle la disponibilidad de un marco teórico que permita la representación matemática idónea y el análisis de los modelos elaborados. De hecho, la elección del tipo de representación es crítica para el éxito de modelo. Para el caso de los sistemas bioquímicos disponemos de algunas de tales representaciones, cada una con sus ventajas y limitaciones. De entre ellas, una que ha mostrado gran versatilidad y que permite realizar ejercicios de optimización que son prácticamente inabordables con otras formulaciones es la Teoría de los Sistemas Bioquímicos.

## II. La Teoría de los Sistemas Bioquímicos: Un marco teórico idóneo para la modelización matemática de los biosistemas

El tipo de representación que conocemos como Teoría de los Sistemas Bioquímicos (TSB) se basa en la posibilidad de expresar las ecuaciones de velocidad de las reacciones y otros procesos bioquímicos como un producto de funciones exponenciales. Matemáticamente este tipo de representación se apoya en el teorema de Taylor, que permite la aproximación de funciones con polinomios.

**La representación de potencias.** Consideremos el caso sencillo de una reacción enzimática que obedece al mecanismo de Michaelis-Menten:



La expresión de la velocidad de esta reacción adopta la forma:

$$v = v(X) = \frac{V_{max} X}{K_M + X}$$

donde X representa la concentración del sustrato y  $K_M$  y  $V_{max}$  las constantes de Michaelis-Menten y velocidad máxima, respectivamente. En el formalismo de potencias esta ecuación de velocidad se representa:

$$v \approx \alpha \cdot X^g$$

Es decir, la función original se aproxima mediante una función potencial con un exponente real y un factor positivo. El exponente, g, se denomina *orden cinético* y el

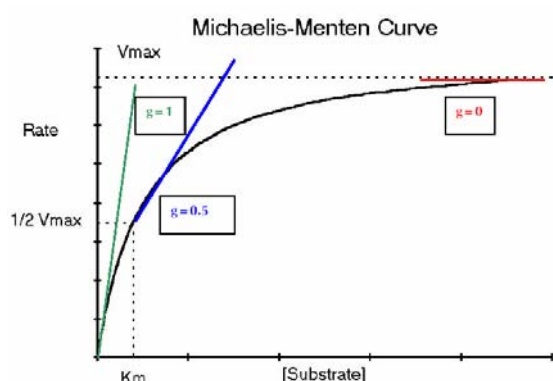
multiplicador,  $\alpha$ , constante de velocidad. El orden cinético,  $g$ , es la derivada del logaritmo de la función  $v$  con respecto al log de la variable  $X$ :

$$g = \frac{d \ln v}{d \ln X} = \frac{d v}{d X} \cdot \frac{X}{v}$$

evaluados en el punto de operación elegido. Una vez calculado el orden cinético, la constante de velocidad se obtiene como:

$$\alpha = \frac{v(X_0)}{(X_0)^g}$$

ya que la función y su aproximación deben ser iguales en el punto de operación,  $X_0$ . Así definido, las características numéricas de la aproximación dependen del punto de operación. Por ejemplo, el orden cinético  $g$  en el caso de la ecuación de Michaelis-Menten disminuye desde 1 a 0 si se elige un punto de operación con concentración más alta de sustrato. En particular, para el caso en que  $X_0 = K_M$  el orden cinético es 0.5.



Una característica fundamental de esta aproximación es que se puede generalizar a cualquier caso sin importar el número de elementos. Por ejemplo, si los intermediarios metabólicos que contribuyen a la velocidad son  $n$ ,  $V(X_1, X_2, X_3, X_4, \dots, X_n)$ , la expresión correspondiente en el formalismo de potencias es un producto de potencias de todos los metabolitos implicados:

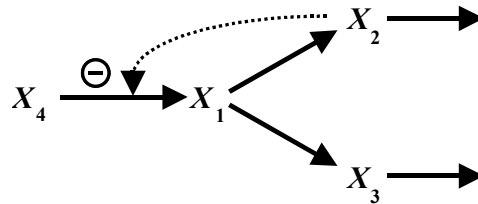
$$V = \alpha \cdot X_1^{g_1} X_2^{g_2} X_3^{g_3} X_4^{g_4} \dots X_n^{g_n} .$$

Sólo aquellas variables que influyen en la velocidad del proceso aparecen en la representación de potencias. Si una variable  $X_k$  no influye sobre  $V$ , la derivada parcial de  $V$  con respecto a  $X_k$  es cero. Por tanto,  $g_k$  es cero y la expresión  $X_k^{g_k}$  igual a 1, independientemente del valor de  $X_k$ . La estructura en forma de producto de la representación hace que todos aquellos factores del tipo  $X_k^{g_k} = 1$  no tengan efecto alguno y sean por tanto superfluos. Cada orden cinético  $g_i$  se puede calcular directamente por derivación *parcial* de  $V$  con respecto a  $X_i$  de la misma manera que se mostró arriba. Una vez que todos los ordenes cinéticos de un flujo o velocidad han sido calculados, la constante de velocidad se puede calcular basados en la equivalencia de la expresión de potencias y el valor de la velocidad en el punto de referencia elegido.

Ocurre que, mientras que los valores numéricos de los órdenes cinéticos y de las constantes de velocidad cambian con el punto de operación elegido, la estructura de la aproximación es siempre la misma, una función potencial del tipo  $\alpha X^g$ . Este hecho implica que

es posible construir un modelo sustentado en esta representación simbólicamente, sin necesidad de conocer la estructura exacta del proceso.

**Los modelos S-system.** Basándonos en la representación de potencias podemos desarrollar una descripción general de los sistemas que constituye la base fundamental de la TSB. Para ello consideremos un proceso (ruta metabólica) ramificado con la siguiente estructura:



En este sistema  $X_1$  se transforma en  $X_2$  y  $X_3$ . Además,  $X_2$  ejerce un efecto inhibitorio sobre la síntesis de  $X_1$ . Un modelo dinámico típico de este sistema consiste en un conjunto de ecuaciones diferenciales en las que cada una representa los cambios que experimenta una variable (positiva) dependiente  $X_i$ . El sistema incluye también variables independientes ( $X_4$  en este caso) que no cambian en cada situación pero que pueden adoptar distintos valores de un caso a otro.

La representación simbólica S-system de este sistema es de la siguiente forma:

$$\begin{aligned} \frac{dX_1}{dt} &= V_1^+ - V_1^- = \alpha_1 X_2^{g_{12}} X_4^{g_{14}} - \beta_1 X_1^{h_{11}} & X_1(0) &= \text{valor inicial de } X_1 \\ \frac{dX_2}{dt} &= V_2^+ - V_2^- = \alpha_2 X_1^{g_{21}} - \beta_2 X_2^{h_{22}} & X_2(0) &= \text{valor inicial de } X_2 \\ \frac{dX_3}{dt} &= V_3^+ - V_3^- = \alpha_3 X_1^{g_{31}} - \beta_3 X_3^{h_{33}} & X_3(0) &= \text{valor inicial de } X_3 \\ X &= \text{constante.} \end{aligned}$$

En la función positiva  $V_i^+$  se representan todos los elementos que influyen, aumentando o disminuyendo, la producción de  $X_i$ , mientras que la función, también positiva,  $V_i^-$  representa los elementos que influyen, aumentando o disminuyendo, la transformación de  $X_i$ . Una característica destacada de todas estas funciones es que tienen una *estructura homogénea*. Las variables, ya sean dependientes o independientes, son tratadas formalmente de la misma manera, a saber, como un factor en el producto  $V_i^+$  o  $V_i^-$ , elevado a una potencia. Este tipo de formulación es muy diferente de las expresiones de velocidad empleadas normalmente, que suelen adoptar la forma de funciones racionales en las que los sustratos y moduladores aparecen en diferentes formas.

Los S-systems pueden usarse para representar cualquier tipo de no-linealidad. Debido a la estructura rígida de estos modelos y a la flexibilidad que muestran para capturar no-linealidades, los modelos S-systems se denominan *modelos canónicos*. La ventaja más

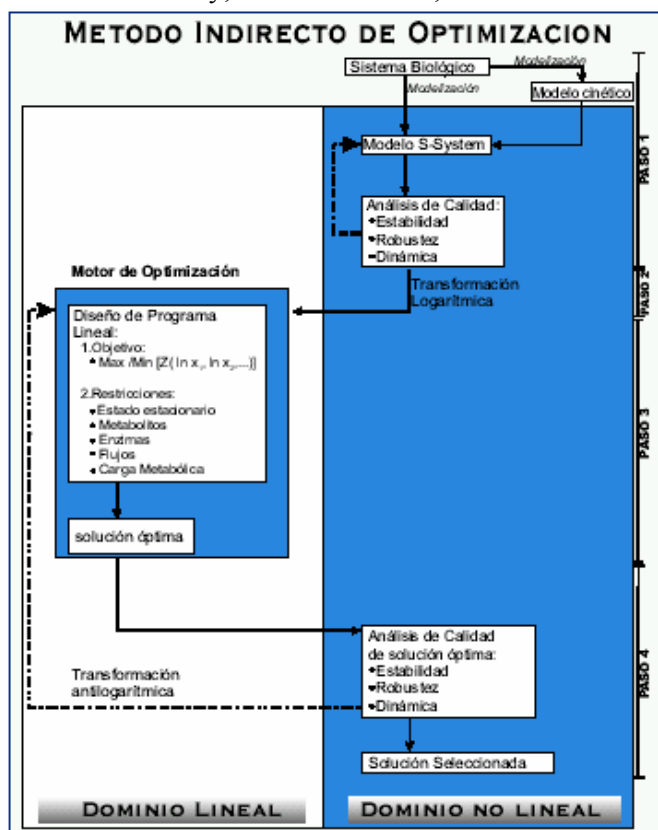
significativa de los modelos canónicos es el hecho de que su estructura homogénea permite la estandarización de los análisis del mismo, que incluyen el diseño del modelo, la estimación de parámetros, el análisis de los estados estacionarios, el análisis de sensibilidades, el estudio de las dinámicas frente a perturbaciones y la optimización.

**Optimización basada en modelos S-system.** Es en el capítulo de la optimización en el que los modelos presentados en la versión S-system presentan algunas de sus ventajas más significativas.

### III. El Método Indirecto de Optimización (MIO)

En su formulación más general, el MIO consiste en 4 pasos.

1. **Diseño del modelo, traslación a la versión S-system y evaluación de la calidad del mismo.** Tal como se indicó antes los modelos se formulan como un conjunto de ecuaciones diferenciales. Si el modelo no ha sido formulado en la versión S-system debe ser traducido a este formalismo y, basándose en él, sometido a los análisis de calidad correspondientes:



estabilidad de estado estacionario, análisis de sensibilidades y dinámica frente a perturbaciones. Este tipo de análisis puede hacerse de manera automática con el programa *PLAS*, diseñado específicamente para la elaboración y análisis de modelos en el formalismo de potencias. Si el modelo muestra una calidad suficiente estaremos en condiciones de proceder al siguiente paso.

2. **Linealización y programación lineal.** La ventaja clave de formular un modelo biológico como un S-system radica en que los estados estacionarios en esta representación vienen descritos por un conjunto de ecuaciones algebraicas lineales cuando se expresan en coordenadas logarítmicas. De

la misma manera, las funciones susceptibles de optimización (funciones objetivo) y las restricciones a los flujos e intermediarios necesarias para garantizar la viabilidad biológica de los sistemas se pueden expresar también como ecuaciones lineales en escala logarítmica. De este modo, la optimización de un sistema no lineal se transforma en un problema de

optimización lineal. La formulación de un problema de optimización lineal adoptaría la siguiente forma:

*Programa lineal*

1. Maximizar  $\ln(\text{flujo})$

Sujeto a las siguientes restricciones:

2. Condición de estado estacionario expresada en término del logaritmo de las variables
3. Límites a las variaciones de las variables:  $\ln(\text{variable}) \leq, \geq, = \text{constante}$
4. Límites a las variaciones de los flujos:  $\ln(\text{flujos}) \leq, \geq, = \text{constante}$

Esta optimización puede hacerse con alguno de los muchos programas de optimización lineal disponibles en el mercado (*LINDO PC*, de *LINDO Systems*).

**3. Transferencia de los resultados al modelo original y análisis de la calidad del modelo.**

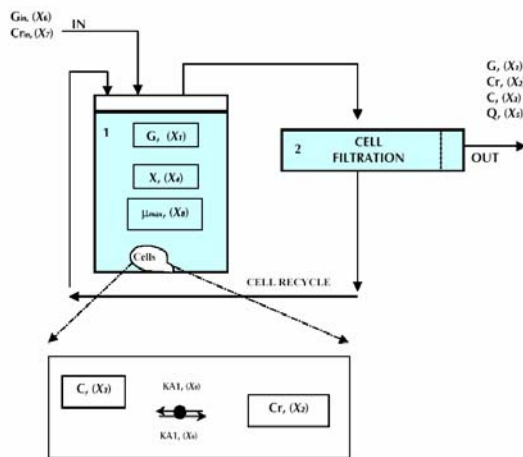
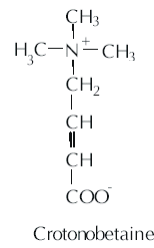
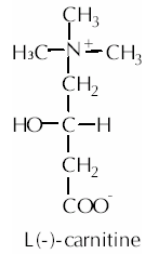
Las soluciones óptimas que se obtienen y sus correspondientes estados estacionarios deben ser evaluadas con respecto a su estabilidad y calidad. Si alguna de las soluciones mostrara ser inestable o no cumpliera con los requisitos mínimos de calidad deberá ser descartada y el proceso de optimización revisado con nuevas restricciones.

**4. Implementación de la solución óptima en el sistema experimental.** La comparación de los resultados obtenidos en el sistema real con las predicciones confirmaría o refutaría la solidez del modelo y del método de optimización.

En lo que se sigue se ilustrará la aplicación del MIO a la mejora de la producción de L-(-)-carnitina a partir de crotonobetaína en un biorreactor en régimen continuo con reciclaje celular.

**IV. Mejora de la producción de L-(-)-carnitina por *E. Coli***

La L-(-)-carnitina [R-(-)-3-hidroxi-4-N-trimetilaminobutirato] es un compuesto quiral, fuertemente polar ampliamente distribuido en la naturaleza y con un amplio rango de aplicaciones médicas. Actualmente, gran parte de la producción de L-carnitina se obtiene por síntesis química, con el inconveniente de producir una mezcla racémica que es preciso y costoso separar. Por ello, y ante el incremento de la demanda de L-carnitina, se ha visto estimulada la búsqueda de métodos para la síntesis de formas ópticamente puras por medio de la biotransformación de precursores aquirales baratos, como es el caso de la crotonobetaína. En esta línea se ha obtenido una cepa de *E. coli* (O44 K74), excelente productora de L-carnitina, y se han caracterizado varias de las enzimas implicadas en la biosíntesis de L-carnitina. Esta cepa ha servido de base para la puesta a punto de un sistema de bioproducción de L-(-)-carnitina en un biorreactor en régimen continuo con reciclaje celular.



**Paso 1. Diseño del modelo, traslación a la versión S-system y evaluación de la calidad del mismo.** El esquema adjunto muestra esquemáticamente el diseño experimental empleado para la biotransformación de la crotonobetaina en L-carnitina mediante *E. coli* O44 K74 en un biorreactor con recirculación celular. Las flechas aparecen junto a los parámetros del modelo. A partir de un modelo previo se obtuvo la versión S-system del mismo con las siguientes equivalencias:

$$\frac{dG/dt = [Q \cdot G_{in}] - \left[ Q \cdot G + \frac{u \max g \cdot G \cdot \left( K_{i1} \cdot 2.3^{-\frac{G}{K_{i2}}} + K_{i2} \right) \cdot X}{(K_{g01} + K_{g02} + G) \cdot \left( Y_{xg01} + Y_{xg1} \cdot e^{-\frac{G}{K_{i2}}} \right)} \right] \cdot X}{=} \alpha_1 X_5^{g1.5} X_6^{g1.6} - \beta_1 X_1^{h1.1} X_4^{h1.4} X_5^{h1.5} X_8^{h1.8} = \frac{dX_1}{dt}$$

$$\frac{dCr/dt = \left[ \frac{X \cdot (K_{A1} + K_{A3}) \cdot C \cdot V_2}{K_2 + C} + Q \cdot Cr_{in} \right] - \left[ \frac{X \cdot (K_{A1} + K_{A3}) \cdot Cr \cdot V_1}{K_1 + Cr} + Q \cdot Cr \right]}{=} \alpha_2 X_3^{g2.3} X_4^{g2.4} X_5^{g2.5} X_7^{g2.7} X_9^{g2.9} - \beta_2 X_2^{h2.2} X_4^{h2.4} X_5^{h2.5} X_9^{h2.9} = \frac{dX_2}{dt}$$

$$\frac{dC/dt = \left[ \frac{X \cdot (K_{A1} + K_{A3}) \cdot Cr \cdot V_1}{K_1 + Cr} \right] - \left[ \frac{X \cdot (K_{A1} + K_{A3}) \cdot C \cdot V_2}{K_2 + C} + Q \cdot C \right]}{=} \alpha_3 X_2^{g3.2} X_4^{g3.4} X_9^{g3.9} - \beta_3 X_3^{h3.3} X_4^{h3.4} X_5^{h3.5} X_9^{h3.9} = \frac{dX_3}{dt}$$

$$\frac{dX/dt = \left[ \frac{u \max g \cdot G \cdot \left( K_{i1} \cdot 2.3^{-\frac{G}{K_{i2}}} + K_{i2} \right) \cdot X}{K_{g01} + K_{g02} + G} \right] - [u \cdot X]}{=} \alpha_4 X_1^{g4.1} X_4^{g4.4} X_8^{g4.8} - \beta_4 X_4^{h4.4} = \frac{dX_4}{dt}$$

Los órdenes cinéticos y constante de velocidad se calcularon a partir de datos experimentales, obteniéndose los siguientes resultados:

$$\frac{dX_1}{dt} = X_5 X_6 - 1.0213 \cdot X_1^{0.9162} X_4^{0.5675} X_5^{0.4324} X_8^{0.5675}$$

$$\frac{dX_2}{dt} = 0.2438 X_3^{0.3538} X_4^{0.9394} X_5^{0.0605} X_7^{0.0605} X_9^{0.9292} - 0.464 \cdot X_2^{0.1424} X_4^{0.9643} X_5^{0.0356} X_9^{0.9537}$$

$$\frac{dX_3}{dt} = 0.3844 \cdot X_2^{0.1106} X_4 X_9^{0.989} - 0.2058 \cdot X_3^{0.3926} X_4^{0.9742} X_5^{0.0257} X_9^{0.9636}$$

$$\frac{dX_4}{dt} = 0.0053 \cdot X_1^{0.8524} X_4 X_8 - 0.08 \cdot X_4$$

*Estabilidad del estado estacionario.* El estado estacionario mostró ser estable. Éste se recuperaba después de perturbaciones macroscópicas. La observación experimental de este comportamiento fue confirmada por la evaluación de los valores propios (-2.0737;  $-4.31 \cdot 10^{-2}$ ; -18.2467; -1).

*Análisis de sensibilidades.* Las sensibilidades del sistema a cambios en los parámetros miden su capacidad para resistir pequeñas fluctuaciones en los mismos. Se definen como:

$$S(X_i, \alpha_j) = \left( \frac{\partial X_i}{\partial \alpha_j} \frac{\alpha_j}{X_i} \right)_0 = \frac{\partial(\log X_i)}{\partial(\log \alpha_j)}$$

$$S(X_i, g_{jk}) = \left( \frac{\partial X_i}{\partial g_{jk}} \frac{g_{jk}}{X_i} \right)_0 = \frac{\partial(\log X_i)}{\partial(\log g_{jk})}$$

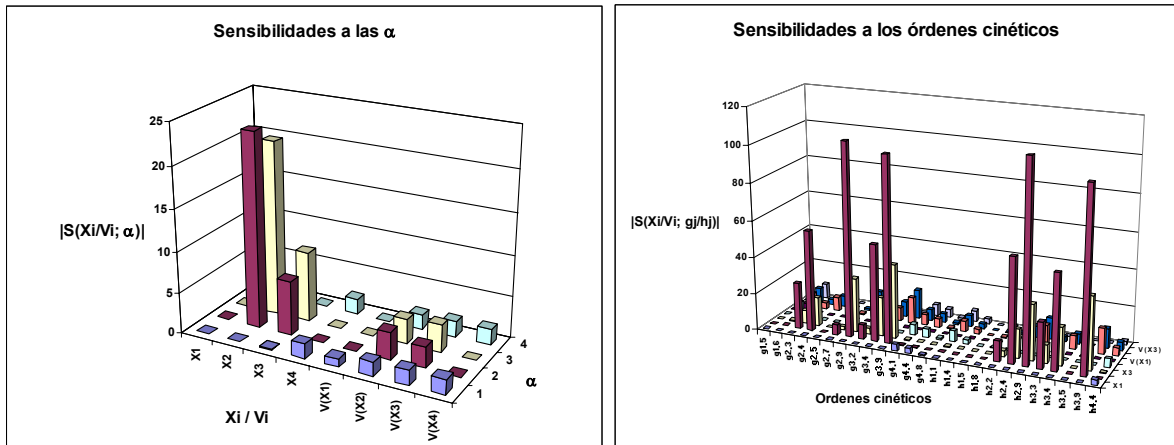
$$S(V_i, \alpha_j) = \left( \frac{\partial V_i}{\partial \alpha_j} \frac{\alpha_j}{V_i} \right)_0 = \frac{\partial(\log V_i)}{\partial(\log \alpha_j)}$$

$$S(V_i, g_{jk}) = \left( \frac{\partial V_i}{\partial g_{jk}} \frac{g_{jk}}{V_i} \right)_0 = \frac{\partial(\log V_i)}{\partial(\log g_{jk})}$$

En todos los casos, si las sensibilidades son mayores que 1 implica la amplificación de la alteración original, mientras que si es menor que 1 es una atenuación. Un valor positivo indica que los cambios se producen en la misma dirección, ya sea un aumento o una disminución. Un signo negativo indica cambios en la dirección opuesta.

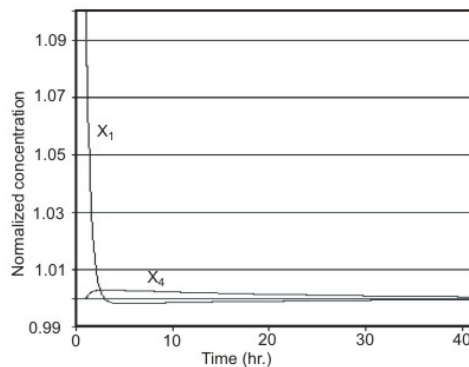
En un buen modelo las sensibilidades a los parámetros son pequeñas. Si éstas fueran muy grandes significaría que el modelo está mal determinado, y requeriría la revisión de las partes del modelo afectadas. Lo observado en nuestro modelo se muestra en las dos figuras siguientes. Las sensibilidades  $S(V(X_j), \alpha_j)$  son pequeñas en general, siendo las únicas excepciones aquellas asociadas con los flujos a través de la crotonobetaína ( $X_2$ ) y carnitina ( $X_3$ ).





Por otro lado, las sensibilidades a los órdenes cinéticos tienen en su mayor parte (84 de 104, 80.8%) valores entre 0 y 10. Los valores más altos son los asociados con las concentraciones de crotonbetaína y carnitina ( $X_2$  y  $X_3$ ) y con la carnitina deshidratasa ( $X_9$ ). Estos resultados sugieren que debe tenerse especial cuidado con la valoración experimental de los órdenes cinéticos afectados ( $g_{2,4}$ ,  $g_{2,9}$ ,  $g_{3,4}$  y  $g_{3,9}$ ), ya que un error en estas valoraciones conduciría a la amplificación de los mismos en el modelo.

*Dinámica.* Como parte de la evaluación de la calidad examinamos la respuesta frente a un incremento del 10% en la variable dependiente  $X_1$  (concentración inicial de glicerol). En la figura siguiente se muestra cómo el sistema retorna rápidamente al estado estacionario original con una variación máxima en las otras variables dependientes de menos de un 1%, sin que se produzcan oscilaciones. Este comportamiento se considera aceptable en un sistema biotecnológico.



**Paso 2. Linealización y programación lineal.** El primer paso en la formulación de un programa de optimización lineal es la definición de la función objetivo.

*Función objetivo.* La carnitina disponible se genera en el biorreactor a una velocidad  $V_c = Q \cdot C$ , que en términos de las variables se expresa como:

$$V_c = X_3 \cdot X_5$$

Esta función expresada en coordenadas logarítmicas se transforma en:

$$y_3 + y_5$$

donde  $y_i$  es  $\ln(X_i)$  para  $i = 3, 5$ .

*Restricciones de estado estacionario.* La condición de estado estacionario una vez linealizada en coordenadas logarítmicas es:

$$\begin{aligned} -0.9162 y_1 - 0.5675 y_4 + 0.5675 y_5 + y_6 - 0.5675 y_8 &= 0.0211 \\ -0.1424 y_2 + 0.3538 y_3 - 0.0248 y_4 + 0.0248 y_5 + 0.0605 y_7 - 0.0245 y_9 &= 0.6433 \\ 0.1106 y_2 - 0.3926 y_3 + 0.0257 y_4 - 0.0257 y_5 + 0.0254 y_9 &= -0.6246 \\ 0.8524 y_1 + y_8 &= 2.706 \end{aligned}$$

donde  $y_i$  es  $\ln(X_i)$  para  $i = 1, 2, 3, \dots, 9$ .

*Restricciones a las variables dependientes e independientes.* En este punto definimos la magnitud en la que se permite variar a las variables y parámetros del modelo. En general, el rango de variación será entre 0.5 y 1.5 veces los valores basales (excepto para  $X_4$ , la biomasa, a la que se le permite aumentar hasta 2 veces el valor basal). De acuerdo con esto obtenemos:

3.0746	≤	$y_1$	≤	4.1732
2.6909	≤	$y_2$	≤	3.7895
2.3277	≤	$y_3$	≤	3.4263
1.8068	≤	$y_4$	≤	3.1931
-0.6931	≤	$y_5$	≤	0.4054
3.9120	≤	$y_6$	≤	5.0106
3.2188	≤	$y_7$	≤	4.3174
-1.1988	≤	$y_8$	≤	-0.1002
4.1214	≤	$y_9$	≤	6.424

**Paso 3. Transferencia de los resultados al modelo original y análisis de la calidad del modelo.** En la Tabla siguiente se muestran los resultados de la búsqueda sistemática, paso a paso, de las mejores soluciones. Junto con la solución obtenida en el modelo S-system se muestran las obtenidas cuando éstas se trasladaron al modelo original (K-M).

Variables		$(X_i)_{opt}/(X_i)_{basal}$									
		1		2		3		4		5	
Dependientes	Basal	S-s	K-M	S-s	K-M	S-s	K-M	S-s	K-M	S-s	K-M
$X_1$ (G)	43.28 mM	1	1	1	1	1	1	0.68	0.69	0.62	0.69
$X_2$ (Cr)	29.49 mM	1	1	1.5	1.49	1.5	1.53	1.5	1.56	1.5	1.55
$X_3$ (C)	20.51 mM	1	1	1.12	1.11	1.24	1.18	1.26	1.19	1.26	1.19
$X_4$ (X)	12.18 g/L	1.5	1.5	1.5	1.5	1.5	1.5	2	1.85	2	1.85
Independientes											
$X_5$ (Q)	1 L/h <sup>-1</sup>	1.5		1.5		1.5		1.5		1.5	
$X_6$ (G <sub>in</sub> )	100 mM	1		1		1		1		0.95	
$X_7$ (Cr <sub>in</sub> )	50 mM	1		1.33		1.39		1.4		1.4	
$X_8$ ( $\mu_{max}$ )	0.603 h <sup>-1</sup>	1		1		1		1.38		1.5	
$X_9$ (KA1)	123.3 U h L <sup>-1</sup>	1		1		5		5		5	
<b>Veloc. Producción</b>	21.1 mM/h <sup>-1</sup>	1.5	1.5	1.68	1.65	1.86	1.77	1.9	1.78	1.9	1.78

Tabla 1.

Se observa que hay un buen acuerdo entre las soluciones S-system y las que rinde el modelo original, y siempre se respetan las restricciones impuestas. Por otra parte, todas las soluciones corresponden a estados estacionarios estables con sensibilidades dentro de rangos aceptables.

De acuerdo con los perfiles óptimos que se muestran en la Tabla 1, el primer parámetro que se debe modificar es  $X_5$  (la velocidad de extracción del biorreactor, Q). El aumento de la misma en un 50% causa un aumento similar en la velocidad de la producción de carnitina. El siguiente parámetro a modificar es la concentración de crotonobetaina ( $Cr_{in}$ ),  $X_7$ . Si aumentamos simultáneamente la velocidad de reciclaje en un 50% y la concentración de crotonobetaina en un 40%, la velocidad de producción de carnitina aumenta en un 68%. Estos parámetros son directamente accesibles en el biorreactor. La siguiente mejora se obtiene amplificando la actividad de la carnitina deshidratasa (KA1),  $X_9$ , 5 veces el valor basal. En este caso, la modificación simultánea de las tres parámetros provoca un aumento de la productividad del 86%, pero un incremento en  $\mu_{max}$  ( $X_8$ ) no produce efectos apreciables.

**Paso 4. Implementación de la solución óptima en el sistema experimental.** Los perfiles óptimos mostrados en la Tabla 1 se implementaron en el biorreactor objeto de estudio. La Tabla 2 muestra que los resultados experimentales coincidieron prácticamente con las predicciones teóricas, tanto en lo referente a los estados estacionarios observados experimentalmente como en lo referente a las velocidades de producción de carnitina.

Variables		$(X_i)_{opt}/(X_i)_{basal}$			
		1		2	
Dependientes	Basal	K-M	Experimental	K-M	Experimental
$X_1$ (G)	43.28 mM	1	0.96	1	0.95
$X_2$ (Cr)	29.49 mM	1	0.97	1.49	1.77
$X_3$ (C)	20.51 mM	1	1.01	1.11	1.13
$X_4$ (X)	12.18 g/L	1.5	1.53	1.5	1.53
<b>Independientes</b>					
$X_5$ (Q)	1 L/h <sup>-1</sup>		1.5		1.5
$X_7$ ( $Cr_{in}$ )	50 mM		1	1.33	1.5
<b>Veloc. Producción</b>	21.1 mM/h <sup>-1</sup>	1.5	1.54	1.65	1.74

Tabla 2.

### V. Conclusiones

A lo largo de esta exposición se ha ilustrado, con un caso de interés biotecnológico, el recorrido completo que va desde la observación de un biosistema, la integración de dichas observaciones en un modelo matemático del sistema, el análisis y evaluación de la calidad de dicho modelo, la predicción, basada en dicho modelo, de los cambios que conducen a la mejora del mismo basados en la optimización del modelo y, finalmente la verificación experimental de las predicciones del mismo.

Las Ciencias Biológicas son ya las más matemáticas e interdisciplinarias de todas las disciplinas. La llamada a las matemáticas surge como consecuencia del elevado número de elementos e interacciones no lineales que están presentes en ellos. Su naturaleza hace imposible la evaluación cuantitativa y el seguimiento de la evolución de las variables por otras vías distintas del cálculo. Por otra parte, esa misma complejidad resiste cualquier abordaje unidimensional. Se requieren las aportaciones de distintas disciplinas, cada una de ellas centrada en el estudio del mismo sistema, pero desde ópticas metodológicas diferentes. Así, la Genética Molecular nos informa sobre los mecanismos básicos implicados en la replicación de los genes, su transducción en proteínas y sobre los sistemas de control y regulación de estos procesos; la Ingeniería Genética desarrolla técnicas que permiten la introducción de genes en los organismos; la Genómica, por su parte, suministra información sobre los genes que se expresan en un organismo en un momento dado; y la Bioinformática organiza y explica los datos genómicos. La aproximación matemática al estudio de los biosistemas es la que permite integrar toda esta información y dar coherencia y racionalidad a la aplicación de las técnicas de manipulación desarrolladas por todas estas disciplinas. Con ellas podemos calcular y predecir la configuración metabólica óptima para la biosíntesis de un producto de interés que posteriormente podremos implementar en los sistemas u organismos con las herramientas que la Ingeniería Genética nos suministra.

El potencial de esta aproximación interdisciplinar es enorme. Por otra parte, la necesaria colaboración entre distintas subespecialidades científicas sugiere fuertemente un inminente cambio de paradigma científico. Este nuevo paradigma, sin duda, dirigirá la evolución de las biociencias en este siglo.

### **Bibliografía**

- F. Alvarez-Vasquez, M. Cánovas, J.L. Iborra, N.V. Torres: Modelling And Optimization Of Continuous L-(-)-Carnitine Production By High-Density *Escherichia Coli* Cultures. *Biotechnology and Bioengineering* **80**, no. 7 (2002), 794-805.
- N.V. Torres, E.O. Voit: *Pathway Analysis and Optimization in Metabolic Engineering*. Cambridge University Press, 2003.

### **En Internet**

<http://webpages.ull.es/users/sympbst>

*Grupo de Tecnología Bioquímica de la Universidad de La Laguna*

<http://www.udl.es/usuarios/q3695988/WebPL/main.htm>

*The Power-Law Formalism*

<http://correio.cc.fc.ul.pt/~aenf/plas.html>

*PLAS (Power-Law Analysis and Simulation)*

## MODELIZACIÓN MATEMÁTICA Y OPTIMIZACIÓN DE BIOPROCESOS: MÉTODOS Y APLICACIONES



Grupo de Tecnología Bioquímica y Control Metabólico  
Departamento de Bioquímica y Biología Molecular

Néstor V. Torres Darías  
Universidad de La Laguna



1

## LA MODELIZACIÓN DE BIOPROCESOS



Bioquímica/Biotecnología/Biomedicina → comprensión y control de los procesos biológicos a nivel molecular

*Reduccionismo*: comprensión del comportamiento macroscópico a partir del conocimiento de las estructuras y propiedades de sus constituyentes microscópicos



Útil, pero insuficiente: "el conocimiento de las propiedades individuales de las enzimas, proteínas y ácidos nucleicos *no basta* para comprender el comportamiento del conjunto"

2

## LA MODELIZACIÓN DE BIOPROCESOS



Modelización matemática: herramienta imprescindible en la fase reconstructiva de un proyecto científico

Beneficios de los modelos:

- La representación más precisa de la realidad: interpretación única, objetiva y cuantitativa
- Permiten organizar y almacenar la información e ideas sobre los sistemas objeto de estudio (muchos elementos y complejas interacciones)
- Permite la predicción de la dinámica de los sistemas
- Permiten la realización de experimentos "in número" y la exploración de escenarios posibles.

3

## LA MODELIZACIÓN DE BIOPROCESOS



Los datos son información si conocemos los procesos implicados.

La información es conocimiento cuando podemos ver cómo opera el sistema.

Pero el conocimiento sólo llega a ser verdadera comprensión cuando podemos predecir la evolución de cualquier sistema y cómo estos se comportan en la realidad.

Peter Allen (1994)

4

## LA MODELIZACIÓN DE BIOPROCESOS



¿Qué se necesita para construir modelos *útiles* de biosistemas?

- Datos sobre los mecanismos y elementos constituyentes
- Un marco que permita la representación matemática idónea y el análisis de los modelos
- Herramientas de computación eficaces



La elección del tipo de representación es crítica

5

## LA MODELIZACIÓN DE BIOPROCESOS



1. La Teoría de los Sistemas Bioquímicos: un marco teórico idóneo para la modelización matemática de los biosistemas.

1.1 El formalismo de leyes de potencias

1.2 La representación S-system

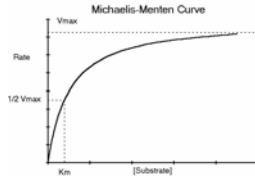
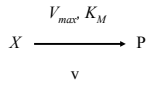
2. Mejora de la producción de L-(-)-carnitina por *E. coli*.

3. Optimización de la producción de L-(-)-carnitina basada en la representación S-systems. El Método Indirecto de Optimización.

3. Conclusiones.

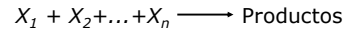
6

## 1.1 El formalismo de las leyes de potencias



$$v = v(X) = \frac{V_{max} X}{K_M + X} \quad \rightarrow \quad v \approx \alpha \cdot X^g$$

7



$V(X_1, X_2, X_3, X_4, \dots, X_n)$



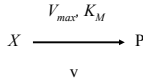
$$V = \alpha \cdot X_1^{g_1} X_2^{g_2} X_3^{g_3} X_4^{g_4} \dots X_n^{g_n}$$

$$\alpha_i = V \cdot X_1^{-g_1} X_2^{-g_2} \dots X_n^{-g_n}$$

La estructura de la aproximación es siempre la misma,  $\alpha X^g$   
Permite la formulación simbólica del modelo sin necesidad de conocer la el mecanismo exacto de los procesos cinéticos implicados

8

$$v \approx \alpha \cdot X^g$$



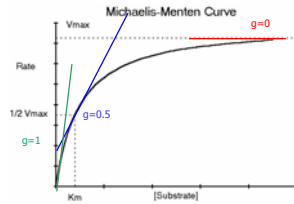
Calculados en el estado estacionario de referencia

$$\alpha = \frac{v(X_0)}{(X_0)^g}$$

Constante de velocidad

$$g = \frac{d \ln v}{d \ln X} = \frac{d v}{d X} \cdot \frac{X}{v}$$

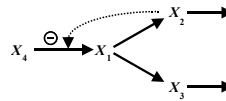
Orden cinético



$0 < g < 1$ ; si  $X_0 = K_m$ ,  $g = 0.5$

9

## 1.2 La representación S-system



Estructura homogénea pero flexible



Modelos canónicos

$$\frac{dX_1}{dt} = V_1^+ - V_1^- = \alpha_1 X_2^{g_{21}} X_4^{g_{41}} - \beta_1 X_1^{h_{11}}$$

$$\frac{dX_2}{dt} = V_2^+ - V_2^- = \alpha_2 X_1^{g_{12}} - \beta_2 X_2^{h_{22}}$$

$$\frac{dX_3}{dt} = V_3^+ - V_3^- = \alpha_3 X_1^{g_{13}} - \beta_3 X_3^{h_{33}}$$

10

## Modelos Canónicos

Estandarización del proceso de modelización

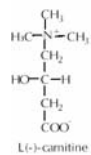
- Diseño del modelo
- Estimación de parámetros y representación numérica
- Evaluación de la calidad del modelo
  - ✓ Estabilidad del estado estacionario
  - ✓ Análisis de sensibilidades: Robustez y Control
  - ✓ Dinámicas



Optimización del modelo

11

## 2. Producción de L(-)-carnitina por *E. coli*



Factor esencial para el transporte de ácidos grasos de cadena larga a través de la membrana mitocondrial interna

Aplicaciones médicas:

- Disfunciones del músculo esquelético
- Hipoglucemia Hemodialisis crónica
- Isquemia cardiaca y otras cardiopatías

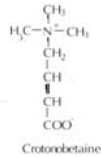
Estimulación del crecimiento de cultivos bacterianos y de levadura



Aumento de la demanda

12

## 2. Producción de L(-)-carnitina por *E. coli*



Creciente demanda de L-carnitina

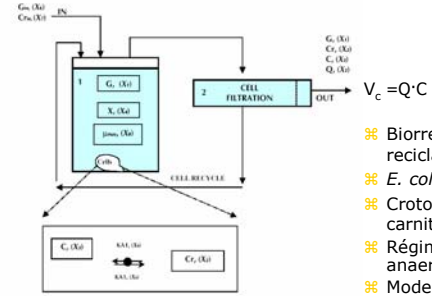
Obtención por biotransformación a partir de precursores baratos,  
accesibles  
aquirales → crotonobetaina

### Elementos biotecnológicos

- *E. coli* O44 K74: la mejor cepa productora de L-carnitina
- Se conoce el metabolismo de la L(-)-carnitina en *E. coli*
- L(-)-carnitina deshidratasa y crotonobetaina reductasa: inducidas anaerobicamente en presencia de crotonobetaina

13

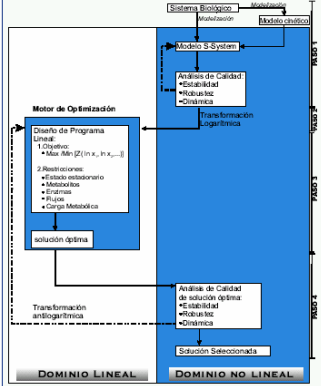
## 2. Producción de L(-)-carnitina por *E. coli*



- ⌘ Biorreactor con reciclaje celular
- ⌘ *E. coli* O44 K74
- ⌘ Crotonobetaina en L-carnitina
- ⌘ Régimen continuo anaeróbico a 37 °C
- ⌘ Modelo previo

14

### METODO INDIRECTO DE OPTIMIZACION



### 3. Optimización de la producción de L(-)-carnitina. El Método Indirecto de Optimización.

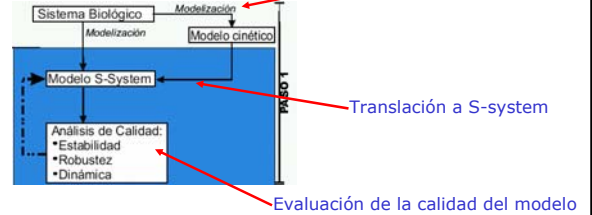
1. Modelización, traslación a la forma S-system y evaluación del modelo
2. Transformación logarítmica y programación lineal
3. Transferencia de la solución óptima al modelo original
4. Implementación en el sistema experimental

15

### Método Indirecto de Optimización

#### Paso 1

Ámbito no-lineal



16

### Método Indirecto de Optimización

#### Paso 1

#### Modelización y traslación a S-system

$$dG/dt = [Q - G_{in}] - \left[ Q - G + \frac{u \max_g \cdot G \cdot \left( K_{i1} \cdot 2.3^{-\frac{G}{K_{i1}}} + K_{i2} \right) \cdot X}{(K_{g0} + G + K_{g0} \cdot 2 + G) \cdot (Y_{gg0} + Y_{gg1} \cdot e^{-\frac{G}{K_{g1}}})} \right] = \alpha_1 \cdot X^{2.5} \cdot Y_1^{2.5} \cdot Y_2^{2.5} \cdot Y_3^{2.5} \cdot Y_4^{2.5} - \beta_1 \cdot X^{2.5} \cdot Y_1^{2.5} \cdot Y_2^{2.5} \cdot Y_3^{2.5} \cdot Y_4^{2.5} = \frac{dX_1}{dt}$$

$$dC/dt = \left[ \frac{X \cdot (K_{d1} + K_{d3}) \cdot C \cdot J^2}{K_2 + C} + Q \cdot Cr_{in} \right] - \left[ \frac{X \cdot (K_{d1} + K_{d3}) \cdot Cr \cdot J^1}{K_1 + Cr} + Q \cdot Cr \right] = \alpha_2 \cdot X^{2.5} \cdot Y_1^{2.5} \cdot Y_2^{2.5} \cdot Y_3^{2.5} \cdot Y_4^{2.5} - \beta_2 \cdot X^{2.5} \cdot Y_1^{2.5} \cdot Y_2^{2.5} \cdot Y_3^{2.5} \cdot Y_4^{2.5} = \frac{dX_2}{dt}$$

$$dC/dt = \left[ \frac{X \cdot (K_{d1} + K_{d3}) \cdot Cr \cdot J^1}{K_1 + Cr} \right] - \left[ \frac{X \cdot (K_{d1} + K_{d3}) \cdot C \cdot J^2}{K_2 + C} + Q \cdot C \right] = \alpha_3 \cdot X^{2.5} \cdot Y_1^{2.5} \cdot Y_2^{2.5} \cdot Y_3^{2.5} \cdot Y_4^{2.5} - \beta_3 \cdot X^{2.5} \cdot Y_1^{2.5} \cdot Y_2^{2.5} \cdot Y_3^{2.5} \cdot Y_4^{2.5} = \frac{dX_3}{dt}$$

$$dX/dt = \left[ \frac{u \max_g \cdot G \cdot \left( K_{i1} \cdot 2.3^{-\frac{G}{K_{i1}}} + K_{i2} \right) \cdot X}{K_{g0} + K_{g0} \cdot 2 + G} \right] - [u \cdot X] = \alpha_4 \cdot X^{2.5} \cdot Y_1^{2.5} \cdot Y_2^{2.5} \cdot Y_3^{2.5} \cdot Y_4^{2.5} - \beta_4 \cdot X^{2.5} = \frac{dX_4}{dt}$$

17

### Método Indirecto de Optimización

#### Paso 1

Evaluación de la calidad del modelo

- Estabilidad local del estado estacionario
- Robustez
- Análisis de la dinámica

#### Estabilidad local del estado estacionario

Autovalores con partes reales negativas:

-2.0737; -4.31 · 10<sup>-2</sup>; -18.2467; -1

Sin oscilaciones

18

## Evaluación de la calidad del modelo

### Robustez

Medida de la resistencia a cambios estructurales

Sensibilidades a las  $\alpha$  y órdenes cinéticos

$$S(X_i, \alpha_j) = \left( \frac{\partial X_i}{\partial \alpha_j} \right)_0 = \frac{\partial(\log X_i)}{\partial(\log \alpha_j)} \quad S(X_i, g_{jk}) = \left( \frac{\partial X_i}{\partial g_{jk}} \right)_0 = \frac{\partial(\log X_i)}{\partial(\log g_{jk})}$$

$$S(V_i, \alpha_j) = \left( \frac{\partial V_i}{\partial \alpha_j} \right)_0 = \frac{\partial(\log V_i)}{\partial(\log \alpha_j)} \quad S(V_i, g_{jk}) = \left( \frac{\partial V_i}{\partial g_{jk}} \right)_0 = \frac{\partial(\log V_i)}{\partial(\log g_{jk})}$$

Atenuación  $\leftarrow 1 > \text{Sensibilidad} > 1 \rightarrow$  Amplificación

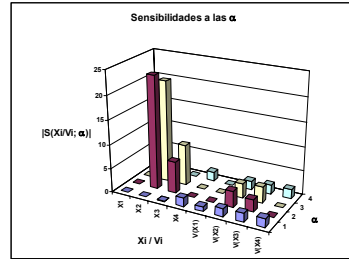
Dirección opuesta  $\leftarrow (-) \text{Sensibilidad} (+) \rightarrow$  Misma dirección

En un buen modelo las sensibilidades son pequeñas

19

## Evaluación de la calidad del modelo

### Robustez



Variables y flujos más sensibles:

Crotonobetaína ( $X_2$ )

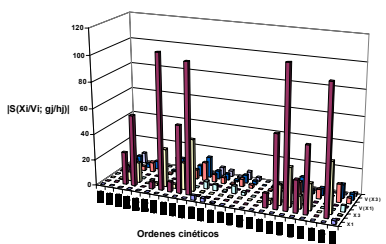
Carnitina ( $X_3$ )

20

## Evaluación de la calidad del modelo

### Robustez

#### Sensibilidades a los órdenes cinéticos



80% entre 0-10

Crotonobetaina ( $X_2$ )

Carnitina ( $X_3$ )

y carnit. desh. ( $X_9$ )

Precaución con:

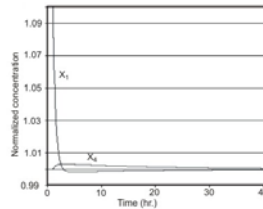
$g_{2,4}, g_{2,9}, g_{3,4}, g_{3,9}$

$\rightarrow$  amplificación error experimental

21

## Evaluación de la calidad del modelo

### Análisis de la dinámica



Necesario para identificar problemas de consistencia y exactitud del modelo

Aumento del 10% en  $X_1$   
(concentración inicial de glicerol)



El sistema retorna rápidamente al estado estacionario con variaciones inferiores al 1% en el valor de las variables, sin oscilaciones

22

## Evaluación de la calidad del modelo

### Valoración global

El comportamiento observado en términos de estabilidad del estado estacionario, robustez y dinámica frente a perturbaciones



Descripción realista y aceptable del sistema experimental



Base fiable para estudios de optimización

23

## Método Indirecto de Optimización

### Paso 2: Linearización y Programación Lineal

Optimización en el dominio no-lineal: enormes dificultades

Optimización en el dominio lineal: accesible (programación lineal)

La ventaja clave de los modelos S-system es que los estados estacionarios vienen descritos por un conjunto de *ecuaciones lineales* cuando se expresan en coordenadas logarítmicas

Igualmente las respuestas susceptibles de optimización (funciones objetivo) y las restricciones a los flujos y variables (necesarias para garantizar la viabilidad de los sistemas) se pueden expresar como *ecuaciones lineales* en escala logarítmica.



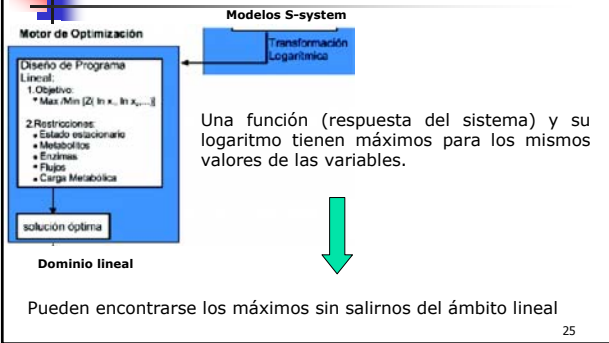
La optimización de un sistema no lineal se transforma en un problema de optimización lineal

24



## Método Indirecto de Optimización

### Paso 2: Linearización y Programación Lineal



25

## Método Indirecto de Optimización

### Paso 2: Linearización y Programación Lineal

#### Programación lineal

Maximizar  $\ln(\text{flujo})$

Sometido a las siguientes restricciones

1. Condición de estado estacionario expresada en el logaritmo de las variables
2.  $\ln(\text{variables}) \leq, \geq, =, \text{constante}$
3.  $\ln(\text{flujo}) \leq, \geq, =, \text{constante}$
4. Otras

26

### Paso 2: Linearización y Programación Lineal

Objetivo: Mejora de la producción de L(-)-carnitina

#### Función objetivo:

Maximizar la producción de L-carnitina:  $V_c = Q \cdot C$

$Q(X_3)$  = Caudal extracción

$C(X_5)$  = Concentración carnitina en el caudal de extracción

Max  $V_c = X_3 \cdot X_5$

En coordenadas logarítmicas: Max  $y_3 + y_5$

$y_i = \ln(X_i)$   $i=3,5$

27

### Paso 2: Linearización y Programación Lineal

Objetivo: Mejora de la producción de L(-)-carnitina

#### Restricciones de estado estacionario

La solución óptima corresponderá a un estado estacionario

$$-0.9162 y_1 - 0.5675 y_4 + 0.5675 y_5 + y_6 - 0.5675 y_8 = 0.0211$$

$$-0.142 y_2 + 0.353 y_3 - 0.024 y_4 + 0.024 y_5 + 0.06 y_7 - 0.024 y_9 = 0.6433$$

$$0.1106 y_2 - 0.3926 y_3 + 0.0257 y_4 - 0.0257 y_5 + 0.0254 y_9 = -0.6246$$

$$0.8524 y_1 + y_8 = 2.706$$

$$y_i = \ln(X_i), \quad i=1,2,3,\dots,9$$

28

### Paso 2: Linearización y Programación Lineal

Objetivo: Mejora de la producción de L(-)-carnitina

#### Restricciones a las variables

Se imponen límites superiores a las concentraciones de los intermediarios  $\rightarrow$  garantizar la viabilidad celular (concentraciones intracelulares muy altas o muy bajas  $\rightarrow$  soluciones fisiológicamente inviables).

El conocimiento del sistema nos informa sobre el valor de estas restricciones.

29

### Paso 2: Linearización y Programación Lineal

Objetivo: Mejora de la producción de L(-)-carnitina

Crotonobetaina ( $X_2$ ) y carnitina ( $X_3$ ): 0.5-1.5 val. base

$X_5$ , (Q): 1-1.5

$X_7$ , (Crotonobetaina) y  $X_9$  (KA1): 1-5

$$2.6909 \leq y_2 \leq 3.7895$$

$$2.3277 \leq y_3 \leq 3.4263$$

$$-0.6931 \leq y_5 \leq 0.4054$$

$$3.2188 \leq y_7 \leq 4.3174$$

$$y_i = \ln(X_i); \quad i=2,3,5,7$$

30

### Método Indirecto de Optimización

Paso 3. Transferencia de la solución óptima al modelo original

Variables		Ratio $(X_i)_{opt}/(X_i)_{basal}$		
Dependientes	Basales	1	2	3
		Modelo	Modelo	Modelo
$X_1$ (G)	43.28 mM	1	1	1
$X_2$ (Cr)	29.49 mM	1	1.49	1.53
$X_3$ (C)	20.51 mM	1	1.11	1.18
$X_4$ (X)	12.18 g/L	1.5	1.5	1.5
<b>Independientes</b>				
$X_5$ (Q)	1 L/h <sup>-1</sup>	1.5	1.5	1.5
$X_6$ (Gln)	100 mM	1	1	1
$X_7$ (Cria)	50 mM	1	1.33	1.39
$X_8$ (I <sub>max</sub> )	0.603 h <sup>-1</sup>	1	1	1
$X_9$ (KA1)	123.3 U/h L <sup>-1</sup>	1	1	5
Veloc.	21.1 mM/h <sup>-1</sup>	1.5	1.65	1.77
Producción				

31

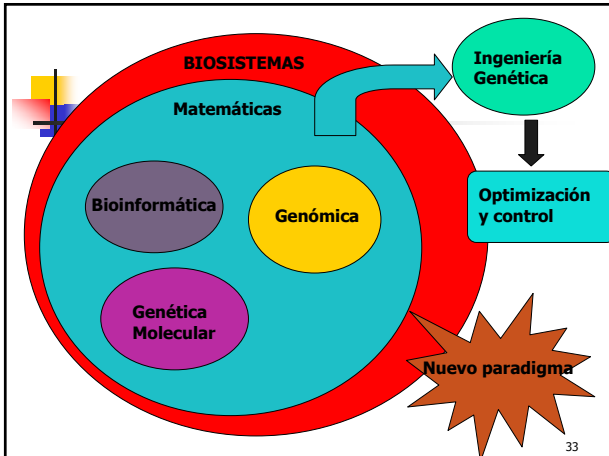
### Método Indirecto de Optimización

Paso 4. Implementación en el sistema experimental

#### Comprobación experimental de las predicciones

Variables		Ratio $(X_i)_{opt}/(X_i)_{basal}$			
Dependientes	Basal	1		2	
		Modelo	Experimental	Modelo	Experimental
$X_1$ (G)	43.28 mM	1	0.96	1	0.95
$X_2$ (Cr)	29.49 mM	1	0.97	1.49	1.77
$X_3$ (C)	20.51 mM	1	1.01	1.11	1.13
$X_4$ (X)	12.18 g/L	1.5	1.53	1.5	1.53
<b>Independientes</b>					
$X_5$ (Q)	1 L/h <sup>-1</sup>		1.5		1.5
$X_7$ (Cria)	50 mM		1		1.33
Veloc.	21.1 mM/h <sup>-1</sup>	1.5	1.54	1.65	1.74
Producción					

32



33

<http://webpages.ull.es/users/sympbst/>

34